Die approbierte Originalversion dieser Dissertation ist in der Hauptbibliothek der Technischen Universität Wien aufgestellt und zugänglich.
http://www.ub.tuwien.ac.at

The approved original version of this thesis is available at the main library of the Vienna University of Technology.

http://www.ub.tuwien.ac.at/eng

TECHNISCHE
UNIVERSITÄT
WIEN

DISSERTATION

# A High Order Discontinuous Galerkin Method for the Boltzmann Equation

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines Doktors der technischen Wissenschaften unter der Leitung von

Univ.Prof. Dipl.-Ing. Dr.techn. Joachim Schöberl
E101
Institut für Analysis und Scientific Computing

eingereicht an der Technischen Universität Wien
an der Fakultät für Mathematik und Geoinformation

von

Dipl.-Ing. Gerhard Kitzler
Matrikelnummer: 0226154
Franz-Guggenberger-Strasse 22/2
2100 Korneuburg

Wien, am 26. November 2015

# Kurzfassung

In der vorliegenden Arbeit, wird eine numerische Methode zur Simulation des Verhaltens eines verdünnten Gases vorgestellt. Das mathematische Modell zur Beschreibung eines solchen Gases ist die Boltzmann Gleichung. Ihre Lösung, die üblicherweise mit $f = f(t, x, v)$ bezeichnet wird, steht für die Anzahl an Teilchen die sich nahe dem Punkt $x$ befinden und Geschwindigkeit nahe zu $v$ haben.

Die im Weiteren vorgestellte Diskretisierung ist eine Petrov-Galerkin Projektion. Für die numerische Approximation der Lösung, welche in $\mathbb{R}^7$ definiert ist, wird ein Tensorprodukt verwendet. Die Testfunktionen sind globale Polynome in der Geschwindigkeitsvariable und lokale, unstetige, stückweise Polynome bezüglich der Orstkoordinate. Diese Testfunktionen liefern die Erhaltung physikalischer Erhaltungsgrößen natürlich. Die Ansatzfunktionen sind ähnlich den Testfunktionen lokale, unstetige, stückweise Polynome bezüglich der Ortsvariablen. In der Geschwindigkeitsvariablen wählen wir den Ansatz globaler Polynome multipliziert mit Gaussfunktionen. Das liefert gute Approximationseigenschaften nahe dem Equilibrium, also nahe der Fluiddynamik.
Aufgrund der Unstetigkeit der Ansatz und Testfunktionen treten Kantenintegrale in der Variationsformulierung auf. Durch die Wahl natürlicher Upwind Flüsse in diesen Integralen wird eine stabile Diskretisierung erzielt.

Der Gausspeak in den Ansatzfunktionen stellt sicher, dass alle Integrale über den unbeschränkten Geschwindigkeitsraum existieren. Für deren Berechnung verwenden wir Gauss Hermite Quadraturformeln. Im Gegensatz zu vielen anderen deterministischen Methoden wird kein zusätzlicher Modellfehler eingeführt, da weder der Geschwindigkeitsraum noch die Integrationsgebiete beschränkt werden müssen.

Die Grundidee wird nun folgendermaßen erweitert: Die Gaussfunktion im Ansatzraum wird elementweise entsprechend der lokalen mittleren Geschwindigkeit und Temperatur geshiftet bzw. skaliert. Diese Parameter werden hierfür aus dem vorigen Zeitschritt ermittelt. Die Approximationseigenschaften des Ansatzraumes werden durch so eine Anpassung sehr stark verbessert, auf der anderen Seite treten Stabilitätsprobleme auf. Durch leichtes Glätten der eben genannten Parameter sind diese jedoch zum großen Teil in den Griff zu bekommen.

Die Berechnung der Kollisionsintegrale hat in numerischen Berechnungen den größten Aufwand. Um diesen zu reduzieren führen wir die numerische Lösung von nodalen auf hierachische Polynome über um den innersten auftretenden Integraloperator in diagonale Form zu bringen. Wir zeigen, wie man spezielle Eigenschaften der Ansatzräume nutzen kann um effizient zwischen den Polynombasen zu transformieren.

Abschließend zeigen wir numerische Ergebnisse als Validierung für das Verfahren. Dabei werden sowohl örtlich homogene als auch inhomogene Probleme gezeigt. Diese zeigen die exzellenten Approximationseigenschaften der angepassten Basisfunktionen, speziell nahe an der Fluiddynamik. Die Rechenzeiten der Beispiele zeigen außerdem den erzielten Geschwindigkeitsvorteil in der Auswertung der Kollisionsintegrale.

# Abstract

In the underlying thesis, we present a numerical method to solve for the behaviour of a dilute gas. The mathematical model behind such a gas is the Boltzmann equation. It's solution, usually denoted by $f = f(t, x, v)$ depends on time, position and velocity and holds the average number of particles having position close to $x$ and velocity close to $v$.

The discretization presented in the sequel is a Petrov-Galerkin projection. To numerically approximate the solution, which is defined in $\mathbb{R}^7$, we use a tensor product. The test functions are global polynomials in the velocity variable and local, discontinuous, piecewise polynomials in the position variable. These test functions yield the conservation of physically conserved properties naturally. The trial functions are similar to the test functions chosen as discontinuous, piecewise polynomials in the spatial variable. In the velocity variable, we take an approach of global polynomials multiplied with Gaussian peaks. This gives good approximation properties of solutions close to equilibrium and thus, close to the fluid regime.
The discontinuities of the trial and test functions yield skeleton integrals in the variational formulation. By choosing natural upwind fluxes in these skeleton integrals a stable discretization is achieved.

The Gaussian peak in the trial functions ensures additionally that all integrals over the unbounded momentum domain exist. For the evaluation we use the Gauss Hermite quadrature rules. In contrast to many other deterministic methods there is no additional modelling error due to domain truncation.

We extend the main idea in the following way: we shift and scale the Gaussian peaks element wise according to the gas' local mean velocity and temperature calculated from the previous time step. The approximation properties of the trial space are greatly enhanced by such a dependency on the solution. On the other hand, stability is decreased. By smoothing the above mentioned parameters mean velocity and temperature slightly, the stability issue can be avoided for the most part.

The evaluation of the collision integrals in actual computations is a critical part since this involves a lot of numerical work. To reduce the complexity in the calculations we transform the solution from nodal to hierarchical polynomials to arrive at an inner integral operator in diagonal form. We show how to use the properties of the trial spaces to execute this transformations efficiently.

Finally we show a lot of numerical examples as a validation for the method. This includes space homogeneous as well as space dependent problems. The results demonstrate the excellent approximation properties of the shifted and scaled basis functions, especially close to the fluid regime. In addition, the computation times show the speed up achieved by the evaluation techniques for the collision integral.

# Acknowledgements

The present thesis was written during my work at Vienna University of Technology at the Institute for Analysis and Scientific Computing.

I would like to express my gratitude to my advisor professor Joachim Schöberl for his encouraged supervision during the past years. Many of our intensive communications ended up in very helpful ideas. I want to appreciate this here once more.
I'm very grateful to professor Christian Schmeiser from the University of Vienna for his readiness to review the thesis.

Next I want to mention my colleagues in our research group Christian Aumayr, Haik Davtjan, Martin Halla, Antti Hannukainen, Matthias Hochsteger, Karl Hollaus, Martin Huber, Philip Lederer, Christoph Lehrenfeld, Lothar Nannen, Markus Schöbinger, Markus Wess, Christoph Wintersteiger and Anna Zechner. I'm grateful for all their scientific support, nice social events, as well as for their motivation during strenuous times.

Moreover, I owe special thanks to all of my friends for taking my mind off mathematical and programming stuff.

I want to thank my parents for their support during my master and PHD study.

Finally I want to express my gratefulness to my wife Daniela for her open ears and patience.

Sincere thanks are given to all mentioned and unmentioned supporters during the last years.

# Contents

**1 Introduction** **1**

1.1 The Boltzmann Equation . . . . . 1
  1.1.1 Deriving The Boltzmann Equation . . . . . 2
  1.1.2 Fields of applications . . . . . 11
1.2 The Problem Setting . . . . . 12
  1.2.1 Boundary conditions . . . . . 13
1.3 Numerical Challenges . . . . . 15
1.4 Outline . . . . . 16

**2 Elementary properties of solutions** **19**

2.1 Moments and Macroscopic Properties . . . . . 19
2.2 Properties of the Collision operator . . . . . 22
  2.2.1 Collision Invariants . . . . . 24
  2.2.2 Solutions of $Q(f) \equiv 0$ . . . . . 26
2.3 Boltzmann H-theorem . . . . . 27
  2.3.1 The Space Homogeneous case . . . . . 28
  2.3.2 The Space Inhomogeneous case . . . . . 29
2.4 Moment equations . . . . . 31
  2.4.1 From Boltzmann to Euler . . . . . 33
  2.4.2 From Boltzmann to Navier-Stokes . . . . . 34

**3 Numerical Methods** **39**

3.1 Stochastic methods . . . . . 39
  3.1.1 Monte Carlo methods – Nanbu - Babovsky scheme . . . . . 39
  3.1.2 Time relaxed Monte Carlo methods . . . . . 42
3.2 Deterministic Methods . . . . . 45
  3.2.1 Fourier series expansion . . . . . 45
  3.2.2 Discrete velocity models . . . . . 50

**4 A Discontinuous Galerkin approach** **55**

4.1 Discontinuous Galerkin Discretization . . . . . 55
  4.1.1 DG for a scalar linear convection equation . . . . . 56
  4.1.2 Applying DG to the Boltzmann transport operator . . . . . 59

4.1.3 Polynomial basis in $\mathbf{V_{h,N}}$ . . . . . . . . . . . . . . . . . . . . . . . . . . . 63
4.2 Application of the collision integrals . . . . . . . . . . . . . . . . . . . . . . . . . . 65
4.3 Efficient algorithm for the collision operator . . . . . . . . . . . . . . . . . . . . . . 68
4.3.1 Approximation on shifted grids . . . . . . . . . . . . . . . . . . . . . . . . 68
4.3.2 The integral w.r.t. the mean velocity . . . . . . . . . . . . . . . . . . . . . 73
4.3.3 Hermite and Polar-Laguerre polynomial bases . . . . . . . . . . . . . . . . 74
4.3.4 Efficient transformation between polynomial bases . . . . . . . . . . . . . 81
4.3.5 Transformations of the test functions . . . . . . . . . . . . . . . . . . . . . 86
4.4 Adaptive choice of element Maxwellians . . . . . . . . . . . . . . . . . . . . . . . . 90

**5 Numerical results** **95**
5.1 Space homogeneous problems . . . . . . . . . . . . . . . . . . . . . . . . . . . . 95
5.1.1 BKW Solution . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 95
5.2 Space dependent problems . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 98
5.2.1 Model problem 1 - Flow around a wedge . . . . . . . . . . . . . . . . . . . 98
5.2.2 Model problem 2 - Flow around a cylinder . . . . . . . . . . . . . . . . . . 103
5.2.3 Flow around NACA 7410 air foil - specular reflection . . . . . . . . . . . . 106
5.2.4 Flow around NACA 7410 air foil - diffuse reflection . . . . . . . . . . . . 113
5.2.5 Mach 3 wind tunnel with step . . . . . . . . . . . . . . . . . . . . . . . . 115
5.2.6 The Knudsen pump . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 118

**6 Summary and Outlook** **123**
6.1 Summary and Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 123
6.2 Outlook . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 124

**7 Bibliography** **127**

# 1 Introduction

In many industrial applications the knowledge of the flow around an obstacle is important: In aeronautics, the resulting flow around an air foil enables one to calculate lift and drag force acting on the air foil. Due to immense expenses needed for wind tunnel experiments and the capability of modern computers to solve for such computational intensive problems, numerical simulation has become more and more relevant. Typically people solve the system of Navier-Stokes or Euler equations to obtain the flow field. The Navier-Stokes equations are capable to model friction and thus include turbulences and also hydrodynamic boundary layers.

In contrast to Navier-Stokes, the Euler equations do not include turbulences, neither boundary layers. However, if no boundary layers or turbulences occur, the Euler equations provide a satisfactory description of the flow.

As the gas becomes more and more dilute, the description by the above models is not satisfactory any more since they are based on a local thermodynamic equilibrium of the flow. This requires a sufficiently large amount of collisions. If the gas becomes more and more rarefied there are too few collisions to guarantee local equilibrium and thus we need an alternative model for such a gas.

## 1.1 The Boltzmann Equation

The Boltzmann equation is a model for transport phenomena in a sufficiently dilute gas. The equation models the transport (free flow) of particles and in addition a relaxation process (particle interaction) to equilibrium. In the simplest case, the interaction can be imagined as hard sphere collisions. In terms of the Boltzmann equation, the gas is described by a density distribution function, usually denoted by $f$. The Boltzmann framework may be seen as a reduced description of the microscopic state (relaxation is modelled in the mean). Clearly, a complete description incorporates the position and momenta of all particles under consideration. For a realistic situation this would incorporate approximately $10^{19}$ particles, making a direct approach unsuitable and thus, gives need to study the Boltzmann equation. The equation can be rigorously derived from a system of $N$ particles which are moving according to Newtons laws. A small gap is left which can be closed by the Boltzmannschen Stozahlansatz.

### 1.1.1 Deriving The Boltzmann Equation

At the beginning we give a short presentation of the derivation of the Boltzmann equation. It is essentially based on the derivation in [Cer90, CIP94].

We consider the motion of $N$ particles with positions $x_i, i = 1 \dots N$ and velocities $v_i, i = 1 \dots N$. If no interaction occurs, the evolution of the particles is given by

$$\dot{x}_i = v_i \qquad \dot{v}_i = X_i, \tag{1.1.1}$$

where $X_i$ is the force acting on the $i$-th particle divided by its mass. We look for an alternative representation of this evolution given by one single function depending on all the positions and velocities

$$C : \begin{cases} \mathbb{R} \times \Omega^N \times \mathbb{R}^{3N} \to \mathbb{R} \\ (t, x_i \dots x_n, v_1, \dots v_N) \mapsto C(t, x_i \dots x_N, v_1, \dots v_N) \end{cases}, \tag{1.1.2}$$

such that $C = 1$ if $x_i = x_i(t), i = 1 \dots N$, $v_i = v_i(t), i = 1 \dots N$ and 0 otherwise. We denote the space $\Omega^N \times \mathbb{R}^{3N}$ as the phase space of the system of the $N$ particles.
The function C can be written as a product of $\delta$ distributions,

$$C(t, x_1 \dots x_N, v_1 \dots v_N) = \prod_{i=1}^{N} \delta(x_i - x_i(t))\delta(v_i - v_i(t)). \tag{1.1.3}$$

Its time derivative satisfies

$$\frac{\partial C}{\partial t} + \sum_{i=1}^{N} v_i \frac{\partial C}{\partial x_i} + \sum_{i=1}^{N} X_i \frac{\partial C}{\partial v_i} = 0. \tag{1.1.4}$$

The last equation provides a governing equation for the evolution of the system in terms of the certainty density $C$. (1.1.4) is called the Liouville equation. Assuming that one is able to obtain the initial data of a set of $N \approx 10^{19}$ particles exactly, the Liouville equation would – supplemented with appropriate boundary conditions – describe the gas entirely for each $t > 0$. However, it is unfortunately not possible to obtain the initial data exactly. Therefore we drop the assumption that we know each particles position and momentum exactly and replace $C$ by a probability density $P = P(t, x_1, \dots x_N, v_1, \dots v_N)$.

$$\text{Prob}((x_1 \dots x_N, v_1 \dots v_N)) \in D) = \int_D P \, dx_1 \dots dx_N \, dv_1 \dots v_N, \tag{1.1.5}$$

where $D$ is a region of the phase space. We note that – under the absence of particle interactions – also $P$ satisfies the Liouville equation (1.1.4). Due to the consideration of identical particles we assume that the function $P$ is symmetric w.r.t. the particles positions and velocities.

**Remark 1.1.1.** *Describing the system by the certainty density $C$ and changing it afterwards to a description by the probability density $P$ is the approach presented in [Cer90]. In [CIP94] Cercignani starts with a description of the system under consideration with the probability density $P$. By a consideration of the time evolution of the points in phase space he also ends up with the Liouville equation for $P$.*

Next we want to consider particle interactions in addition. For simplicity we consider hard sphere molecules which interact like billiard balls, thus with perfectly elastic collisions. A collision happens if we are at a point in phase space such that $\exists \, i, j \in \{1 \dots N\}, i \neq j$ with $|x_i - x_j| = \sigma$, where $\sigma$ is the molecular diameter. If we consider only points in phase space such that $|x_i - x_j| > \sigma, i, j = 1 \dots N, i \neq j$, then no particle interaction takes place and the evolution is still governed by the Liouville equation

$$\frac{\partial P}{\partial t} + \sum_{j=1}^{N} v_j \frac{\partial P}{\partial x_j} = 0, \tag{1.1.6}$$

with $X_i = 0$. Note that $P \equiv 0$ if $\exists \; i, j : \; |x_i - x_j| < \sigma$, since such a state of the system can not be attained.

Our goal is now to derive an evolution equation for the one particle distribution function, that is the probability density to find a certain particle 1 at position $x_1$ having a velocity $v_1$ and the other particles to have any position and velocity. This function is given by

$$P^{(1)}(t, x_1, v_1) := \int_{\mathbb{R}^{3N-3}} \int_{\Omega^{N-1}} P(t, x_1 \dots x_N, v_1 \dots v_N) \, dx_2 \dots dx_N \, dv_2 \dots dv_N. \tag{1.1.7}$$

It is intuitive how we can generalize this definition to a $s$-particle probability density function.

In order to obtain an evolution equation for $P^{(1)}$ we integrate the Liouville equation over the positions and velocities of all particles except the first one. The integration domain for the velocities $v_j, j = 2\dots N$ is the whole space $\mathbb{R}^{3N-3}$, while for the positions $x_j, j = 2\dots N$ we have the integration domain $\Omega^{N-1}$ without the points that satisfy $|x_i - x_j| < \sigma$ for at least one pair $(i, j)$, $i, j = 1\dots N, i \neq j$. In order to perform the following steps we have to assume sufficient regularity on the probability density $P$.
We start with integrating the time derivative of $P$ resulting in

$$\int\int \frac{\partial P}{\partial t}\, dx_2 \dots dx_N\, dv_2 \dots dv_N = \frac{\partial P^{(1)}}{\partial t}, \tag{1.1.8}$$

where we just had to interchange the order of differentiation and integration.

Integrating the first term $v_1 \frac{\partial P}{\partial x_1}$ in the sum of (1.1.6) has to be done carefully if differentiation and integration shall be exchanged since the boundaries for the positions also depend upon $x_1$. By applying Leibnitz' rule we obtain the correct result

$$\int\limits_{\mathbb{R}^3}\int\limits_{B_N} v_1 \cdot \frac{\partial P}{\partial x_1}\, dx_N\, dv_N = v_1 \cdot \left( \frac{\partial}{\partial x_1} \int\limits_{\mathbb{R}^3}\int\limits_{B_N} P\, dx_N\, dv_N - \int\limits_{\mathbb{R}^3}\int\limits_{\partial B_N} v_{\partial B_N}^T n P\, ds(x_N) dv_N \right). \tag{1.1.9}$$

The integration domain $B_N$ for the $N$-th particle is given by $\{x \in \Omega : |x_i - x_N| > \sigma, i = 1\dots N-1\}$. $n$ is the outer unit normal vector to the boundary $\partial B_N$ and $v_{\partial B_N}$ is the "velocity" of the boundary w.r.t $x_1$. This velocity is given by the identity matrix for the part of $\partial B_N$ with $|x_1 - x_N| = \sigma$ and is zero for the other parts of $\partial B_N$. Using this fact, we can restrict the boundary integral to the sphere $S_{x_1,\sigma}$ with center at $x_1$ and radius $\sigma$. Now we integrate both sides of (1.1.9) w.r.t. $x_{N-1}, v_{N-1}$ and apply Leibnitz' rule again. By this procedure, we can interchange the order of integration w.r.t. $x_{N-1}, v_{N-1}$ and differentiation w.r.t $x_1$. After $(N-2)$ steps we end up with

$$\int v_1 \cdot \frac{\partial P}{\partial x_1}\, dx_2 \dots dx_n\, dv_2 \dots dv_N = v_1 \cdot \frac{\partial P^{(1)}}{\partial x_1} - \sum_{j=2}^{N} \int\limits_{\mathbb{R}^3}\int\limits_{S_{x_1,\sigma}} v_1 \cdot n_j P^{(2)}\, ds(x_j) dv_j. \tag{1.1.10}$$

In the above equation, $n_j$ denotes the inner normal vector to the sphere $S_{x_1,\sigma}$ at the point $x_j$. Having a closer look at the above equation, we note that the integrals are independent of the index $j$ and therefore we can omit it in the sequel. Instead of $x_j$ and $v_j$ we write $x^*$ and $w$ in the sequel.

For the above sum we now obtain

$$\int v_1 \cdot \frac{\partial P}{\partial x_1}\, dx_2 \dots dx_n\, dv_2 \dots dv_N = v_1 \cdot \frac{\partial P^{(1)}}{\partial x_1} - (N-1) \int\limits_{\mathbb{R}^3} \int\limits_{S_{x_1,\sigma}} v_1 \cdot n P^{(2)}\, ds(x^*) dw. \tag{1.1.11}$$

The arguments of $P^{(1)}$ are $t, x_1, v_1$ and for $P^{(2)}$ we have the arguments $t, x_1, x^*, v_1, w$.

Next, we want to interchange the order of differentiation and integration in the terms with $j > 1$ in (1.1.6). Now the derivative is taken w.r.t. one if the integration variables in contrast to the term with $j = 1$. Therefore, one can directly apply the Gaussian theorem yielding

$$\begin{aligned} &\int v_j \cdot \frac{\partial P}{\partial x_j}\, dx_2 \dots dx_N\, dv_2 \dots dv_N \\ &= \int \left( \int\limits_{\mathbb{R}^3} \int\limits_{\partial B_j} n_{\partial B_j} \cdot v_j P ds(x_j) dv_j \right) dx_2 \dots x_{j-1} x_{j+1} \dots dx_N\, dv_2 \dots dv_{j-1} dv_{j+1} \dots dv_N \end{aligned} \tag{1.1.12}$$

In the above equation, the integration domain $B_j$ for $x_j$ is given by $\{x \in \Omega : |x_i - x_j| > \sigma, i = 1 \dots N, i \neq j\}$, $n_{\partial B_j}$ denotes the outer unit normal vector at a point in $\partial B_j$. We note that the boundary integral over $\partial B_j$ can be split in a sum of integrals over the spheres $S_{x_i,\sigma}$ what we do in the next step. To be more precise, $\int_{\partial B_j}$ consists of one additional part, the integral over the boundary of the domain $\Omega$ itself. As is shown in [CIP94] this term vanishes if the boundary is such that particles are perfectly reflected, see (1.2.4a).

In the integral consisting of the part of the boundary where $|x_1 - x_j| = \sigma$, we execute the integration w.r.t. all variables except $x_j$ and $v_j$. In the other terms consisting of the boundaries $|x_i - x_j| = \sigma, 1 < i \leq N$ we execute the integration w.r.t. all variables except $x_j$, $x_i$, $v_j$ and $v_i$. This leads to

$$\begin{aligned} &\int v_j \cdot \frac{\partial P}{\partial x_j}\, dx_2 \dots x_N\, dv_2 \dots dv_N \\ &= \int\limits_{\mathbb{R}^3} \int\limits_{S_{x_1,\sigma}} n \cdot v_j P^{(2)}\, ds(x_j) dv_j + \sum_{\substack{i=2 \\ i \neq j}}^{N} \int\limits_{\mathbb{R}^3} \int\limits_{B_i^1} \int\limits_{\mathbb{R}^3} \int\limits_{S_{x_i,\sigma}} n_i \cdot v_j P^{(3)}\, ds(x_j) dv_j dx_i dv_i. \end{aligned} \tag{1.1.13}$$

As before, $n$ and $n_i$ are the inner unit normal vectors of the spheres with center $x_1$ and $x_i$ respectively. The notation for the integration domain $B_i^1$ of $x_i$ shall reflect that $x_i$ is integrated over all positions $x \in \Omega$, excluding those points where $|x_i - x_1| < \sigma$. In the first integral $P^{(2)}$ is evaluated at $t, x_1, x_j, v_1, v_j$. In the second term $P^{(3)}$ is evaluated at $t, x_1, x_j, x_i, v_1, v_j, v_i$. We note, that the first term in the above equation is independent of the index $j$ and thus, if we sum according to 1.1.6 from $j = 2 \dots N$, we obtain this term $N-1$ times. As we did before, we write $x^*$ and $w$ instead of $x_j$ and $v_j$ in this term in the sequel.

In the next step we want to show that the last integral in (1.1.13) which involves the 3 particle distribution function vanishes. The first thing we do to come to this end is rewriting the integral.

$$\int\limits_{\mathbb{R}^3}\int\limits_{B_i^1}\int\limits_{\mathbb{R}^3}\int\limits_{S_{x_i,\sigma}} n_i \cdot v_j P^{(3)}\, ds(x_j) dv_j dx_i dv_i = \frac{1}{2}\int\limits_{\mathbb{R}^3}\int\limits_{B_i^1}\int\limits_{\mathbb{R}^3}\int\limits_{S_{x_i,\sigma}} n_i \cdot (V_{ji}) P^{(3)}\, ds(x_j) dv_j dx_i dv_i \tag{1.1.14}$$

with the relative velocity $V_{ji} = v_j - v_i$.
The second ingredient we need is a condition on the behaviour of $P$ when 2 particles collide. Since $P$ has to be constant along the trajectories of a point $z$ in phase space this condition is

$$\begin{aligned} &P(t, x_1 \dots x_N, v_1 \dots v_i \dots v_j \dots v_N) \\ &= P(t, x_1 \dots x_N, v_1 \dots v_i - n(n \cdot V_{ij}) \dots v_j + n(n \cdot V_{ij}) \dots v_N) \quad \text{if} \quad |x_i - x_j| = \sigma, \end{aligned} \tag{1.1.15}$$

with the relative velocity $V_{ij} = v_i - v_j$ and the unit vector $n = \frac{x_i - x_j}{|x_i - x_j|}$.
In order to see that the integral vanishes, we now consider 2 velocities $v_i$, $v_j$, such that $n_i \cdot V_{ji} > 0$. The velocities $v_i' := v_i - n_i(n_i \cdot V_{ij})$ and $v_j' := v_j + n_i(n_i \cdot V_{ji})$ satisfy $n_i \cdot (v_i' - v_j') = -n_i \cdot V_{ji} < 0$. Therefore, by the transform $(v_i, v_j) \mapsto (v_i', v_j')$, we can map each point on the hemisphere defined via $n_i \cdot V_{ji} > 0$ to a point on the hemisphere $n_i \cdot V_{ji} < 0$, such that $P^{(3)}$ has the same value at these two points, but the factor in front of $P^{(3)}$ has opposite sign at these two points. Thus, the two contributions cancel each other.

The remaining terms yield

$$\frac{\partial P^{(1)}}{\partial t} + v_1 \frac{\partial P^{(1)}}{\partial x_1} = (N-1) \int\limits_{\mathbb{R}^3}\int\limits_{S_{x_1,\sigma}} n \cdot (v_1 - w) P^{(2)} ds(x^*) dw, \tag{1.1.16}$$

where $P^{(1)}$ and its derivative are evaluated at $(t, x_1, v_1)$ and $P^{(2)}$ in the integral is evaluated at $(t, x_1, x^*, v_1, w)$. In order to simplify notation we drop the subscript 1 from $x_1$ and $v_1$ in the sequel since it is not needed any more.
To end up with the desired result, it is convenient to split the integral over the sphere $S_{x,\sigma}$. For a given pair of velocities $(v, w)$ we consider the contributions from $S_+ := \{y \in S_{x,\sigma} : n \cdot (v - w) > 0\}$ and $S_- := S_{x,\sigma} \setminus S_+$ separate. Due to a better readability we omit the dependency of these hemispheres on $v - w$ in the notation.

The contribution from $S_-$ corresponds to those particle pairs which are going to collide, from the plus hemisphere we obtain the contribution of the collisions that have immediately happened.
In both hemispheres we now interchange the surface element $ds$ of the sphere with radius $\sigma$ by that of the unit sphere denoted by $dn$, with the relation to $ds$ given by $\sigma^2 dn = ds$. This transforms the right hand side of (1.1.16) into

$$(N-1)\sigma^2 \left( \int\limits_{\mathbb{R}^3} \int\limits_{S_+} P^{(2)} |n \cdot (v-w)| \, dn(x^*) dw - \int\limits_{\mathbb{R}^3} \int\limits_{S_-} P^{(2)} |n \cdot (v-w)| \, dn(x^*) dw \right). \quad (1.1.17)$$

To obtain an evolution equation for $P^{(1)}$ we have to get rid of $P^{(2)}$ in the collision integrals on the right hand side of (1.1.16). To come to that end we consider Boltzmanns Stozahlansatz. To that end we think of the following situation: Let the gas be enclosed in a box of size 1 cm$^3$ and assume that the number of molecules inside the box is large $\approx 10^{20}$, while it's size is rather small $\sigma \approx 10^{-8}$ cm. Thus, the volume occupied by the particles is in the order of $N\sigma^3 \approx 10^{-4}$ cm$^3$ and is very small compared to the size of the box. Therefore, if we fix a pair of particles, we find that a collision between them is a very rare event. Thus, we can think of two particles that are going to collide as 2 randomly chosen particles which moved independent from each other. Thus, for particles which are going to collide we write

$$P^{(2)}(t,x,x^*,v,w) = P^{(1)}(t,x,v) P^{(1)}(t,x^*,w) \quad \text{if} \quad n \cdot (v-w) < 0. \quad (1.1.18)$$

We can use the above factorization, stating statistical independence on the minus hemisphere. Clearly, the collision creates a strong correlation between these two particles and we can not apply the above splitting to the plus hemisphere (As explained in [Cer90], one would obtain a zero collision contribution if one does). In order to factorise $P^{(2)}$ also on the plus hemisphere, we express the collision in terms of the ingoing configuration $v' = v - nn \cdot (v-w)$ and $w' = w + nn \cdot (v-w)$. According to (1.1.15) this does not change $P^{(2)}$. Now we have $P^{(2)}$ evaluated at an ingoing collision configuration, and therefore we can again apply the idea of statistical independence and write on the plus hemisphere

$$P^{(2)}(t,x,x^*,v',w') = P^{(1)}(t,x,v') P^{(1)}(t,x^*,w') \quad \text{if} \quad n \cdot (v-w) > 0. \quad (1.1.19)$$

If we accept these simplifying arguments of Boltzmann we obtain for the collision integral

$$(N-1)\sigma^2 \int \int\limits_{S_+} \Big( P^{(1)}(t,x,v')P^{(1)}(t,x-n\sigma,w') - P^{(1)}(t,x,v)P^{(1)}(t,x+n\sigma,w) \Big) |n\,V|\, dn dw, \tag{1.1.20}$$

where $V = v - w$, $v' = v - n(nV)$ and $w' = w + n(nV)$. In the above equation we have interchanged the direction of the normal vector on the minus hemisphere to write it as an integral over the plus hemisphere also. Additionally we used that the points $x^*$ on the hemispheres are given by $x \pm n\sigma$.

Finally we perform the Boltzmann-Grad limit, that is $N \to \infty$, $\sigma \to 0$ such that $N\sigma^2$ remains finite. Accordingly we have $x \pm n\sigma \to x$, resulting in

$$\frac{\partial P^{(1)}}{\partial t} + \mathrm{div}_x(vP^{(1)}) + \int \int\limits_{S_+} \Big( P^{(1)}(v')P^{(1)}(w') - P^{(1)}(v)P^{(1)}(w) \Big)\, |n\,V|\, dn dw = 0. \tag{1.1.21}$$

Here we have omitted the time and space dependency in the integrand since it is $t$ and $x$ throughout. In addition we may also omit the superscript $(1)$ from $P^{(1)}$ since it is not needed any more.
As stated in [CIP94], the integration over the plus hemisphere can be extended to the whole surface of the sphere which requires a multiplication of the result by $\frac{1}{2}$.

It shall be noted, that there are different representations for the collision operator. The above representation uses the unit vector $n$, joining the centers of the 2 spheres at the instant of the collision. This vector is often denoted as the angle of collision. Another way to describe the collision is by the use of the scattering vector which we denote by $e'$. The scattering vector represents the directions of the outgoing particles relative to their mean velocity. In Figure 1.1.1 these 2 approaches for describing a collision are sketched.
The connection between the angle of collision and the scattering vector is

$$e' = \frac{v-w}{|v-w|} - 2n\, n \cdot \frac{v-w}{|v-w|}. \tag{1.1.22}$$

The post collision velocities are represented by the scattering vector via

$$v' = \frac{v+w}{2} + e'\frac{|v-w|}{2} \qquad \text{and} \qquad w' = \frac{v+w}{2} - e'\frac{|v-w|}{2}, \tag{1.1.23}$$

and the collision operator reads

$$Q(P)(t,x,v) := \int_{\mathbb{R}^3}\int_{S^2} B(v,w,e')\left[P(v')P(w') - P(v)P(w)\right] de'\, dw, \tag{1.1.24}$$

with the collision kernel $B(v,w,e')$ given by $B(v,w,e') = |v-w|$. In the sequel we deal with the collision representation in terms of the scattering vector $e'$.

Since rotational symmetry is demanded from the interaction law, the dependency of $B$ on its input arguments reduces to a dependency on the relative velocity and on the angle formed by the scattering direction and the relative velocity:

$$B(v,w,e') = B(|v-w|, \frac{(v-w)\cdot e'}{|v-w|}).$$

The kernel we obtained in derivation corresponding to hard sphere collisions is the physically best justified interaction law for binary particle interaction. In the derivation we implicitly assumed a finite interaction distance given by the molecular diameter $\sigma$. In contrast to hard sphere interaction, an interaction law in terms of an inverse power law consists of an infinite interaction distance, leading to a collision kernel of the following form [RW05]

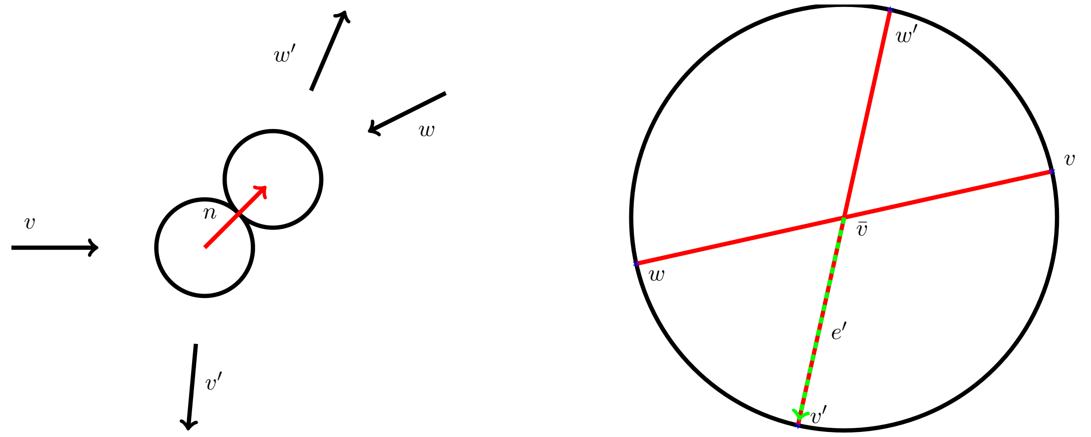
$$B(v,w,e') = |v-w|^{1-\frac{4}{\alpha}} b_\alpha(\theta)\sin(\theta)^{-1},$$

with $\theta = \arccos\left(\frac{e'\cdot(v-w)}{|v-w|}\right)$. Such interaction laws lead to a huge number of grazing collisions, i.e. collisions with $v' \approx v$ and $w' \approx w$. As a consequence, the differential cross section $|v-w|^{-\frac{1}{4}} b_\alpha(s)$ becomes singular at $s = 0$ and is even not integrable. In that case, the usual splitting of the collision operator into its gain and loss term is not applicable what is a crucial point in many numerical approaches.

Often one does not solve for the one particle probability density $P$, but one introduces a quantity close related to $P$ and denotes it by $f$ usually.

$$f(t,x,v) = P(t,x,v)Nm, \tag{1.1.25}$$

where $N$ is the number of particles under consideration and $m$ the mass of the particles. The unknown $f$ is per construction the mass density at a single point $(x,v)$ in phase space at time $t$.



(a) A graphical representation of the angle of collision. The circles represent colliding particles, the incoming velocities are $v$ and $w$ respectively, the outgoing are denoted with as $v'$ and $w'$. The angle of collision $n$ is the connection of the 2 midpoints at the instant of collision.

(b) Here we show a graphical representation of the scattering vector $e'$. W. r. t. the mean velocity, the pre as well as the post collision velocities lie diametrically opposite on a circle with its center at the mean velocity and its diameter being the relative velocity.

Figure 1.1.1: Two different representations of a binary collision.

Thus,

$$\int\limits_{\Omega\times\mathbb{R}^3} f(t,x,v)\,d(x,v) = Nm \tag{1.1.26}$$

gives the total mass of the gas. In addition, also $\frac{f(t,x,v)}{m}$ giving the number density in phase space is used in literature.

**Remark 1.1.2.** *The Boltzmann-Grad limit, $N \to \infty$, $\sigma \to 0$, $N\sigma^2$ stays finite, means that the total volume occupied by the gas tends to $0$, i.e. is of 0 Lebesgue measure. Thus, it should be possible to "squeeze" all particles inside the domain of interest into a plane.*

**Remark 1.1.3.** *It shall be noted here, that the presented derivation of the collision integral via integrating the $N$-particle distribution function was not yet done by Boltzmann. He combined arguments of probabilistic nature, e.g:*

*(1) How many particles can undergo a certain collision with prescribed outcome?*

*(2) Assume that particles before a collision are statistically uncorrelated.*

*(3) Use the direct relation between pre and post collision velocities to argue for the splitting of $P^{(2)}$ at those particles that have recently collided.*

### 1.1.2 Fields of applications

In kinetic theory of dilute gases the Boltzmann equation forms a fundamental basis. Recent developments in aerospace, demand highly accurate solutions of the equation, for instance in re entry problems.

Besides classical gases Boltzmann considered, proper generalizations of the Boltzmann equation are found in electron transport in solids and plasmas also. The transport of charged electrons through a semi conductor device can be described by a kinetic transport model, in addition external forces acting on the electrons (the electric field) are present. The actual length scales of such devices require an additional treatment by equations from quantum mechanics [Jün09].

Besides electron transport, neutron transport in nuclear reactors, phonon transport in super fluids, and radiative transfer in planetary and stellar atmospheres can be modelled by kinetic transport models.

In addition to classical applications, kinetic equations have attracted other scientific disciplines as for instance applications in biology used for swarm modelling. A swarm in that sense is a collection of at least 10, up to millions of individuals. Classically, mathematical models are derived from first principles of swarming. These include the social tendency to form groups (attraction), the space around each individual to feel comfortable in the group (collision avoidance) and the synchronisation with a group (alignment). Classical models are based on simulation of each individual resulting in a large system of Odes. Besides them, kinetic models are likely used – the individuals are described by a density distribution as in the theory of rarefied gases. Particle interactions are based on the above considerations about attraction, collision avoidance and alignment. These interaction model is known as a three zone model. Figure 1.1.2, which appears in [CFTV10] represents these zones.

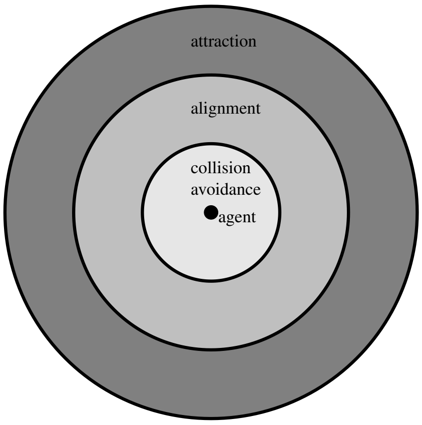


Figure 1.1.2: A graphical representation of the three zone model. The ”particle interaction” is defined separately for each of those zones.

## 1.2 The Problem Setting

In the current section we present the exact form of the equation we are solving numerically, the collision kernels we are able to handle in our method and the boundary conditions we considered in our implementation. The numerical method presented in section 4 is presented for 2 spatial and 2 momentum dimensions. In the remaining sections we will work with $d$ spatial and also $d$ momentum dimensions with $d = 2$ or $d = 3$.
We denote the spatial domain by $\Omega \subset \mathbb{R}^d$. We describe the gas by the one particle distribution function which we denote by $f = f(t, x, v) \geq 0$. The form of the Boltzmann equation we use reads

$$\frac{\partial}{\partial t} f + \mathrm{div}_x(vf) = \frac{1}{\mathrm{Kn}} Q(f) \qquad x \in \Omega, v \in \mathbb{R}^d, t \geq 0, \tag{1.2.1}$$

with $\mathrm{div}_x$ being the divergence operator with respect to the spatial coordinate $x$. $Q(f)$ denotes the Boltzmann collision operator, given by:

$$Q(f)(t, x, v) := \int\limits_{\mathbb{R}^d} \int\limits_{S^{d-1}} B(v, w, e') \left[ f(t, x, v') f(t, x, w') - f(t, x, v) f(t, x, w) \right] de\, dw. \tag{1.2.2}$$

The Knudsen number is defined as $\mathrm{Kn} := \frac{\lambda}{L}$, where $\lambda$ is the mean free path of the particles between subsequent collision and $L$ is a typical length scale of the problem. Note that $x$, $v$ and $t$ are scaled by typical length, velocity and time of the problem. The post collision velocities are defined as in the previous section in terms of the scattering vector $e' \in S^{d-1}$ via

$$v' = \frac{v + w}{2} + e' \frac{|v - w|}{2} \qquad w' = \frac{v + w}{2} - e' \frac{|v - w|}{2} \qquad e' \in S^{d-1}. \tag{1.2.3}$$

For the numerical method presented in section 4, we have to assume a separable collision kernel $B$ such that

$$B(v, w, e') = b_r(|v - w|) b_\theta\left(\tfrac{(v-w)\cdot e'}{|v-w|}\right).$$

In order to treat the gain and loss terms separate, the differential cross section has to be integrable, thus the function $b_\theta$ has to satisfy Grad's cut off assumption in addition:

$$\int_0^\pi b_\theta(s)\, ds < \infty.$$

Finally, we assume a power law dependency of $b_r$ on its argument $|v-w|$, such that

$$b_r(r) = r^\beta,$$

for some exponent $\beta \in (-3, 1)$. These assumptions are natural for a wide range of collision kernels, including (Pseudo) Maxwellian gases as well as variable hard sphere gases.

### 1.2.1 Boundary conditions

As a consequence of the first order time and space derivative the Boltzmann equation has to be supplied with initial- as well as boundary conditions.
The initial condition reads

$$f(0, x, v) = f_0(x, v),$$

describing the gas at time $t = 0$. The boundary conditions shall reflect the interaction of the particles with $\partial\Omega$ which may be a wall or an open boundary as well. As a preparation we introduce the incoming and outgoing directions at a certain point $x \in \partial\Omega$.

$$\mathbb{R}^d_{\text{in}} := \{v \in \mathbb{R}^d : v \cdot n < 0\} \quad \text{and} \quad \mathbb{R}^d_{\text{out}} := \mathbb{R}^d \setminus \mathbb{R}^d_{\text{in}},$$

where $n$ is the outer normal vector in a point $x \in \partial\Omega$. Due to readability we suppress the $x$ dependency of these sets.
We consider the following conditions at $\partial\Omega$:

- specular reflection

  $$f(t, x, v) = f(t, x, v - 2n(x) \cdot vn(x)) \quad \forall v \in \mathbb{R}^d_{\text{in}}. \tag{1.2.4a}$$

  Particles hitting the wall behave like billiard balls. The tangential component of the particles' velocity does not change, while the normal component is multiplied by -1. Since $v - 2v \cdot nn \in \mathbb{R}^2_{\text{out}}$, this boundary condition states a direct relation between incoming and outgoing particles.

- diffuse reflection

$$f(t,x,v) = ce^{-\left|\frac{v-V_{\text{bnd}}}{\sqrt{T_{\text{bnd}}}}\right|^2} \int\limits_{\mathbb{R}^d_{\text{out}}} f(t,x,w)w \cdot n\, dw \quad \forall v \in \mathbb{R}^d_{\text{in}}, \tag{1.2.4b}$$

with $c > 0$ being a normalization constant for the Maxwell distribution or more precisely for the flux of the Maxwellian distribution function (i.e. $c = \left(\int_{\mathbb{R}^d_{\text{in}}} e^{-\left|\frac{v-V_{\text{bnd}}}{T_{\text{bnd}}}\right|^2} |w \cdot n(x)|\, dw\right)^{-1}$). The normalization guarantees that the total incoming and outgoing flux are the same.

$$\int_{\mathbb{R}^d_{\text{in}}} f(t,x,w)|w \cdot n(x)|\, dw = \int_{\mathbb{R}^d_{\text{out}}} f(t,x,w)|w \cdot n(x)|\, dw.$$

This is the only relation between in and outgoing values of $f$ in this case. The behaviour of particles hitting the wall is affected by the temperature and velocity of the wall and also the total outgoing flux. Note that $V_{\text{bnd}}$ is tangential to $\partial\Omega$ at the point $x$.

**Remark 1.2.1.** *As stated in [CIP94], a more general way to describe interaction of the paricles with rigid obstacles is given by the following condition.*

$$|v' \cdot n| f(t,x,v') = \int\limits_{\mathbb{R}^2_{out}} R(v \mapsto v')|v \cdot n| f(t,x,v)\, dv \quad \forall v' \in \mathbb{R}^2_{in}.$$

*In the above equation $n$ is the outer unit normal vector at $\partial\Omega$. $R(v \mapsto v')$ is the probability density that a molecule hitting the boundary with velocity $v$ is re-emitted from $\partial\Omega$ with velocity $v'$. By its meaning, $R$ is demanded to satisfy positivity for valid pairs $(v, v')$. Moreover, the wall shall not produce or capture particles, resulting in a normalization requirement for $R$:*

$$\int\limits_{\mathbb{R}^2_{in}} R(v \mapsto v')\, dv = 1 \quad \forall v' \in \mathbb{R}^2_{out}.$$

*The choice of $R(v \mapsto v') = \delta(v' - v + 2nn \cdot v)$ yields the specular reflection condition. Choosing $R$ equal to an appropriate Maxwell distribution, one obtains a diffuse reflecting wall.*

- inflow boundary condition

$$f(t,x,v) = f_{\text{in}}(t,x,v) \quad \forall v \in \mathbb{R}^2_{\text{in}}, \tag{1.2.4c}$$

with $f_{\text{in}}$ being some given, non-negative distribution function at the boundary. In this case, there is no relation between incoming and outgoing values of $f$.

## 1.3 Numerical Challenges

The Boltzmann equation provides a heavy task when being solved numerically. It is defined in a seven dimensional space, the space is unbounded in three of these directions. The interaction of particles, modelled as the Boltzmann collision operator is non linear in the solution function, usually providing a quadratic form on the discrete level with a quite complicated matrix.

Present numerical methods can in general be split in deterministic and probabilistic approaches. Despite of their simplicity, probabilistic methods, such as the Monte Carlo method yield highly accurate results only if a huge number of particles is simulated, making them unsuitable if high accuracy is desired. Moreover, these methods have to deal with stochastic fluctuations.
Many of the deterministic approaches are based on Fourier techniques, such as truncated Fourier series expansion and Fourier transformation. Both, strong as well as weak formulations are present in literature. In contrast to probabilistic methods, these methods are typically accurate as spectral methods. On the other hand, using Fourier techniques binds oneself to deal with domain truncation. Due to that, additional attention has to be paid to the conservation properties.

The method presented in section 4 is deterministic. It is based on a Discontinuous Galerkin projection in position and momentum space. We use global basis functions w.r.t. the momentum, enabling us to treat the integration over the unbounded domains without truncation. Conservation properties of the Boltzmann Collision operator suggest that a Petrov-Galerkin method allowing for global polynomials as test functions in momentum direction, naturally yields the conservation properties of the Boltzmann equation on the discrete level. To have good approximation properties, we let the basis functions within each cell depend on the actual solution.
The concept of global basis functions in the velocity domain, coupled with a Galerkin projection is not new in the context of kinetic transport models. For the semiconductor Boltzmann equation an expansion to Hermite polynomials was also used in [RSZ01].
A similar discretization for the transport operator is presented in [DDCS12, HGMM12].

Another deterministic approach in solving the Boltzmann equation is in terms of discrete velocity methods. In such methods, the particles attain only velocities from a discrete set of values, particle interactions have to be in such a way, that the interaction outcome belongs again to the above mentioned set of discrete values. Also in such methods the conservation properties on the discrete level have to be investigated carefully of course, but for such models exist well known conditions for the parameters of the method, i.e. the ”collision coefficients” such that the conservation properties and the H-theorem hold on the discrete level.

Finally we want to mention methods which are based on moment equations. A popular set of these equations are Grad’s 13 moment equations, also known as the R13 equations [Gra49,Gra58,ST03]. These are evolution equations for the moments of the distribution function. The $i$-th resulting equation incorporates the $i$-th and $(i+1)$-st moment. Thus, a finite set of equations is not complete since it involves more moments than equations. This problem is solved by applying so called constitutive equations which state relations between higher and lower order moments. This relations can result from first principles (e.g. Fourier’s law, providing a relation between the second and third moment). A numerical treatment of these equations is found in [RTS13].

Summarizing, when dealing with the Boltzmann equation from a numerical point of view, the following issues have to be addressed:

- **High dimensionality.** As has already been seen, the Boltzmann equation is defined on a 7 dimensional, unbounded space. In other words, each spatial position $x \in \mathbb{R}^d$ is equipped with a distribution function. We work on that space by introducing a tensor product basis.

- **Integrals over unbounded domains** have to be calculated. The pairing of our trial and test functions enables us to calculate these integrals by Gauss Hermite quadrature rules.

- Evaluating the **Boltzmann Collision operator** provides a huge quantity of numerical work. Moreover, it forms the basis of the macroscopic behaviour and thus its evaluation has to be done with care.

- **Spatial transport** needs a stable discretization. We are going to work on that issue by using a Discontinuous Galerkin method with upwind fluxes.

## 1.4 Outline

The outline of the underlying thesis is organized as follows:

- In **section 2** some basic properties of solutions and of the collision term are collected. We introduce the concept of collision invariant functions, state the kernel of the collision operator (the Maxwell distributions) and present the H-theorem. The second part of this section is devoted to the connection of the Boltzmann equation with the Euler and Navier- Stokes equations.

- **Section 3** provides a general view on existing numerical methods and presents various Monte Carlo methods as well as certain deterministic approaches.

- **Section 4** is the main part of the underlying work. Here the method developed as part of the thesis of the author is presented. We formulate the basic method and show a technique to calculate the collision integrals efficiently. This is based on various basis transformations in the trial and test space, orthogonality relations shared by the different bases and their tensor product structure.

- Finally, in **section 5** we present numerical examples as a validation for the method.

# 2 Elementary properties of solutions

In the current section we collect some basic properties of solutions of the Boltzmann equation and of the Boltzmann collision operator. We begin with the definition of the moments and of the macroscopic properties of the flow. In the main part of this section we focus on the Boltzmann collision operator with the presentation based on the textbooks of Cercignani [Cer90, CIP94]; some of the results are also presented in [RW05]. As a preparation, we show well known representations for the variational form of the collision integrals and introduce the concept of collision invariants which leads to the conservation properties of the collision operator. As an application of the collision invariants and the above mentioned representations we show how to obtain the kernel of the collision operator and present the $H$-theorem. At the end of this section we show a system of partial differential equations satisfied by the moments of the distribution function. This system leads to Euler or Navier-Stokes equations.

## 2.1 Moments and Macroscopic Properties

In the following we define the moments and macroscopic properties of the distribution function.

**Definition 2.1.1.** *The moments of the distribution function are denoted by the quantities* $m^{(i)}$. *These are defined by*

$$m^{(i)}_{j_1\dots j_i} := \int\limits_{\mathbb{R}^d} v_{j_1}\dots v_{j_i} f(t,x,v)\,dv \qquad j_i \in \{1,\dots,d\}.$$

*The* $i-$*th moment is therefore an* $i-$*dimensional, symmetric tensor, i.e.:*

$$m^{(i)}_{j_1\dots j_i} = m^{(i)}_{\pi(j_1)\dots\pi(j_i)} \qquad \forall \textit{ permutations } \pi$$

*Naturally, the* $0-th$ *order moment is interpreted as a scalar* $m^{(0)}$*, the first moment as a vector* $m^{(1)} = (m^{(1)}_1, \dots, m^{(1)}_d)^T$*, and the second order moment as a matrix*

$$m^{(2)}_{(3d)} = \begin{pmatrix} m^{(2)}_{11} & m^{(2)}_{12} & m^{(2)}_{13} \\ m^{(2)}_{21} & m^{(2)}_{22} & m^{(2)}_{23} \\ m^{(2)}_{31} & m^{(2)}_{32} & m^{(2)}_{33} \end{pmatrix} \quad \textit{resp.} \quad m^{(2)}_{(2d)} = \begin{pmatrix} m^{(2)}_{11} & m^{(2)}_{12} \\ m^{(2)}_{21} & m^{(2)}_{22} \end{pmatrix}.$$

The moments of the distribution function are closely connected to the macroscopically perceivable properties of the flow.

**Definition 2.1.2.** *The macroscopic behaviour of the gas is expressed by the following quantities, which are referred to as the macroscopic properties of the gas. Note that these quantities can be written in terms of the moments* $m^{(i)}_{j_1 \dots j_i}$*,* $i \leq 3$*, see section 2.4.*

$$\begin{aligned} \rho(t,x) &:= \int\limits_{\mathbb{R}^d} f(t,x,v)\,dv && \textit{mass density}, \\ V(t,x) &:= \frac{1}{\rho(t,x)} \int\limits_{\mathbb{R}^d} v f(t,x,v)\,dv && \textit{mean velocity}, \\ P_{ij}(t,x) &:= \int\limits_{\mathbb{R}^d} (v_i - V_i)(v_j - V_j) f(t,x,v)\,dv && \textit{stress tensor}, \\ p(t,x) &:= \frac{1}{d} \sum_{i=1}^{d} P_{ii} && \textit{pressure}, \\ T(t,x) &:= \frac{p(t,x)}{\rho(t,x)} && \textit{temperature}, \\ q_i(t,x) &:= \frac{1}{2} \int\limits_{\mathbb{R}^d} (v_i - V_i)|v - V|^2 f(t,x,v)\,dv && \textit{heat flux}. \end{aligned} \tag{2.1.1}$$

The definition of the mass density follows from the definition of the distribution function. The mean velocity $V$ is what one perceives macroscopically from the random motion of particles and describes the transport of mass.

The term stress tensor is justified by the fact that $P_{i,j}$ plays the same role in the macroscopic equations obtained from the Boltzmann equation, as the stress tensor does in conservation equations

derived from macroscopic considerations.
Another justification is stated in [Cer90] and is due to a comparison of the quantities $\int_{\mathbb{R}^d} v_i v_j f\,dv$ and $V_i V_j \rho$ which are in relation via

$$\begin{aligned}\int\limits_{\mathbb{R}^d} v_i v_j f\,dv &= \int\limits_{\mathbb{R}^d} (v_i - V_i + V_i)(v_j - V_j + V_j) f\,dv \\ &= V_i V_j \rho + \int\limits_{\mathbb{R}^d} (v_i - V_i)(v_j - V_j) f\,dv.\end{aligned}$$

The temperature is defined using the equation of state for an ideal gas. To be precise, we have not defined the physical temperature, but have defined the temperature multiplied with the Boltzmann constant divided by the particles mass.

As in the case of the stress tensor, the term heat flux is justified by its role in the macroscopic description. Another justification for the term heat flux – similar as for the stress tensor – can be found in [Cer90].

Finally we give the expression for the energy density $E$:

$$E := \frac{1}{2}\int\limits_{\mathbb{R}^d} |v|^2 f\,dv = \frac{1}{2}\rho|V|^2 + \frac{1}{2}\int\limits_{\mathbb{R}^d} |v - V|^2 f\,dv.$$

Thus, according to the above equation, the energy density consists of the macroscopic kinetic energy $\rho|V|^2$ and an additional term which will be referred to as the internal energy of the gas. Even if the macroscopic kinetic energy vanishes, the microscopic description still yields a non-vanishing kinetic energy. This remainder part arises due to the arbitrary motion of the particles around the mean velocity.

We conclude with an example:

**Example 2.1.3.** *The macroscopic properties of a Gaussian peak* $f(v) = \frac{\rho}{(2\pi T)^{\frac{d}{2}}} e^{-|\frac{v-V}{\sqrt{2T}}|^2}$ *are* *given by:*

$$\begin{aligned} \int_{\mathbb{R}^d} f(v)\,dv &= \rho \\ \int_{\mathbb{R}^d} v f(v)\,dv &= V \\ \int_{\mathbb{R}^d} (v_i - V_i)(v_j - V_j) f(v)\,dv &= \delta_{i,j}\rho T \end{aligned} \tag{2.1.2}$$

*Thus, the constants denoted by* $\rho$*,* $V$ *and* $T$ *are in accordance with the macroscopic properties. In the context of the Boltzmann equation we denote the above defined* $f$ *as Maxwellians.*

## 2.2 Properties of the Collision operator

In the current subsection we are dealing with the collision operator which is a local operator in time $t$ and position $x$, but acts globally on the velocity variable $v$. Thus, in order to study the properties of $Q$ it is sufficient to consider a simplified equation.

**Definition 2.2.1.** *Let* $f = f(t,v) \geq 0$*. We denote the following initial value problem as the spatially homogeneous Boltzmann equation:*

$$\frac{\partial}{\partial t} f = Q(f) \qquad f(0,v) = f_0(v). \tag{2.2.1}$$

**Remark 2.2.2.** *The initial value problem for the homogeneous equation (2.2.1) is quite well studied. It is well posed in the following sense: For initial data* $f_0$ *such that for* $s \in \{2,4\}$ *there holds* $\int_{\mathbb{R}^d} (1+|v|^2)^{\frac{s}{2}} f_0(v)\,dv < \infty$*, the problem has a unique solution* $f \in C^1([0,T];L^1)$ *with* $L^1$*-valued time derivative. Moreover,* $f$ *satisfies* $\int_{\mathbb{R}^d}(1+|v|)^2 f(t,v)\,dv < \infty$*. Finally, the solution is Lipschitz continuous w.r.t the initial data. However, a presentation of the quite long proof of these statements is out of the focus of this thesis.*

To study the properties of $Q$, the following theorem which provides different representations for $\int Q(f)\phi(v)\,dv$ is very useful. Most of the subsequent results are based on it.

**Theorem 2.2.3.** *Let $v'$ and $w'$ be defined via (1.2.3). Then for any functions $\phi, f$ for which the integrals on both sides and $\int B(v,w,e')f(v')f(w')\phi(v)\,de'dwdv$ exist, there holds*

$$\begin{aligned}\int\limits_{\mathbb{R}^d} Q(f)\phi(v)\,dv &= \int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}^d}\int\limits_{S^{d-1}} B(v,w,e')f(v)f(w)[\phi(v')-\phi(v)]de'dwdv\\ &= \frac{1}{2}\int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}^d}\int\limits_{S^{d-1}} B(v,w,e')f(v)f(w)\times\\ &\qquad [\phi(v')+\phi(w')-\phi(v)-\phi(w)]de'dwdv\\ &= \frac{1}{4}\int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}^d}\int\limits_{S^{d-1}} B(v,w,e')[f(v)f(w)-f(v')f(w')]\times\\ &\qquad [\phi(v')+\phi(w')-\phi(v)-\phi(w)]de'dwdv.\end{aligned}$$

*Proof.* The proof of the theorem is done in [Cer90, RW05] for example. The proof in [RW05] is a little more general: It is shown there, that for any suitable function $\varphi(v,w,e') = \varphi(|v-w|,(v-w)\cdot e', v, w, v', w')$ there holds

$$\begin{aligned}&\int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}^d}\int\limits_{S^{d-1}} \varphi(|v-w|,(v-w)\cdot e', v, w, v', w')\,de'dwdv\\ &\qquad = \int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}^d}\int\limits_{S^{d-1}} \varphi(|v-w|,(v-w)\cdot e', v', w', v, w)\,de'dwdv.\end{aligned} \tag{2.2.2}$$

This can be obtained by transforming to mean and relative velocity. This yields $v = \bar{v} + \frac{1}{2}\hat{v}$ and $v' = \bar{v}+e'|\hat{v}|$, for $w$ and $w'$ the only difference is a minus instead of a plus. Then one transforms the relative velocity $\hat{v} = |\hat{v}|e$ to polar coordinates to obtain $v = \bar{v}+\frac{1}{2}|\hat{v}|e$ and $v' = \bar{v}+\frac{1}{2}|\hat{v}|e'$. Now one combines $|\hat{v}|e'$ into a new variable $\tilde{v}$. At that point, we have actually exchanged the pre and post collision velocities. To conclude, one transforms the result back to $v$ and $w$ as integration variables. Note, that the transform $(v,w) \mapsto (\bar{v},\hat{v})$ and also its inverse have Jacobian determinant equal to 1. Finally one notes that under the above transformations $|v-w|$ and $\frac{\hat{v}\cdot e'}{|\hat{v}|}$ remain unchanged.

The first assertion of the theorem then follows by letting $\varphi(v,w,e') = B(v,w,e')f(v')f(w')\phi(v)$. The second can be obtained by simply interchanging $v$ and $w$ in $\int Q(f)\phi(v)\,dv$ and using

$B(w, v, -e') = B(v, w, e')$. The third representation results again by applying (2.2.2) to the second assertion of the theorem. ☐

**Remark 2.2.4.** *To study the properties of the collision operator it is often written as a bilinear mapping*

$$Q(f,g) = \frac{1}{2} \int_{\mathbb{R}^d} \int_{S^{d-1}} B(v,w,e') \left( f(v')g(w') + f(w')g(v') - f(v)g(w) - f(w)g(v) \right).$$

*Also for the above bilinear form, properties similar to those stated in theorem 2.2.3 can be shown, the proof follows essentially the above lines. The bilinear notation, or more precisely the collision contributions on $f$ exerted by $g$ and vice versa are used in mixtures of gases.*

The third representation of $\int Q(f)\phi$ in theorem 2.2.3 is of particular interest. We directly obtain that $\int Q(f)\phi$ vanishes independent of the particular function $f$, as soon as $\phi$ satisfies $\phi(v) + \phi(w) - \phi(v') - \phi(w') \equiv 0$. This leads us to the definition of the collision invariants.

### 2.2.1 Collision Invariants

**Definition 2.2.5.** *Let $v'$ and $w'$ be defined according to (1.2.3). A continuous function $\phi$ is called a collision invariant if it satisfies*

$$\phi(v) + \phi(w) = \phi(v') + \phi(w') \qquad v, w \in \mathbb{R}^d. \tag{2.2.3}$$

Using the definition of a collision invariant one can easily verify that $\Phi_0(v) \equiv 1$, $\Phi_j(v) = v_j,\ j = 1 \ldots d$ and $\Phi_{d+1}(v) = |v|^2$ are collision invariants which will be referred to as elementary collision invariants. On the other hand, it can be shown that all continuous collision invariants are linear combinations of the elementary collision invariants $\Phi_j,\ j = 0 \ldots d+1$. The requirements can be even further relaxed as shown in [CIP94]. Already Boltzmann investigated the solutions of (2.2.3). He already found its solutions under the additional assumption that they are in $C^2(\mathbb{R}^d, \mathbb{R})$.

**Theorem 2.2.6.** *Let $\Phi \in C(\mathbb{R}^d, \mathbb{R})$. $\Phi$ is a collision invariant if and only if it is a linear combination of the $d+2$ elementary collision invariants $\Phi_j,\ j = 0 \ldots d+1$.*

The proof under the above requirements is done in [CIP94]. Therein it is additionally shown that the continuity requirement can be further relaxed to measurable functions $\Phi$ which satisfy 2.2.3 almost everywhere and are finite almost everywhere. A purely physical argumentation of the proof is stated by Cercignani in [Cer90]. With higher smoothness assumptions on $\Phi$, i.e. $\Phi \in C^2$, the proof is worked out in [Bre].

Now we want to introduce a physical meaning for the $d+2$ elementary collision invariants $\Phi_j,\ j = 0 \dots d+1$. By definition of the collision invariants and by the use of theorem 2.2.3 one finds that

$$\int\limits_{\mathbb{R}^d} Q(f)\Phi(v)\,dv = 0 \tag{2.2.4}$$

holds for each collision invariant $\Phi$, thus especially for $\Phi_j, j = 0 \dots d+1$. This relation forms the basis of the macroscopic behaviour of the gas. Let $f$ be a solution of (2.2.1). By testing with the elementary collision invariant $\Phi_0 \equiv 1$ one directly obtains

$$\frac{\partial}{\partial t}\rho(t) = \frac{\partial}{\partial t}\int f\,dv = \int \frac{\partial}{\partial t} f\,dv = \int Q(f)\,dv \equiv 0.$$

This expresses the conservation of the total mass. In a similar way, conservation of momentum and energy are obtained by testing with the remaining elementary collision invariants. Thus, for a solution of the homogeneous Boltzmann equation the quantities

$$\begin{aligned} \rho(t) &= \rho(0) \\ V(t) &= V(0) \\ E(t) &= E(0) \end{aligned} \tag{2.2.5}$$

are conserved over time.

As the above considerations show, the conservation properties of the collision operator are the basis of the macroscopic behaviour of the gas. Therefore, a numerical approach should address these properties. At that point there are even multiple sources for inappropriate macroscopic behaviour. On the one hand, the elementary collision invariants should be collision invariants on the discrete level also. On the other hand, by discretization no additional collision invariants shall occur. Otherwise, a moment of the distribution function which is physically not conserved would be conserved on the discrete level.

Next, the kernel of the collision operator and thus stationary solutions of the homogeneous problem shall be obtained. Again, a suitable representation for the collision operator in theorem 2.2.3 is chosen.

### 2.2.2 Solutions of $Q(f) \equiv 0$

Since the proof for the following theorem also gives a hint on how to arrive at Boltzmanns well known $H$-theorem we decided to present it here.

**Theorem 2.2.7.** *A strictly positive, continuous and integrable function* $f : \mathbb{R}^d \to \mathbb{R}$ *satisfies* $Q(f) = 0$ *if and only if there exists a density* $\rho$*, a velocity* $V$ *and a temperature* $T$ *such that*

$$f(v) = \frac{\rho}{(2\pi T)^{d/2}} e^{-\left|\frac{v-V}{\sqrt{2T}}\right|^2}.$$

*Proof.* Consider a positive density function $f$ satisfying $Q(f)(v) \equiv 0$. Since $f > 0$, $\log(f)$ can be used as a test function. The third part of theorem 2.2.3 gives

$$\begin{aligned} &\int\limits_{\mathbb{R}^d} Q(f)(v) \log(f)(v)\, dv \\ =&\frac{1}{4} \int\limits_{\mathbb{R}^d} \int\limits_{\mathbb{R}^d} \int\limits_{S^{d-1}} [f(v)f(w) - f(v')f(w')] \times \\ &\qquad\qquad [\log(f(v')) + \log(f(w')) - \log(f(v)) - \log(f(w))] de'\, dw\, dv \\ =&\frac{1}{4} \int\limits_{\mathbb{R}^d} \int\limits_{\mathbb{R}^d} \int\limits_{S^{d-1}} f(v)f(w) \left(1 - \frac{f(v')f(w')}{f(v)f(w)}\right) \log\left(\frac{f(v')f(w')}{f(v)f(w)}\right) de' dw dv = 0. \end{aligned}$$

Since $(1-z)\log(z) < 0$ holds for each positive $z \in \mathbb{R} \setminus \{1\}$, $(1-z)\log(z) = 0 \Leftrightarrow z = 1$ we obtain that the integrand is non positive. By the continuity of $f$ we conclude that the integrand has to vanish completely in order to obtain a vanishing integral. This means that the $\frac{f(v')f(w')}{f(v)f(w)} \equiv 1$ has to hold, leading to

$$\log\left(\frac{f(v')f(w')}{f(v)f(w)}\right) = 0 \Leftrightarrow \log(f(v')) + \log(f(w')) = \log(f(v)) + \log(f(w)).$$

The last equation states that $\log(f)$ is a collision invariant. Thus, there exists $a, c \in \mathbb{R}$, $b \in \mathbb{R}^d$ with

$$\log(f(v)) = a + b \cdot v + c|v|^2 \Leftrightarrow f(v) = e^{a+b\cdot v+c|v|^2}.$$

Note that due to integrability of $f$, the constant $c$ in the above equation has to be negative. Taking this into account we can rewrite $f$ as

$$f(v) = \frac{\rho}{(2\pi T)^{d/2}} e^{-\left|\frac{v-V}{\sqrt{2T}}\right|^2}.$$

□

The above mentioned Gaussian peaks forming the kernel of the collision operator are termed Maxwellians in context of the Boltzmann equation. James Clark Maxwell already studied the problem of thermal equilibrium of a gas. He was able to prove – even without knowing the Boltzmann equation – that the one particle distribution function is a Maxwellian if the gas is in thermal equilibrium, see [Cer90] for more information.

The proof of the above theorem shows that for any solution $f$ of the homogeneous Boltzmann equation $\int Q(f)\log(f)\,dv \leq 0$ is satisfied. This forms the basis of the $H$-theorem.

## 2.3 Boltzmann H-theorem

**Definition 2.3.1.** *We start by defining the functionals*

$$\begin{aligned}
\mathcal{H}(f)(t,x) &:= \int\limits_{\mathbb{R}^d} f(t,x,v)\log(f(t,x,v))dv \\
\mathcal{H}_i(f)(t,x) &:= \int\limits_{\mathbb{R}^d} v_i f(t,x,v)\log(f(t,x,v))dv \quad i = 1\ldots d \\
H(f)(t) &:= \int\limits_{\Omega} \mathcal{H}(f)(t,x)\,dx.
\end{aligned}$$

### 2.3.1 The Space Homogeneous case

Similar to the proof of theorem 2.2.7 one obtains the following statement.

**Lemma 2.3.2.** *For a solution $f$ of the space homogeneous Boltzmann equation, the $H$-functional satisfies*

$$\frac{\partial \mathcal{H}(f)}{\partial t}(t) \leq 0.$$

*The equal sign is achieved if and only if $f$ is a Maxwellian $f_{eq}(v) = \frac{\rho}{(2\pi T)^{d/2}} e^{-\left|\frac{v-V}{\sqrt{2T}}\right|^2}$. The functional then evaluates to*

$$\mathcal{H}(f_{eq}) = \rho\left(\log\left(\frac{\rho}{(\pi T)^{d/2}}\right) - \frac{d}{2}\right).$$

*Proof.* We have almost everything collected to proof the above theorem. The only thing that is left, is a suitable representation for the time derivative of $f\log(f)$. We obtain

$$\frac{\partial}{\partial t}\left(f\log(f)\right) = \frac{\partial f}{\partial t}\left(\log(f)+1\right) = Q(f)(\log(f)+1), \tag{2.3.1}$$

where the second equal sign is due to $f$ being a solution of the homogeneous Boltzmann equation. Now we integrate both sides of the above equation with respect to $v$ and interchange the order of differentiation and integration on the left hand side to arrive at

$$\frac{\partial \mathcal{H}(f)}{\partial t}(t,x) = \frac{\partial}{\partial t}\int\limits_{\mathbb{R}^d}\left(f\log(f)\right)dv = \int\limits_{\mathbb{R}^d} Q(f)(\log(f)+1)dv = \int\limits_{\mathbb{R}^d} Q(f)\log(f)dv \leq 0.$$

This shows the first part of Lemma 2.3.2. From the proof of 2.2.7 it is clear, that the equal sign in the above equation is obtained if and only if $f$ is a Maxwellian function. In this case, the value of the functional is given by

$$\begin{aligned}\int\limits_{\mathbb{R}^d} f\log(f)dv &= \frac{\rho}{(2\pi T)^{d/2}}\int\limits_{\mathbb{R}^d} e^{-\left|\frac{v-V}{\sqrt{2T}}\right|^2}\left(\log(\rho(2\pi T)^{-d/2}) - \left|\frac{v-V}{\sqrt{2T}}\right|^2\right)dv\\ &= \frac{\rho}{\pi^{d/2}}\int\limits_{\mathbb{R}^d} e^{-|v|^2}\left(\log(\rho(\pi T)^{-d/2}) - |v|^2\right)dv\\ &= \frac{\rho}{\pi^{d/2}}\left(\log(\rho(2\pi T)^{-d/2})\pi^{d/2} - \pi^{d/2}\frac{d}{2}\right) = \rho\left(\log\Big(\frac{\rho}{(2\pi T)^{d/2}}\Big) - \frac{d}{2}\right).\end{aligned}$$

This completes the proof. □

### 2.3.2 The Space Inhomogeneous case

A generalization of the $H$-theorem to space dependent problems is under certain boundary conditions also possible. To this end, just as before we assume that $f = f(t,x,v) > 0$ is a solution of the Boltzmann equation in our spatial domain $\Omega$. For simplicity we consider the boundary $\partial\Omega$ of the domain to be a specular reflecting wall.

The first observation is that – as for the time derivative – there holds

$$\frac{\partial}{\partial x_i}(v_i f\log(f)) = \left(\frac{\partial}{\partial x_i}v_i f\right)(1+\log(f)).$$

Now we multiply both sides of the Boltzmann equation with $(1+\log(f))$, use the relations for the derivatives w.r.t. $x_i$ and $t$; (2.3.2) and (2.3.1) respectively and integrate the result over the velocity space to obtain

$$\frac{\partial\mathcal{H}}{\partial t} + \mathrm{div}_x\begin{pmatrix}\mathcal{H}_1\\ \vdots\\ \mathcal{H}_d\end{pmatrix} = \int\limits_{\mathbb{R}^d} Q(f)\log(f)\,dv \le 0. \tag{2.3.2}$$

In the next step we integrate both sides with respect to the spatial coordinate over the domain $\Omega$ and apply the divergence theorem

$$\frac{\partial H}{\partial t} + \int\limits_{\partial\Omega}\begin{pmatrix}\mathcal{H}_1\\ \vdots\\ \mathcal{H}_d\end{pmatrix}\cdot n \le 0. \tag{2.3.3}$$

If the boundary condition prohibits inflow of $\mathcal{H}$, then $\int\limits_{\partial\Omega} \begin{pmatrix} \mathcal{H}_1 \\ \vdots \\ \mathcal{H}_d \end{pmatrix} \cdot n \geq 0$, since we may interpret the boundary integral as the outflow of $\mathcal{H}$. As we stated in the beginning of our considerations, the gas is enclosed in a box with a specular reflecting wall. For that case it is shown in [Cer90] that the boundary integral in (2.3.3) vanishes completely. Moreover, in the case of a diffuse reflecting wall, the boundary integral has at least a positive sign such that we end up with

$$\frac{\partial H}{\partial t} \leq 0. \tag{2.3.4}$$

Clearly, the equal sign is achieved if

$$\int\limits_{\partial\Omega} \begin{pmatrix} \mathcal{H}_1 \\ \vdots \\ \mathcal{H}_d \end{pmatrix} \cdot n = 0 \quad \text{and} \quad \int\limits_{\Omega\times\mathbb{R}^d} Q(f)\log(f) = 0. \tag{2.3.5}$$

In accordance with the properties of the collision operator, $\int_{\mathbb{R}^d\times\mathbb{R}^d} Q(f)\log(f)$ vanishes in the case of a point wise Maxwellian distribution. Otherwise it is non-positive such that we may interpret the collision process as a negative source for the quantity $H$. Finally, we arrived at

**Lemma 2.3.3.** *Let the gas of interest be enclosed in a box $\Omega$ with either perfectly smooth or diffuse reflecting walls. Define $\mathcal{H}, \mathcal{H}_i, H$ as in definition 2.3.1 and let $f = f(t, x, v)$ be a solution of the Boltzmann equation on $\Omega$. Then the $H$-functional satisfies*

$$\frac{\partial H}{\partial t} \leq 0. \tag{2.3.6}$$

*Proof.* The proof is already sketched in the above lines, a rigorous derivation is found in [Cer90]. □

The $H$-theorem brought a lot of criticism to Boltzmann. It can be interpreted as some sort of irreversibility, what was the discussion starter. We consider a set of $N$ particles $(x_i, v_i), i = 1 \dots N$. If we – at a certain time $t_0$ interchange the velocity of each particle to $-v_i$, then the system will go back to its initial state, just with negative velocities. If we consider the same situation w.r.t. to our statistical description, then let $f(t_0, x, v)$ be the density of the number density before we interchanged the velocities. The system with interchanged velocities shall be described by $\tilde{f}(t, x, v)$. We find that $f(t_0, x, v) = \tilde{f}(t_0, x, -v)$ holds. If we assume for the moment that – as in the Newton dynamics – the Boltzmann equation brings $\tilde{f}$ back to the initial state with mirrored velocities, i.e. $\tilde{f}(2t_0, x, v) = f(0, x, -v)$. Then we would have $H(\tilde{f}(2t_0) = H(f)(0)$. On the other hand since the $H$-functional is strictly decreasing we have $H(f)(0) > H(f)(t_0) = H(\tilde{f}(t_0) > H(\tilde{f}(2t_0)$, thus, $H(f)(0) > H(\tilde{f}(2t_0)$, and we have obtained the contradiction $H(f)(0) > H(\tilde{f}(2t_0) = H(f)(0)$. Thus, the Boltzmann equation will not bring $\tilde{f}$ back to the initial configuration. This irreversibility is a result from our manipulation of the gain term of the collision operator in the derivation of the Boltzmann equation. For a detailed discussion on the $H$-theorem we refer once more to [Cer90, CIP94].

## 2.4 Moment equations

In this section we consider a gas close to equilibrium. We use the properties of the collision invariants to derive a system of five partial differential equations with 13 unknowns (in 3 dimensions). To obtain a closed set of equations we have to postulate equations that relate the heat flux $q$ and the stress tensor $P$ to the lower order moments. These relations are called closure relations.

From the basic properties of the collision operator we obtain for a solution of the Boltzmann equation

$$\begin{aligned}
&\int\limits_{\mathbb{R}^d} \frac{\partial f}{\partial t} + \mathrm{div}_x(vf)\,dv = 0\\
&\int\limits_{\mathbb{R}^d} \frac{\partial f}{\partial t} v_j + \mathrm{div}_x(vf)v_j\,dv = 0 \quad j = 1\dots d\\
&\int\limits_{\mathbb{R}^d} \frac{\partial f}{\partial t}|v|^2 + \mathrm{div}_x(vf)|v|^2\,dv = 0,
\end{aligned} \tag{2.4.1}$$

by testing the Boltzmann equation with the elementary collision invariants. Interchanging the orders of differentiation and integration yields conservation equations for the moments

$$\begin{aligned}
&\frac{\partial m^{(0)}}{\partial t} + \mathrm{div}_x(m^{(1)}) = 0\\
&\frac{\partial m^{(1)}}{\partial t} + \mathrm{div}_x(m^{(2)}) = 0\\
&\frac{\partial \mathrm{tr}(m^{(2)})}{\partial t} + \mathrm{div}_x \begin{pmatrix} m^{(3)}_{111} + \dots + m^{(3)}_{1dd}\\ \vdots \\ m^{(3)}_{d11} + \dots + m^{(3)}_{ddd}\end{pmatrix} = 0.
\end{aligned} \tag{2.4.2}$$

Note that the divergence is applied row wise to the matrix $m^{(2)}$ in the second equation. By tr( . ) we denote the trace of a matrix, i.e. $\text{tr}(m^{(2)}) = \sum_i m^{(2)}_{ii}$.

The conservation equations for the moments form the basis of the ongoing calculations. Our goal is to express the moments and their divergence respectively in terms of the macroscopic properties. There holds

$$m^{(0)} = \int_{\mathbb{R}^d} f\,dv = \rho \qquad m^{(1)}_i = \int_{\mathbb{R}^d} vf\,dv = \rho V$$

$$m^{(2)}_{ij} = \int_{\mathbb{R}^d} v_i v_j f\,dv = P_{i,j} + V_i V_j \rho \qquad \text{tr}(m^{(2)}) = (dT + |V|^2)\rho.$$

For the calculation of $m^{(3)}_{i,j,j}$ we express the heat flux vector $q$ as

$$\begin{aligned} 2q_i &= \int_{\mathbb{R}^d} (v_i - V_i)|v - V|^2 f\,dv = \sum_{j=1}^{d} \int_{\mathbb{R}^d} (v_i - V_i)(v_j - V_j)^2\,dv \\ &= \sum_{j=1}^{d} \int_{\mathbb{R}^d} v_i v_j^2 f - V_i v_j^2 f - 2V_j v_i v_j f + 2V_i V_j v_j f + V_j^2 v_i f - V_i V_j^2 f\,dv \\ &= \sum_{j=1}^{d} m^{(3)}_{ijj} - V_i m^{(2)}_{jj} - 2V_j m^{(2)}_{ij} + 2V_i V_j^2 \rho \\ &= \sum_{j=1}^{d} m^{(3)}_{ijj} - V_i \text{tr}(m^{(2)}) - 2V_j m^{(2)}_{ij} + 2V_i |V|^2 \rho. \end{aligned}$$

Finally, writing $m^{(2)}_{ij} = P_{ij} + \rho V_i V_j$ in terms of the macroscopic properties and expressing $\sum_j m^{(3)}_{ijj}$ from the above equation, one finds

$$\sum_{j=1}^{d} m^{(3)}_{ijj} = 2q_i + \rho V_i (dT + |V|^2) + 2\sum_{j=1}^{d} V_j P_{ij}.$$

Now rewriting the moment conservation (2.4.2) in terms of the macroscopic quantities results in

$$\begin{aligned}
&\frac{\partial \rho}{\partial t} + \operatorname{div}(\rho V) &&= 0\\
&\frac{\partial (\rho V)}{\partial t} + \operatorname{div}(P + \rho V \otimes V) &&= 0\\
&\frac{\partial (dT\rho + |V|^2 \rho)}{\partial t} + \operatorname{div}(2q + \rho(dT + |V|^2)V + 2PV) &&= 0.
\end{aligned} \tag{2.4.3}$$

The system 2.4.3 is the already mentioned unclosed system of equations. In 3 dimensions there are five equations for 13 unknowns: 1 for $\rho$, 3 for both, $V$ and $q$, and finally the pressure tensor $P$ requires 6 unknowns due to it's symmetry. In 2 dimensions one obtains 4 equations for 8 unknowns (1 + 2 + 2 + 3). Thus, in fact we need to solve the Boltzmann equation to arrive at the missing information. If we have solved it, we can already extract all the quantities in the above equation from the distribution function itself.

In order to obtain a useful set of equations, so called constitutive or closure relations have to be postulated. These relations express additional (experimentally obtained) relations between macroscopic properties [Cer90].

### 2.4.1 From Boltzmann to Euler

The constitutive equations for the Euler equations can be obtained by assuming that the gas has a Maxwellian distribution function:

$$f(t,x,v) = \frac{\rho(t,x)}{\sqrt{2\pi T(t,x)}^{d}} e^{-\frac{|v-V(t,x)|^2}{2T(t,x)}}.$$

The heat flux and the pressure tensor can be evaluated to

$$q \equiv 0 \quad \text{and} \quad P = p(t,x)I = TI\rho.$$

Thus, (2.4.3) turns into

$$\begin{aligned}
&\frac{\partial \rho}{\partial t} + \operatorname{div}(\rho V) &&= 0\\
&\frac{\partial(\rho V)}{\partial t} + \operatorname{div}(\rho(TI + V \otimes V)) &&= 0\\
&\frac{\partial(\frac{d}{2}T\rho + \frac{1}{2}|V|^2\rho)}{\partial t} + \operatorname{div}(\rho V(\tfrac{d+2}{2}T + \tfrac{1}{2}|V|^2)) &&= 0.
\end{aligned} \tag{2.4.4}$$

The above equations are called Euler equations, describing a so called Euler fluid.

### 2.4.2 From Boltzmann to Navier-Stokes

When deriving the Euler equations from the Boltzmann equation, the constitutive equations arise naturally by the ansatz of a Maxwellian distribution function. For the Navier-Stokes equations we assume that Fourier's law holds

$$q = -\kappa \nabla T. \tag{2.4.5}$$

Additionally we assume that (i) $P$ depends linearly on the deformation, (ii) $P$ is isotropic and (iii) $P_{ij} = -p\delta_{ij}$ if no deformation occurs. This yields the following expression for $P$

$$P = p(t,x)I - 2\mu\epsilon(V) - \lambda \operatorname{div}(V)I, \tag{2.4.6}$$

with the symmetric gradient $\epsilon(V)_{i,j} := \frac{1}{2}\left(\frac{\partial V_j}{\partial x_i} + \frac{\partial V_i}{\partial x_j}\right)$. The parameter $\mu$ is the dynamic viscosity, $\lambda$ is called volume viscosity.

**Remark 2.4.1.** *The viscous stress tensor as well as Fourier's law can be systematically derived from the Boltzmann equation in context of the Chapman-Enskog expansion. This would be the proper way to derive the Navier-Stokes equations from the Boltzmann equation. However, this derivation is out of the focus of the thesis. We refer to [CIP94] and the references therein for more information.*

As for the Euler equations, the continuity equation reads

$$\frac{\partial \rho}{\partial t} + \operatorname{div}(\rho V) = 0. \tag{2.4.7}$$

For the momentum balance equation the divergence of $P$ is required. If we take into account that the divergence of $pI$ and $\mathrm{div}(V)I$ are given by the gradients of the quantities $p$ and $\mathrm{div}(V)$ respectively, then the divergence of $P$ evaluates to

$$\mathrm{div}(P) = \nabla p - \mu \mathrm{div}(\epsilon(V)) - \lambda \nabla(\mathrm{div}(V)). \tag{2.4.8}$$

Note that the divergence is applied row-wise to the matrix $\epsilon(V)$. The $i$-th component of the divergence of $\rho(V \otimes V)$ results in

$$\begin{aligned} \mathrm{div}(\rho V \otimes V)_i &= \sum_{j=1}^{d} \frac{\partial \rho}{\partial x_j} V_i V_j + \rho \left( \frac{\partial V_i}{\partial x_j} V_j + \frac{\partial V_j}{\partial x_j} V_i \right) \\ &= V_i \left( \sum_{j=1}^{d} \frac{\partial \rho}{\partial x_j} V_j + \rho \frac{\partial V_j}{\partial x_j} \right) + \rho \sum_{j=1}^{d} \frac{\partial V_i}{\partial x_j} V_j \\ &= V_i \, \mathrm{div}(\rho V) + \rho \nabla V_i \cdot V. \end{aligned}$$

If we use the continuity equation $\frac{\partial \rho}{\partial t} = -\mathrm{div}(\rho V)$, we can rewrite the divergence of $\rho(V \otimes V)$ as

$$\mathrm{div}(\rho(V \otimes V)) = -\frac{\partial \rho}{\partial t} V + \rho \nabla V \, V. \tag{2.4.9}$$

In the above equation $\nabla V$ denotes the Jacobian of $V$ w.r.t. the spatial variables.
Finally we consider the time derivative of $\rho V$ resulting in

$$\frac{\partial(\rho V)}{\partial t} = V \frac{\partial \rho}{\partial t} + \rho \frac{\partial V}{\partial t}. \tag{2.4.10}$$

Plugging (2.4.8), (2.4.9) and (2.4.10) into the second line of (2.4.3) one obtains the Navier-Stokes momentum balance

$$\rho \left( \frac{\partial V}{\partial t} + \nabla V \, V \right) - 2\mu \mathrm{div}(\epsilon(V)) - \lambda \nabla(\mathrm{div}(V)) = -\nabla p. \tag{2.4.11}$$

Next we consider the energy balance. The divergence of the heat flux results in

$$\mathrm{div}(2q) = -2\kappa \Delta T, \tag{2.4.12}$$

the divergence of $d\rho TV$ evaluates to

$$\text{div}(d\rho TV) = d\left(\rho V \cdot \nabla T + \text{div}(V\rho)T\right), \tag{2.4.13}$$

and the divergence of $|V|^2\rho V$ is

$$\text{div}(|V|^2\rho V) = |V|^2\text{div}(\rho V) + \rho V \cdot \nabla(|V|^2). \tag{2.4.14}$$

Now we calculate the divergence of $PV$. There holds $PV = pV - \lambda\text{div}(V)V - 2\mu\epsilon(V)V$, such that we can write

$$\text{div}(PV) = \nabla p \cdot V + p\text{div}(V) - \lambda\nabla\text{div}(V) \cdot V - \text{div}(V)^2 - 2\mu\text{div}(\epsilon(V)V). \tag{2.4.15}$$

The divergence of $\epsilon(V)V$ can be written as

$$\text{div}(\epsilon(V)V) = \text{tr}(\nabla V\epsilon(V)) + \text{div}(\epsilon(V)) \cdot V. \tag{2.4.16}$$

From the momentum balance equation one obtains

$$-2\mu\text{div}(\epsilon(V)) \cdot V = \left(-\nabla p + \lambda\nabla(\text{div}(V)) - \rho\left(\tfrac{\partial V}{\partial t} + \nabla V\, V\right)\right) \cdot V,$$

such that we can write div$(PV)$ as

$$\text{div}(PV) = p\text{div}(V) - 2\mu\text{tr}(\nabla V\,\epsilon(V)) - \lambda\text{div}(V)^2 - \rho\left(\tfrac{\partial V}{\partial t} + \nabla V\, V\right) \cdot V. \tag{2.4.17}$$

The time derivative of $dT\rho + |V|^2\rho$ is given by

$$\frac{\partial(dT\rho + |V|^2\rho)}{\partial t} = d\rho\frac{\partial T}{\partial t} + dT\frac{\partial \rho}{\partial t} + \rho\frac{\partial |V|^2}{\partial t} + |V|^2\frac{\partial \rho}{\partial t}. \tag{2.4.18}$$

Now we plug all the representations found above into the third moment equation, use the continuity equation and the identities $V \cdot \nabla(|V|^2) = 2\nabla VV \cdot V$ as well as $\frac{\partial|V|^2}{\partial t} = 2\frac{\partial V}{\partial t} \cdot V$ to arrive at

$$d\rho\frac{\partial T}{\partial t} + 2(p\text{div}(V) - 2\mu\text{tr}(\nabla V\,\epsilon(V)) - \lambda\text{div}(V)^2) + dV\rho\nabla T - 2\kappa\Delta T = 0.$$

Finally we rewrite $p\text{div}(V) = T\rho\text{div}(V) = T(\text{div}(\rho V) - \nabla\rho \cdot V) = T(-\frac{\partial \rho}{\partial t} - \nabla\rho \cdot V)$, introduce the definition of the internal energy $e = \frac{d}{2}T$ and the material derivative $\frac{D.}{Dt} := \frac{\partial .}{\partial t} + V \cdot \nabla(.)$ to end up with the common form of the energy balance equation for the Navier- Stokes system.

$$\rho\frac{De}{Dt} - T\frac{D\rho}{Dt} = 2\mu\text{tr}(\nabla V \epsilon(V))) - \text{div}(q) + \lambda\text{div}(V)^2. \tag{2.4.19}$$

The obtained system of equations now reads

$$\begin{aligned} \frac{D\rho}{Dt} &= -\rho\text{div}(V) \\ \rho\frac{DV}{Dt} &= 2\mu\text{div}(\epsilon(V)) - \lambda\nabla(\text{div}(V)) - \nabla p \\ \rho\frac{De}{Dt} - T\frac{D\rho}{Dt} &= 2\mu\text{tr}(\nabla V \epsilon(V))) - \text{div}(q) + \lambda\text{div}(V)^2. \end{aligned} \tag{2.4.20}$$

# 3 Numerical Methods

In the following section we give a general view of numerical methods for the Boltzmann equation. Even though some of them have been considered for the full Boltzmann equation, we only present the discretization in the velocity domain. Thus, we only treat space homogeneous problems in this section. Moreover we restrict the presentation of the methods to the dimension of the velocity space used in the presentation in the literature.
In the first subsection we deal with a description of stochastic methods, in the second subsection deterministic methods are presented.

## 3.1 Stochastic methods

Monte Carlo methods are widely used for the numerical treatment of the Boltzmann equation. These methods impress by their simplicity and efficiency, the proposed algorithms are of computational effort $N$, where $N$ is the number of simulated particles. The schemes are based on the simulation of a subset of the particles of interest, resulting in simple algorithms. On the other hand the methods have to deal with low accuracy and stochastic fluctuations. We present here the methods proposed by Nanbu and Babovsky [Nan80, Bab86]. We also refer the reader to the approach by Bird [Bir95] and to the textbook of Rjasanow and Wagner [RW05].

### 3.1.1 Monte Carlo methods – Nanbu - Babovsky scheme

We start with the homogeneous equation and apply the usual splitting into the gain and loss term to the collision operator.

$$\begin{aligned}\frac{\partial f}{\partial t} &= \frac{1}{\mathrm{Kn}}Q(f)\\ &= \frac{1}{\mathrm{Kn}}\left(\int\limits_{\mathbb{R}^3}\int\limits_{S^2} B(v,w,e')f(v')f(w')\,de'\,dw - f(v)\int\limits_{\mathbb{R}^3}\int\limits_{S^2} B(v,w,e')f(w)\,de'\,dw\right).\end{aligned} \tag{3.1.1}$$

We assume $f$ is a probability density, i.e. $f \geq 0$ and $\int f = 1$. For simplicity we consider Maxwellian molecules with $B(v, w, e') = \frac{1}{4\pi}$. In that case, the above equation simplifies to

$$\frac{\partial f}{\partial t} = \frac{1}{\mathrm{Kn}} \underbrace{\frac{1}{4\pi} \int_{\mathbb{R}^3} \int_{S^2} f(v')f(w')\, de'\, dw}_{:=Q^+(f)} - \frac{1}{\mathrm{Kn}} f(v). \tag{3.1.2}$$

If one denotes by $f^{(n)}$ the solution at time $t = n\Delta t$ and applies a forward Euler scheme with time step $\Delta t$ to (3.1.2) one obtains

$$\begin{aligned} \frac{f^{(n+1)} - f^{(n)}}{\Delta t} &= \frac{1}{\mathrm{Kn}} Q^+(f^{(n)}) - \frac{1}{\mathrm{Kn}} f^{(n)} \Leftrightarrow \\ f^{(n+1)} &= (1 - \frac{\Delta t}{\mathrm{Kn}}) f^{(n)} + \frac{\Delta t}{\mathrm{Kn}} Q^+(f^{(n)}). \end{aligned} \tag{3.1.3}$$

The method is based on a probabilistic interpretation of (3.1.3). To this end we note that $Q^+(f) \geq 0$ and that $\int Q^+(f) = 1$, thus $Q^+(f)$ is a probability density. Now, a particle sampled from $f^{(n+1)}$ is sampled from $Q^+(f^{(n)})$ with probability $\frac{\Delta t}{\mathrm{Kn}}$ and from $f^{(n)}$ with probability $1 - \frac{\Delta t}{\mathrm{Kn}}$. Roughly speaking, the particle does not collide with probability $1 - \frac{\Delta t}{\mathrm{Kn}}$ in the time interval $[n\Delta t, (n+1)\Delta t]$. From this interpretation one derives Algorithm 1. We denote by $N$ the number of simulated particles and by $n_t$ the number of time steps.

```
for $i = 1 : N$ do
    Set $v_i^{(0)}$ by sampling it from the initial distribution $f_0(v)$.
end
for $t = 1 : n_t$ do
    for $i = 1 : N$ do
        Sample a uniform distributed random number $\xi \in [0, 1]$
        if $\xi < 1 - \frac{\Delta t}{\mathrm{Kn}}$ then
            Set $v_i^{(t+1)} = v_i^{(t)}$
        else
            Sample a uniform distributed random integer $j \in [1, N]$
            Calculate $v_i'$ according to a collision with particle $j$, i.e.
            Sample a uniform distributed $e' \in S^2$
            Set $v_i^{(t+1)} = \frac{v_i^{(t)} + v_j^{(t)}}{2} + e' \frac{|v_i^{(t)} - v_j^{(t)}|}{2}$
        end
    end
end
```

**Algorithm 1:** A first version of the Monte Carlo method

**Remark 3.1.1.** *The method proposed by Nanbu [Nan80] is quite similar to the presented algorithm. Therein a stochastic law for the density function at time $\Delta t$ is derived under the assumption that the initial distribution is of the form $f(0,v) = \frac{1}{N}\sum_{i=1}^{N}\delta(v-v_i)$, based on a first order expansion of $f$ at $t=0$ w.r.t. time. The scheme is presented for a larger class of collision kernels, yielding individual collision probabilities for individual particles and their collision partners. In our setting with constant collision kernel $B(v,w,e') = \frac{1}{4\pi}$, the method can be formulated as in Algorithm 1.*

We note the strategy to sample from $Q^+(f^{(n)})$. In fact we perform a collision according to (1.2.3) and sample the scattering vector uniformly from the sphere $S^2$.
Having a closer look at the collision process one notes that only particle $i$ changes its velocity according to a two particle collision and thus momentum and energy are not conserved at a single collision. A simple modification leading to the desired conservation properties at each single collision can be obtained by the following considerations. The idea is to select independent collision pairs without repetition and change both velocities according to the interaction law. The expected number of collisions in a small time interval $\Delta t$ is given by $N\frac{\Delta t}{\text{Kn}}$. Thus, the number of expected collision pairs is $N\frac{\Delta t}{2\text{Kn}}$. With this considerations, Algorithm 1 can be reformulated as

```
for $i = 1 : N$ do
    Set $v_i^{(0)}$ by sampling it from the initial distribution $f_0(v)$.
end
for $t = 1 : n_t$ do
    Select $N_c := \frac{N\Delta t}{2\text{Kn}}$ independent pairs $(i,j)$ from all possible pairs
    for $(i,j)$ collision pair do
        Compute $v_i'$, $v_j'$ according to a collision of $v_i$ and $v_j$
        Set $v_i^{(t+1)} = v_i'$ and $v_j^{(t+1)} = v_j'$
    end
    Set $v_i^{(t+1)} = v_i^{(t)}$ for all particles which were not selected
end
```

**Algorithm 2:** A conservative version of the Monte Carlo scheme.

In Algorithm 2 both particles $i$ and $j$ interchange their state according to a binary collision and thus, momentum and energy are conserved at each collision in this version of the algorithm. The computation of the post collision velocities is done as is in Algorithm 1.

We note that these algorithms have the time step restriction $1 - \frac{\Delta t}{\text{Kn}} > 0 \Leftrightarrow \Delta t < \text{Kn}$ in order to guarantee a probabilistic interpretation, what is critical if the Knudsen number becomes small. This problem can be avoided with a suitable time discretization, yielding to time relaxed Monte Carlo methods.

### 3.1.2 Time relaxed Monte Carlo methods

The time relaxation approach is based on a suitable series expansion of the solution, a truncation of this series and a replacement of higher order terms by the equilibrium solution. A detailed description of such methods can be found in [PT05], the main aspects are presented in the sequel.

We consider the homogeneous Boltzmann equation for the probability density $f$ in the form

$$\frac{\partial f}{\partial t} = \frac{1}{\mathrm{Kn}} \left( Q^+(f,f) - f \right), \qquad f(0,v) \text{ given.} \tag{3.1.4}$$

Here we used the same collision kernel as before, $B(v,w,e') = \frac{1}{4\pi}$. The bilinear operator $Q^+$ is defined as $Q^+(f,g) = \frac{1}{8\pi} \int_{\mathbb{R}^3} \int_{S^2} f(v')g(w') + g(v')f(w')$. Now the function $f$ can be expanded to the following formal series, which is known as Wild's sum [Wil51].

$$f(t,v) = e^{-\frac{t}{\mathrm{Kn}}} \sum_{k=0}^{\infty} (1 - e^{-\frac{t}{\mathrm{Kn}}})^k f_k(v), \tag{3.1.5}$$

with $f_k$ given by

$$\begin{aligned} f_0(v) &= f(0,v) \\ f_{k+1}(v) &= \frac{1}{k+1} \sum_{u=0}^{k} Q^+(f_u, f_{k-u}), \quad k = 0, 1 \dots \end{aligned} \tag{3.1.6}$$

To construct a numerical method on the basis of the above expansion, the Wild sum is truncated at a fixed $k = m$ and the coefficient of $(1 - e^{-\frac{t}{\mathrm{Kn}}})^{m+1}$ is replaced by the equilibrium $M(v)$, where $M$ denotes the normalized Maxwellian with appropriate velocity and temperature. The numerical solution at time $\Delta t$ is defined via

$$f(\Delta t, v) = e^{-\frac{\Delta t}{\mathrm{Kn}}} \sum_{k=0}^{m} (1 - e^{-\frac{\Delta t}{\mathrm{Kn}}})^k f_k(v) + (1 - e^{-\frac{\Delta t}{\mathrm{Kn}}})^{m+1} M(v). \tag{3.1.7}$$

Truncating at $m = 1$ leads to

$$f(\Delta t, v) = A_0 f_0 + A_1 f_1 + A_2 M \qquad f_0 = f(0,v),\ f_1 = Q^+(f_0, f_0)(v). \tag{3.1.8}$$

The coefficients $A_j$ satisfy $A_j \geq 0$ as well as $A_0 + A_1 + A_2 = 1$. They are given by $A_j = \tau^j(1-\tau),\ j = 0, 1$ and $A_2 = \tau^2$, with $\tau = 1 - e^{-\frac{\Delta t}{\mathrm{Kn}}}$.
A probabilistic interpretation of (3.1.8) can be stated as follows: A particle sampled from $f$ at time $\Delta t$ is sampled from $f_1$ with probability $A_1$, it is sampled from $f_0$ with probability $A_0$. Finally it is sampled from a Maxwellian distribution with probability $A_2$. A numerical scheme based on this interpretation can be implemented as presented in Algorithm 3.

**Remark 3.1.2.** *By the above time discretization, the probabilistic interpretation of (3.1.8) holds uniform in $\frac{1}{\mathrm{Kn}}$, enabling larger time steps. Besides the stochastic approach presented in [PT05], the time relaxation approach can also be applied to deterministic methods, as is done in [GPT97], where the time relaxation is combined with a discrete velocity model.*

```
for $i = 1 : N$ do
    Set $v_i^{(0)}$ by sampling it from the initial distribution $f_0(v)$.
end
for $t = 1 : n_t$ do
    Compute $\tau = 1 - e^{-\frac{\Delta t}{\mathrm{Kn}}}$ and $A_i,\ i = 0, 1, 2$.
    Sample $N_c := \frac{NA_1}{2}$ pairs $(i, j)$ from all possible collision pairs
    Set $N_r := NA_2$
    for $(i, j)$ in collision pairs do
        Compute $v_i'$, $v_j'$ according to a collision of $v_i$ and $v_j$
        Set $v_i^{(t+1)} = v_i'$ and $v_j^{(t+1)} = v_j'$
    end
    Sample $N_r$ indices $i$ from $\{1 \dots N\}$, store them in $I$
    Calculate mean velocity $V$ and temperature $T$ of the sampled particles in $I$
    Sample $N_r$ particles $w_i, i = 1 \dots N_r$ from a Maxwellian $M_{V,T}$
    for $i \in I$ do
        Set $v_i^{(t+1)} = w_i$;
    end
    Set $v_i^{(t+1)} = v_i^{(t)}$ for all $i : i \notin I$ and for all particles which have not collided.
end
```

**Algorithm 3:** A time relaxed version of the Monte Carlo scheme.

To implement the algorithms just presented, one needs to be able to sample from a given distribution function in order to:

- Choose a uniform distributed vector on the unit sphere for sampling the post collision velocities.
- Sample from the initial distribution to construct the initial data.

- Sample $N_c$ collision pairs from a bivariate uniform distribution.
- Sample from a Maxwellian in the time relaxed schemes.

A possible way construct random variables which are distributed as a distribution function $F$ with density $f$ is inverse sampling. To this end consider a random number $\zeta \in [0, 1]$ with uniform distribution $U$. The random number $x = F^{-1}(\zeta)$ is distributed as $F$:

$$P(x \leq t) = P(F^{-1}(\zeta) \leq t) = P(\zeta \leq F(t)) = F(t). \tag{3.1.9}$$

Thus, we construct a uniform distributed random number $\zeta \in [0, 1]$ and solve $x = F^{-1}(\zeta)$ to end up with a random number $x$ sampled from $F$.

As an example we show how to

**Sample uniform distributed points from the unit sphere**

We denote the point on the unit sphere $S^2$ by its Polar coordinates $(\phi, \theta) \in [0,\, 2\pi] \times [0,\, \pi]$. The probability that $x$ lies in an area $A$ is given by

$$P(x \in A) = \int\limits_A \frac{1}{4\pi} ds = \int\limits_{\phi_0}^{\phi_1} \frac{1}{2\pi} d\phi \int\limits_{\theta_0}^{\theta_1} \frac{\sin(\theta)}{2} d\theta. \tag{3.1.10}$$

According to the above equation, if the Polar coordinates are independent and distributed uniform w.r.t. $\phi$ and distributed as $F_\theta(\theta) := \frac{1-\cos(\theta)}{2}$ w.r.t. $\theta$, then $x$ is distributed uniformly on the sphere $S^2$. Thus, in order to sample such points it is sufficient to sample $\phi$ from a uniform distribution and $\theta$ from $F_\theta$. For that purpose we use the above ideas of inverse sampling and introduce two uniform random numbers $\zeta_1, \zeta_2 \in [0, 1]$ with

$$\zeta_1 = \frac{\phi}{2\pi} \quad \text{and} \quad \zeta_2 = \frac{1 - \cos(\theta)}{2}. \tag{3.1.11}$$

Solving for $\phi$ and $\theta$ results in two random variables $\phi, \theta$ which are sampled from the desired distributions

$$2\pi\zeta_1 = \phi \quad \text{and} \quad \arccos(1 - 2\zeta_2) = \theta. \tag{3.1.12}$$

Finally, the point $(x, y, z)$ on the sphere is computed via

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \cos(2\pi\zeta_1)\sin(\arccos(1-2\zeta_2)) \\ \sin(2\pi\zeta_1)\sin(\arccos(1-2\zeta_2)) \\ \cos(\arccos(1-2\zeta_2)) \end{pmatrix}. \tag{3.1.13}$$

It is clear, that this technique is restricted to very special cases, when the effort for the inversion of the distribution function is acceptably moderate. If this is not the case, other techniques, e.g. acceptance-rejection techniques are preferred to sample from a given distribution $F$.
We note that this procedure can also be used to sample from a Maxwellian.

## 3.2 Deterministic Methods

When turning to deterministic approaches, the high complexity of the collision operator is a real challenge. A popular class of methods is based on truncated Fourier series [PP96,PR00,FM10,PR]. Typically these methods require a truncation of the momentum domain for the solution function and also for the collision integrals resulting in a perturbation of the conservation properties. By adding periodicity to the solution function additional errors may occur in the calculation of the collision integrals as we will see in the sequel.

In addition, we present discrete velocity models. The main idea of such methods is to replace the continuous momentum domain by a discrete lattice. This yields evolution equations for $f_i(t) := f(t, v_i)$. For these methods conditions can be derived to ensure that a discrete $H$-theorem holds and that mass, momentum and energy are conserved. Specific attention has to be paid on the integration over the sphere when calculating the collision integrals.

### 3.2.1 Fourier series expansion

We present the spectral method published in [PP96, PR00] for the space homogeneous Boltzmann equation. The first problem one faces in the approximation is the unbounded domain the distribution function $f$ is defined on and the unbounded domain of integration in the collision integrals. A pure restriction of the integration domain for the collision integrals violates the conservation laws for mass, momentum and energy. By means of the following proposition it is possible to truncate the integration domain for solutions with compact support without violating the conservation properties.

**Proposition 3.2.1.** *Let* $\operatorname{Supp}(f) \subset \mathcal{B}(0,R)$*, where* $\mathcal{B}(x,r)$ *denotes the sphere with center at* $x$ *and radius* $r$*. Then there holds*

*(i)* $\operatorname{Supp}(Q(f)) \subset \mathcal{B}(0, \sqrt{2}R)$

*(ii)*

$$Q(f)(v) = \int\limits_{\mathcal{B}(0,2R)} \int\limits_{S^2} B(|g|, g \cdot e')[f(v')f(w') - f(v)f(v-g)]\, de' dg,$$

*where* $g := v - w$ *and* $v', w', v - g \in \mathcal{B}(0, (2+\sqrt{2})R)$.

We assume that the initial distribution has compact support in $\mathcal{B}(0,R)$. We approximate it on a 3 dimensional cube $[-T,T]^3$ by its truncated Fourier series and extend it by periodicity. By adding periodicity we have to choose the cube large enough, in order to calculate the collision integrals without contributions from the neighbouring cubes. This situation can be sketched as in Figure 3.2.1, which is found in [PR00]: if the distribution has support in $B(0,R)$, then we need to evaluate $f$ inside the sphere $\mathcal{B}(0,(2+\sqrt{2})R)$ to calculate the collision integral. In order to avoid errors due to the periodic extension, this sphere should not overlap with the support of the copies of the distribution. This is achieved if $T = \frac{3+\sqrt{2}}{2}R$. Note that the support of $f$ grows w.r.t. time, what needs to be considered during time stepping.

In order to avoid scaling in the argument of the Fourier series expansion, the presentation of the method is restricted to the case where $T = \pi$ and consequently $R = \frac{2}{3+\sqrt{2}}\pi$.
The Fourier series expansion on the cube $[-T,T]^3$ is denoted as

$$f_N(v) = \sum_{k=-N}^{N} \hat{f}_k e^{iv\cdot k}, \quad \text{where} \quad \hat{f}_k = \int\limits_{[-\pi,\pi]^3} f(v) e^{ik\cdot v}\, dv. \tag{3.2.1}$$

In the above equation, $N$ is the order of expansion. Due to readability the 3 dimensional sum over the Fourier modes is denoted as a single sum over the expansion order $N$. Thus, we identify

$$\sum_{k=-N}^{N} \dots \quad \text{with} \quad \sum_{k_x=-N}^{N} \sum_{k_y=-N}^{N} \sum_{k_z=-N}^{N} \dots \ . \tag{3.2.2}$$

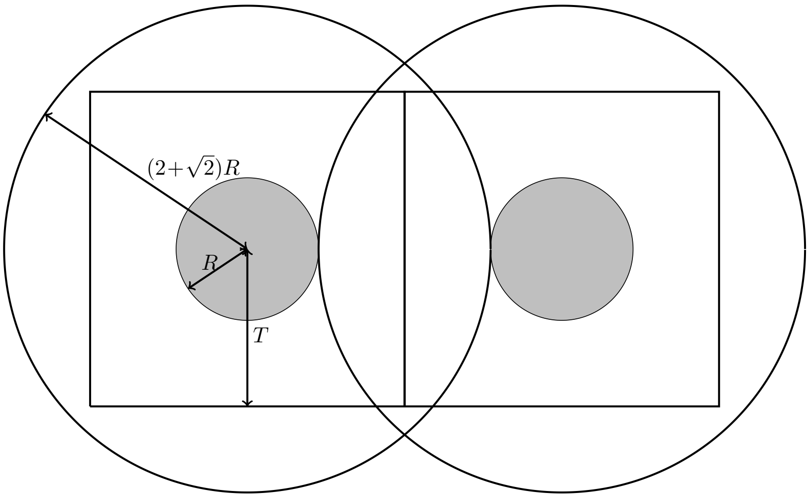


Figure 3.2.1: Here we show the support of the distribution function, as well as a copy due to the periodic extension in two dimensions. The gray shaded area is the actual support of $f$, the large circle corresponds to the domain, where $f$ has to be evaluated to calculate the collision integrals. The squares are the 2d cubes $[-T,T]^2$, with $T = \frac{1}{2}(3+\sqrt{2})R$. On the cube the Fourier approximation is performed.

In addition, by $\hat{f}_k$ the quantity $\hat{f}_{(k_x,k_y,k_z)}$ is denoted.
With the expansion (3.2.1) one obtains for the collision integral

$$Q(f_N)(v) = \sum_{h,k=-N}^{N} \hat{B}(h,k)\hat{f}_h\hat{f}_k e^{iv\cdot(h+k)}, \tag{3.2.3}$$

where the kernel modes $\hat{B}(h,k)$ are given by

$$\hat{B}(h,k) = \int\limits_{\mathcal{B}(0,2R)} \int\limits_{S^2} B(|g|, g\cdot e') \left[ e^{-i\frac{g}{2}\cdot(h+k)+i\frac{|g|}{2}e'\cdot(h-k)} - e^{-ig\cdot k} \right] de'dg. \tag{3.2.4}$$

The kernel modes are independent of the specific function $f$ and also independent of the evaluation point $v$, depending only on the interaction law. Moreover, one can prove the following statement about their dependency on the arguments $h$ and $k$ [PR00].

**Proposition 3.2.2.** *$\hat{B}(h,k)$ is a function of $|h-k|$, $|h+k|$ and $(h-k)\cdot(h+k)$. For a variable hard sphere gas where $B(v,w,e') = C|v-w|^\alpha$, this dependency reduces even to $|h-k|$ and $|h+k|$.*

The previous proposition states for which pairs $(h, k)$ one needs to store information in actual computations. For the case of the variable hard sphere model, the values of $\hat{B}$ depend on two parameters and thus, $\hat{B}$ can be stored in a matrix of size $\mathcal{O}(N^6)$. Considering symmetry, this number reduces in practice.
To obtain an ODE system for the coefficients we expand $Q(f_N)$ to its Fourier series of order $N$ and denote this projection by $Q_N$.

$$
\begin{gathered}
Q_N(f_N)(v) = \sum_{l=-N}^{N} \hat{Q}_l e^{iv\cdot l} \\
\text{where} \\
\hat{Q}_l := \sum_{\substack{h,k=-N \\ h+k=l}}^{N} \hat{f}_h \hat{f}_k \hat{B}(h,k),
\end{gathered}
\tag{3.2.5}
$$

The above equation shows that $Q_N(f_N)(v)$ can be evaluated with $\mathcal{O}(N^6)$ operations, where $N$ is the expansion order. For two dimensions $\mathcal{O}(N^4)$ operations are obtained.
Now we plug the expansion (3.2.1) as well as the representation for the collision integral (3.2.5) into $\frac{\partial f_N}{\partial t}(v) = Q_N(f_N)(v)$ to obtain a system of ODEs for the Fourier coefficients.

$$
\frac{\partial}{\partial t} \sum_{k=-N}^{N} \hat{f}_k(t) e^{ik\cdot v} = \sum_{l=-N}^{N} \hat{Q}_l e^{iv\cdot l}, \qquad v = \frac{\pi}{N} j,\ j \in \{-N \dots N\}^3. \tag{3.2.6}
$$

The initial condition for the Fourier coefficients is obtained from $f(0, v)$. We note that the evaluation of (3.2.6) at all nodes $v_j = j\frac{\pi}{N}$ can be performed using the fast Fourier transform resulting in $\mathcal{O}(N^3 \log(N))$ operations [JWC65].

**Remark 3.2.3.** *In [PP96], the authors give a comment on the calculation of $\hat{Q}_l$ for Maxwellian molecules. For collision kernels independent of the relative velocity, they use a representation of the collision operator found by Bobylev [Bob88]. This representation suggests that $\hat{Q}_l$ in (3.2.5) is well approximated if the sum is restricted to those $h, k$ such that $h \cdot k = 0$.*

### Galerkin Projection

The above expansion of the distribution function can also be used in a Galerkin method as presented in [PR00]. In addition to the previous approach, the usual splitting of the collision operator to its gain and loss term is used. As before, we assume that the initial distribution has compact support in $B(0, R)$. The approximation is also done on a cube $[-T, T]^3$, with $T = \frac{3+\sqrt{2}}{2}R$. We restrict the presentation again to the case where $T = \pi$ to simplify the presentation. According to proposition 3.2.1, the splitting reads

$$\begin{aligned} Q(f_N) &= Q^+(f_N) - f_N L(f_N), \\ &\text{where} \\ Q^+(f_N) &= \int\limits_{\mathcal{B}(0,2R)} \int\limits_{S^2} B(|g|, g \cdot e') f(v') f_N(w')\, de' dg, \\ L(f_N) &= \int\limits_{\mathcal{B}(0,2R)} \int\limits_{S^2} B(|g|, g \cdot e') f_N(v-g)\, de' dg. \end{aligned} \tag{3.2.7}$$

Instead of plugging the expansion directly into the homogeneous equation, the residual is required to be orthogonal to all trigonometric polynomials of order smaller or equal to $N$ w.r.t. each direction.

$$\int\limits_{[-\pi,\pi]^3} \left( \frac{\partial f_N}{\partial t} + f_n L(f_N) - Q^+(f_N) \right) e^{i(j \cdot v)}\, dv = 0, \qquad j \in \{-N \dots N\}^3. \tag{3.2.8}$$

The expressions for $Q^+(f_N)$ and $f_N L(f_N)$ result again in a double sum over the Fourier coefficients.

$$\begin{aligned} Q^+(f_N) &= \sum_{h,k=-N}^{N} \hat{f}_k \hat{f}_h \hat{B}(h,k) e^{i(h+k)\cdot v} \\ f_N L(f_N) &= \sum_{h,k=-N}^{N} \hat{f}_k \hat{f}_h \hat{B}(k,k) e^{i(h+k)\cdot v} \\ &\text{with} \\ \hat{B}(h,k) &= \int\limits_{\mathcal{B}(0,2R)} \int\limits_{S^2} B(|g|, e') e^{-i\frac{g}{2}(h+k) + i\frac{|g|}{2} e' \cdot (h-k)}\, de' dg. \end{aligned} \tag{3.2.9}$$

The different notation for the kernel modes is due to the splitting of $Q$.
Now by the orthogonality of the trigonometric functions, i.e.

$$\int\limits_{[-\pi,\pi]^3} e^{ik\cdot v} e^{il\cdot v}\, dv = (2\pi)^3 \delta_{k_x,l_x}\delta_{k_y,l_y}\delta_{k_z,l_z} \tag{3.2.10}$$

and the expressions (3.2.9) for the collision operator, (3.2.8) turns into

$$\frac{\partial \hat{f}_j}{\partial t} + \sum_{\substack{h,k=-N\\ h+k=j}}^{N} \hat{f}_k \hat{f}_h \hat{B}(k,k) = \sum_{\substack{h,k=-N\\ h+k=j}}^{N} \hat{f}_k \hat{f}_h \hat{B}(h,k), \qquad j \in \{-N \dots N\}^3. \tag{3.2.11}$$

Taking into account the restriction of the summation indices and extending the Fourier coefficients by zero, the scheme can be written as

$$\frac{\partial \hat{f}_j}{\partial t} + \sum_{k=-N}^{N} \hat{f}_{j-k} \hat{f}_k \hat{B}(k,k) = \sum_{k=-N}^{N} \hat{f}_{j-k} \hat{f}_k \hat{B}(j-k,k) \qquad j \in \{-N \dots N\}^3. \tag{3.2.12}$$

The initial condition for the $k$-th Fourier coefficient is obtained as the $k-$th Fourier coefficient of $f0,v)$. We note the the evaluation of (3.2.12) needs $\mathcal{O}(N^6)$ operations in general.

For the analysis of the method and numerical results we refer to [PR00].

### 3.2.2 Discrete velocity models

Discrete velocity models are based on the assumption that the particles under consideration can only have velocities on a discrete set of values $V = \{\zeta_i, i = 1 \dots N\} \subset \mathbb{R}^3$. The distribution function $f(t,x,v)$ is replaced by the discrete values $f_i(t,x)$ which are identified with $f(t,x,\zeta_i)$. The components $f_i$ evolute according to

$$\begin{aligned} &\frac{\partial f_i}{\partial t} + \zeta_i \cdot \nabla f_i = Q_i(f) \\ &\text{with} \\ &Q_i(f) = \sum_{j,k,l \in \{1 \dots N\}} A_{ij}^{kl} (f_k f_l - f_i f_j) \quad i = 1 \dots N, \end{aligned} \tag{3.2.13}$$

where $A_{ij}^{kl}$ are constants. They are interpreted as the rates of those collisions transferring particles $(\zeta_i, \zeta_j)$ into $(\zeta_k, \zeta_l)$.

Now it can be shown that

$$\sum_i Q_i(f) \begin{pmatrix} 1 \\ \zeta_i \\ |\zeta_i|^2 \end{pmatrix} = 0 \quad \text{and} \quad \sum_i Q_i(f) \log(f_i) \leq 0 \tag{3.2.14}$$

hold, provided that the coefficients $A_{ij}^{kl}$ satisfy the symmetries

$$A_{ij}^{kl} = A_{ji}^{lk} \quad \text{and} \quad A_{ij}^{kl} = A_{kl}^{ij}, \tag{3.2.15}$$

and that momentum and energy are collision invariants.

$$\zeta_i + \zeta_j = \zeta_k + \zeta_l \quad \text{and} \quad |\zeta_i|^2 + |\zeta_j|^2 = |\zeta_k|^2 + |\zeta_l|^2, \quad \forall (i,j,k,l) : A_{ij}^{kl} \neq 0. \tag{3.2.16}$$

In other words, the conservation properties of the Boltzmann equation are satisfied on the discrete level and a discrete $\mathcal{H}$-theorem holds.
The method described in the sequel was presented in [Bue96].

The velocity space $\mathbb{R}^3$ is discretized as

$$v_i = ih, \qquad i = (i_x, i_y, i_z) \in \mathbb{Z}^3, \; h > 0. \tag{3.2.17}$$

The collision integrals $\int_{\mathbb{R}^3} \int_{S^2} \ldots de' dw$ evaluated at $v = v_i$ are approximated via a sum over the set of discrete velocities.

$$\begin{aligned} &\int_{\mathbb{R}^3} \int_{S^2} B(v, w, e')(f(v')f(w') - f(v)f(w))\, de' dw \\ &\qquad \approx h^3 \sum_{j \in \mathbb{R}^3} \int_{S^2} B(v_i, v_j, e')(f(v_i')f(v_j') - f(v_i)f(v_j))\, de'. \end{aligned} \tag{3.2.18}$$

For the loss term only values of $f$ at the lattice are needed. The gain term depends on the values of $f$ via an integral over the unit sphere. In the sequel, the integral over the surface of the unit sphere shall be approximated by a quadrature rule. The biggest issue when doing so is finding integration nodes on the unit sphere $S^2$ under the restriction that for given pre collision velocities $(v_i, v_j)$ the post collision velocities $(v_i', v_j')$ shall again be nodes on the lattice, i.e. $\exists\, k, l : (v_i', v_j') = (v_k, v_l)$.

For given pre collision velocities $v_i, v_j$ one sees that $v_i', v_j'$ vary on a sphere of radius $\frac{1}{2}|v_i - v_j|$ with center at $\frac{1}{2}(v_i + v_j)$ if $e'$ varies on the unit sphere. If $v_i'$ is found on the lattice then also $v_j'$, which is located diametrically opposite, is a node of the lattice. For such post collision velocities there exists a unique $e_{ij}^{kl} \in S^2$ such that

$$\tfrac{1}{2}(v_i + v_j) + \tfrac{1}{2}e_{ij}^{kl}|v_i - v_j| = v' = v_k \tag{3.2.19}$$

and

$$\tfrac{1}{2}(v_i + v_j) - \tfrac{1}{2}e_{ij}^{kl}|v_i - v_j| = w' = v_l, \quad l = i + j - k. \tag{3.2.20}$$

For given pre collision velocities $(v_i, v_j)$ these post collision velocities are characterized by the set $S_{ij}$ which is defined as

$$S_{ij} := \{(k, l) \in \mathbb{Z}^3 \otimes \mathbb{Z}^3 : i + j = k + l,\ |k|^2 + |l|^2 = |i|^2 + |j|^2\}, \tag{3.2.21}$$

expressing conservation of momentum and energy. In addition to the sets $S_{ij}$, a set of integration nodes corresponding to $S_{ij}$ is defined via

$$I_{ij} = \{e_{ij}^{kl} \in S^2 : \text{(3.2.19) and (3.2.20) hold}\}. \tag{3.2.22}$$

The weights are chosen equal to $\frac{|S^2|}{|I_{ij}|}$, where $|S^2| = 4\pi$ denotes the surface of the unit sphere $S^2$ and $|I_{ij}|$ denotes the cardinality of the set $I_{ij}$.

With the above notation, for a fixed value of $j \in \mathbb{R}^3$, the integral over the unit sphere is approximated via

$$\begin{aligned}
&\int\limits_{S^2} B(v_i, w_i, e')(f(v_i')f(v_j') - f(v_i)f(v_j))\, de \\
&\qquad \approx \frac{4\pi}{|I_{ij}|} \sum_{e_{ij}^{kl} \in I_{ij}} B(v_i, v_j, e_{ij}^{kl})(f(v_i')f(v_j') - f(v_i)f(v_j)) \\
&\qquad = \frac{4\pi}{|I_{ij}|} \sum_{(k,l) \in S_{ij}} B(v_i, v_j, e_{ij}^{kl})(f(v_k)f(v_l) - f(v_i)f(v_j)).
\end{aligned} \tag{3.2.23}$$

Consequently, (3.2.18) reads

$$Q(f)(v_i) \approx 4\pi h^3 \sum_{j\in\mathbb{Z}^3} \frac{1}{|I_{ij}|} \sum_{(k,l)\in S_{ij}} B(v_i, v_j, e_{ij}^{kl})(f_k f_l - f_i f_j). \tag{3.2.24}$$

The above equation can be rewritten to fit into the general notation of discrete velocity methods.

$$\begin{gathered} Q(f)(v_i) \approx \sum_{(j,k,l)\in(\mathbb{Z}^3)^3} A_{ij}^{kl}(f_k f_l - f_i f_j) \\ \text{where} \\ A_{ij}^{kl} = \begin{cases} \frac{4\pi B(v_i,v_j,e_{ij}^{kl})}{|I_{ij}|} & \text{if } (k,l) \in S_{ij} \\ 0 & \text{otherwise} \end{cases}. \end{gathered} \tag{3.2.25}$$

The coefficients $A_{ij}^{kl}$ satisfy the symmetry properties

$$A_{ij}^{kl} = A_{ji}^{kl}, \tag{3.2.26}$$

expressing that 2 pre collision particles are indistinguishable. By the relation

$$A_{ij}^{kl} = A_{ij}^{lk}, \tag{3.2.27}$$

the same statement is obtained for the post collision particles. Moreover, there also holds

$$A_{ij}^{kl} = A_{kl}^{ij}. \tag{3.2.28}$$

Combining the final three equations, one obtains that (3.2.14) is satisfied (Note that momentum and energy are collision invariants by construction), and therefore the conservation laws are satisfied on the discrete level and a discrete $H$-theorem holds.

One may ask about the quality of the approximation of the unit sphere integral. In [Bue96] a result about splitting an integer into a sum of 3 squared integers [HW79] is adapted in order to show that $|S_{i,j}|$ behaves like $\frac{4\pi|i-j|}{6}$, the distribution of the integration nodes on the sphere seems to be quite arbitrary, as reported in [Bue96]. Rigorous error estimates are not known to the author for such quadrature formulas.

**Remark 3.2.4.** *In [Bue96], the analysis of the method is done on the unbounded lattice. For actual computations, the lattice as well as the sets $S_{ij}$ have to be restricted to a bounded domain $V$. The sets $S_{ij}$ are replaced by $\tilde{S}_{ij}$, defined via*

$$\tilde{S}_{ij} := \{(k,l) \in S_{ij} : v_k, v_l \in V\} \quad \textit{with } v_i, v_j \in V. \tag{3.2.29}$$

**Remark 3.2.5.** *A possibility to avoid integration over the surface of the unit sphere is presented in [PH02]. In this approach the Carleman representation of the collision integrals is used, which provides an expression for the collision integrals without integrals over the surface of a sphere. We do not present details of this approach, refer to [PH02] for more information.*

# 4 A Discontinuous Galerkin approach

This section is devoted to the presentation of the method developed in the context of the underlying thesis. At the beginning we present the concept of Discontinuous Galerkin methods for a standard linear transport problem. For the Boltzmann transport operator we do a reformulation to fit it in the above mentioned class of transport operators in a 2+2 dimensional setting. We stabilize our DG discretization by choosing upwind fluxes in the arising skeleton integrals.

The main part of the section deals with efficient application of the collision integrals. Our idea is based on a reformulation of the integrals in terms of the mean and relative velocity of the colliding particles. For the reformulated integrals we then propose a polynomial basis in which the innermost integrals are diagonal and thus easy and fast to apply. In order to keep efficiency, we show a technique to deal efficiently with the necessary transformation between the bases. The key ingredient to come to that end will be a splitting of the transformation into 2 cheaper ones.

Our discretization consists of 2 parameters within each space element which are closely related to temperature and mean velocity. Actual computations show that the stability of the method is in strong correlation with the choice of these parameters. Thus, we conclude this section with a discussion about their choice.

At that point we switch to a 2-dimensional presentation for both, $x$ and $v$.

## 4.1 Discontinuous Galerkin Discretization

In the context of hyperbolic conservation laws, DG methods are well established. Already in 1973, Reed and Hill [RH73] proposed an approximation of the neutron transport equation by discontinuous Ansatz functions. Convergence proofs for equations of the form $\mathrm{div}(bu) + \alpha u = f$ were then obtained by Johnson and Pitkäranta [JP86]. They proved error estimates of the form $\|e(T)\|_{L_2(0,1)} \leq C|u_0|_{H^{k+1}(0,1)} h^{k+\frac{1}{2}}$, with $e$ denoting the error $e(T) := u(T) - u_h(T)$ at time $T$. A similar result was also obtained by LeSaint and Raviart [LR74]. In the case of non linear conservation laws Cockburn and Shu presented in a series of papers so called Runge Kutta DG methods [CS89, CLS89, CHS90, CS01, CKS00]. These methods consist of the typical DG space discretization paired with explicit Runge Kutta time stepping schemes. A comparison of different techniques including discontinuous as well as continuous Galerkin methods for first order linear hyperbolic equations was presented by Falk in [Fal00]. We additionally refer the reader to the finite element textbooks of Johnson [Joh12], Ern and Guermond [EG04] and the monograph of Hesthaven and Warbuton [HW08]. Applications specifically for Euler and Navier-Stokes equations are presented in [BR97b, BR97a].

The impact of DG methods was significantly raised in the last decades, since besides their mathematical properties, these methods are very well suited for modern hardware architecture: The schemes present in literature allow a straight forward implementation in parallel. Adaptivity in terms of different polynomial degrees, the ability to handle non conformal meshes are in addition among the benefits of DG methods.

In the case of elliptic equations, continuous Galerkin methods are known to perform very well, also the theory is quite satisfactory [Joh12,Bra92]. Due to the need of solving convection-diffusion problems with dominant convection, DG methods were also developed for elliptic equations, due to their good properties concerning the convective part. We refer the reader to [ABCM01] for DG methods for elliptic equations.

### 4.1.1 DG for a scalar linear convection equation

We present a short derivation of the DG method for a linear hyperbolic first order PDE in the sequel. Let us consider the equation

$$\begin{aligned} \frac{\partial u}{\partial t}(t,x) + \operatorname{div}(b(x)u(t,x)) = 0 \quad & x \in \mathbb{R}^d \\ u(0,x) = u_0(x) \quad & x \in \mathbb{R}^d, \end{aligned} \tag{4.1.1}$$

with $\operatorname{div}(b) = 0$. The discontinuous Galerkin space discretization for the above problem is described in the sequel. First we assume a subdivision of $\mathbb{R}^d = \bigcup\limits_{K \in \mathcal{T}_h} K$ is given. Now we look for a discontinuous approximation $u_h$ of the solution $u$. For this purpose we define the space

$$V_h^{DG} := \{u_h : u_h\big|_K \in P^p(K) \,\forall\, K \in \mathcal{T}_h\}, \tag{4.1.2}$$

where $P^p(K)$ denotes the polynomial space on the element $K$ of degree at most $p$. The space $V_h^{\mathrm{DG}}$ is the usual DG approximation space. Let us denote by $\phi_{j,K}$, $j = 1 \ldots \mathrm{ndof}_K$ the $j-$th basis polynomial on the element $K$, extended with 0 outside of $K$. Thus, we expand $u_h$ to

$$u_h(t,x) = \sum_{K\in\mathcal{T}_h} \sum_{j=1}^{\mathrm{ndof}_K} u_{j,K}(t)\phi_{j,K}(x). \tag{4.1.3}$$

For $u_h \in V_h^{\mathrm{DG}}$ we now require the transport equation to be satisfied in a weak sense in each element $K \in \mathcal{T}_h$. Thus, we multiply by a test function $v \in V_h^{\mathrm{DG}}$, integrate over the element $K$ and finally sum over all elements in the triangulation. The restriction to single elements is necessary since we want to integrate by parts. Due to the discontinuities of $u_h$ this holds only on the element level. We obtain our variational formulation as

$$\sum_{K\in\mathcal{T}_h} \frac{\partial}{\partial t}\int\limits_K u_h v\,dx + \int\limits_{\partial K} b\cdot n u_h v\,ds - \int\limits_K b u_h \cdot \nabla v\,dx = 0 \quad \forall v \in V_h^{\mathrm{DG}}. \tag{4.1.4}$$

In terms of the basis polynomials and the expansion (4.1.3), the above formulation reads

$$\begin{aligned}\sum_{K\in\mathcal{T}_h} \sum_{j=1}^{\mathrm{ndof}_K} \Bigg( \frac{\partial}{\partial t}\int\limits_K & u_{j,K}\phi_{j,K}\phi_{j',K} + \\ & \int\limits_{\partial K} b\cdot n u_{j,K}\phi_{j,K}\phi_{j',K}\,ds - \int\limits_K u_{j,K}\phi_{j,K} b\cdot\nabla\phi_{j',K}\,dx \Bigg) = 0.\end{aligned} \tag{4.1.5}$$

The first term results in a mass matrix multiplied with the time derivative of the coefficient vector, also the third term doesn't reveal unexpected troubles. The second term needs to be considered with more care. If we use (4.1.5) exactly as it is, we realize immediately that there is no information exchange across element boundaries. For a transport equation this sounds already inappropriate, the transport would take place within each element only, but not globally. In other words, the outflow from one element has no connection to the inflow of a neighbouring element. Apart from these empirical observations the method becomes highly unstable without additional modification of the skeleton terms.

The above considerations lead to the definition of the upwind flux. For this purpose fix an element $K$ in the mesh and split its boundary $\partial K = \bigcup_{i=1}^{n_{\text{edges}}} E_i$. Moreover we denote by $K_i, i = 1 \ldots n_{\text{edges}}$ the elements sharing the edge $E_i$ with our actual element $K$. The upwind value is defined the following way:

$$u_h^{\mathrm{up},K,E_i}(x) := \begin{cases} u_h\big|_K(x) & b(x)\cdot n(x) > 0, \\ u_h\big|_{K_i}(x) & \text{otherwise.} \end{cases}$$

With the upwind flux defined we can state the final DG formulation that is solved. In fact we simply replace $u_h$ in the skeleton integrals in (4.1.4) by the upwind value $u_h^{\mathrm{up}}$.

$$\sum_{K\in\mathcal{T}_h} \frac{\partial}{\partial t}\int\limits_K u_h v\,dx + \int\limits_{\partial K} b\cdot n u_h^{\mathrm{up}} v\,ds - \int\limits_K b u_h \cdot \nabla v\,dx = 0 \quad \forall v \in V_h^{\mathrm{DG}}. \tag{4.1.6}$$

The definition of the upwind value provides a quite intuitive solution for the evaluation of the skeleton integrals in (4.1.4). In addition for a continuous solution the upwind value reduces to the function value on the boundary. This implies that the DG formulation with upwind fluxes is consistent.

The benefits of DG methods are of different nature. First of all there is no coupling between the local basis functions from one element and another one. This provides an easy way to use different polynomial orders on different elements, making the method highly suitable for adaptivity.
Second, a time discretization by a forward Euler method with time step $\tau$ leads to

$$\begin{aligned} M\frac{u^{(n+1)}-u^{(n)}}{\tau} + Fu^{(n)} = 0 \quad &\Leftrightarrow \\ -\tau M^{-1}Fu^{(n)} + u^{(n)} = u^{(n+1)}&. \end{aligned} \tag{4.1.7}$$

In the above equation, $M$ denotes the mass matrix, i.e. $M_{ij} = \int_\Omega \phi_i\phi_j$ where $\phi_i, i = 1\ldots\mathrm{dim}(V_h^{\mathrm{DG}})$ denotes the basis functions in $V_h^{\mathrm{DG}}$. $F$ is the discrete analogy to the fluxes and $u$ denotes the coefficient vector in the expansion (4.1.3). Thus, in order to calculate $u^{(n+1)}$, the inverse of the mass matrix needs to be applied. Due to the basis functions being non trivial only on one element, the mass matrix $M$ results in a block diagonal matrix with non overlapping blocks. $M = \mathrm{diag}(M^{(1)},\ldots,M^{n_{\text{elements}}})$, where $M^{(K)} \in \mathbb{R}^{\mathrm{ndof}_K\times\mathrm{ndof}_K}$ is the block connected with the element $K$. This matrix is inverted cheaply, only the element matrices $M^{(K)}$ need actual inversion. Since $M^{-1}$ is again a block diagonal matrix, it can be applied efficiently. Note that there is no need to invert the matrix connected with the coupling terms what would need a global inversion. The situation is similar when explicit higher order Runge Kutta schemes are used. The inverse mass matrix has to be calculated, the coupling terms need only a forward application.
Third, the missing coupling between two elements also enables one to use non conformal meshes. However, the numerical integration in such a case is somewhat tricky: If the edge $E$ of the element $K$ is connected with the edges of two different elements $\tilde{K}$, $\hat{K}$, then the upwind value is not a

polynomial along the edge $E$, but a piecewise polynomial only. We use as usual Gauss quadrature formulas to numerically evaluate the integrals, these formulas are exact for polynomials but not for piecewise polynomials. Therefore one has to calculate the skeleton term separate on $E \cap \tilde{K}$ and $E \cap \hat{K}$ and sum up the contributions.

### 4.1.2 Applying DG to the Boltzmann transport operator

In the sequel we want to derive the discontinuous Galerkin formulation applied to the Boltzmann equation (1.2.1). Just as usual we multiply it with a test function $\phi = \phi(x, v)$ and integrate the result w.r.t. to $x$ over $\Omega$ and w.r.t. $v$ over $\mathbb{R}^2$:

$$\frac{\partial}{\partial t} \int\limits_{\Omega\times\mathbb{R}^2} f\phi \, d(x,v) + \int\limits_{\Omega\times\mathbb{R}^2} \mathrm{div}_x(vf)\phi \, d(x,v) = \int\limits_{\Omega\times\mathbb{R}^2} Q(f)\phi \, d(x,v) \quad \forall \text{ suitable } \phi. \tag{4.1.8}$$

As before we use discontinuous polynomial test and trial functions on a finite element mesh $\mathcal{T}_h$ in the spatial variable $x$. As we have seen in (2.2.4) and (2.4.1), conservation of mass, momentum and energy arises by testing the Boltzmann equation with the collision invariants, which are polynomials. This has an important meaning for our method: If the collision invariants are included in the test space, then the discrete problem inherits the conservation properties directly from the continuous one. Thus, our discretization satisfies the same conservation equations for the macroscopic properties as the continuous equation.

The choice of the trial functions in the velocity direction deals with the kernel of the collision operator $Q(f(t, x, .))(v)$, the Maxwell distributions $M_{V,T}(v) = \frac{\rho}{T\pi} e^{-\left|\frac{v-V}{\sqrt{T}}\right|^2}$. We are particularly interested in solutions close to equilibrium. Thus, to have accurate approximation in such situations, we choose the trial functions as polynomials multiplied with the local equilibrium $M_{V,T}(v)$ in the velocity variable. A key ingredient – as confirmed by numerical examples – is the ability to vary the parameters $V$ and $T$ over space and time.

To fix notation, we define the mesh $\mathcal{T}_h := \{K_1, \dots K_r\}$, $h$ being the usual mesh size parameter. In addition we denote the space of polynomials of partial order at most $N$ on $\mathbb{R}^2$ via

$$V_N := Q^N(\mathbb{R}^2),$$

and the polynomial space of degree $k$ on a single element $K \in \mathcal{T}_h$ via

$$V_K := P^k(K).$$

Note that $V_K$ is the space of polynomials of total degree at most $k$. The global space on the spatial domain $\Omega$ is denoted as

$$V_h^{\mathrm{DG}} := \prod_{K \in \mathcal{T}_h} V_K,$$

being the usual DG approximation space in the spatial variable $x$. The full test space in momentum and position variables is defined as a tensor product of the DG approximation space and the polynomial space on $\mathbb{R}^2$.

**Definition 4.1.1.** *We define the test space on the domain* $\tilde{\Omega} := \Omega \times \mathbb{R}^2$ *via*

$$V_{h,N} := V_h^{DG} \otimes V_N.$$

*The trial space depends on additional parameters* $\overline{V}(x) \in \mathbb{R}^2$ *and* $0 < \overline{T}(x) \in \mathbb{R}$, *closely related to the macroscopic quantities bulk velocity* $V(x)$ *and temperature* $T(x)$. *These parameters are assumed to be piecewise constants with notation* $\overline{V}\big|_K \equiv \overline{V}_K$ *and* $\overline{T}\big|_K \equiv \overline{T}_K$. *The resulting space is denoted as*

$$\widetilde{V}_{h,N} := \prod_{K \in \mathcal{T}_h} V_K \otimes e^{-\left|\frac{v-\overline{V}_K}{\sqrt{\overline{T}_K}}\right|^2} V_N.$$

The space $\widetilde{V}_{h,N}$ has dimension ndof $:= \sum_{K \in \mathcal{T}_h} \dim(V_K)\dim(V_N)$. The shorthand notation for element wise degrees of freedom in $x$ direction is $\dim(V_K) =: \mathrm{ndof}_x$ and for the momentum space $\dim(V_N) =: \mathrm{ndof}_v$. The Maxwellian weight for the polynomials is termed element Maxwellian and is denoted by $M_{\overline{V}_K, \overline{T}_K}$.

The discontinuous Galerkin method is – as before – obtained by element-wise integration by parts of the transport term in (4.1.8). Assuming sufficient regularity, the exact solution $f$ satisfies:

$$\begin{aligned}\sum_{K \in \mathcal{T}_h} \frac{\partial}{\partial t} \int\limits_{K \times \mathbb{R}^2} f \phi \, d(x,v) + \int\limits_{\partial(K) \times \mathbb{R}^2} v \cdot n f \phi \, d(x,v) - \int\limits_{K \times \mathbb{R}^2} f v \cdot \nabla_x \phi \, d(x,v) \\ = \sum_{K \in \mathcal{T}_h} \int\limits_{K \times \mathbb{R}^2} Q(f) \phi \, d(x,v),\end{aligned} \tag{4.1.9}$$

for all test functions $\phi \in V_{h,N}$. In the above equation, $n$ is the unit outer normal vector to the spatial element $K$.

For the definition of the upwind flux it is convenient to interpret the transport operator as a standard linear transport operator in $\mathbb{R}^{2+2}$. This is achieved by introducing the new variable $y := (x, v) \in \mathbb{R}^4$, the wind vector $b := (v, 0) \in \mathbb{R}^4$, the new domain $\tilde{\Omega} := \Omega \times \mathbb{R}^2$ and the mesh $\tilde{\mathcal{T}}_h := \prod_{K \in \mathcal{T}_h} K \times \mathbb{R}^2$ with elements $\tilde{K}$. In that setting, the outer normal vector to an element of $\tilde{\mathcal{T}}_h$ results in $\tilde{n}(y) = \tilde{n}(x, v) = (n(x), 0)$, where $n$ is the outer normal vector to $K$ at position $x \in \partial K$. It is easy to see, that $\operatorname{div}_x(vf) = \operatorname{div}_y(bf)$ and $b \cdot \tilde{n} = v \cdot n$ hold. In the new variables the variational problem for the pure transport equation reads

$$
\begin{aligned}
&\text{Find } f_h \in \widetilde{V}_{h,N} : \\
&\sum_{\tilde{K} \in \tilde{\mathcal{T}}_h} \frac{\partial}{\partial t} \int\limits_{\tilde{K}} f_h \phi \, dy + \int\limits_{\partial(\tilde{K})} b \cdot \tilde{n} f_h \phi dy - \int\limits_{\tilde{K}} f_h b \cdot \nabla_y \phi \, dy = 0, \quad \forall \phi \in V_{h,N}.
\end{aligned} \tag{4.1.10}
$$

This is a problem of the form (4.1.4) in a 2+2 dimensional space. Since the trial and test space do not coincide, there is still a small difference. As for the lower dimensional case, we use the upwind flux for the evaluation of the skeleton integrals in the transport operator. Thus, we replace $f_h$ in the skeleton integrals by its upwind value $f_h^{\text{up}}$, with $f_h^{\text{up}}$ given by

$$
f_h^{\text{up}} := \begin{cases} f_h\big|_K & b \cdot \tilde{n} > 0, \\ f_h\big|_{\tilde{K}} & \text{otherwise.} \end{cases}
$$

In the above equation, $\tilde{K}$ is the corresponding neighbour element to $K$. The definition of the upwind flux is sketched in Figure 4.1.1.

The above definition of the upwind function holds for inner edges of the mesh $\mathcal{T}_h$ only. For a boundary edge the upwind value $f_h^{\text{up}}$ has to be adjusted. Instead of the solution value from the neighbouring element, the boundary value is used. For an inflow boundary condition (1.2.4c) this ends up in

$$
f_h^{\text{up}}(x, v) := \begin{cases} f_h\big|_K(x, v) & v \cdot n > 0, \\ f_{\text{in}}(x, v) & \text{otherwise,} \end{cases}
$$

while for the specular reflection (1.2.4a) obtains

$$
f_h^{\text{up}}(x, v) := \begin{cases} f_h\big|_K(x, v) & v \cdot n > 0, \\ f_h\big|_K(x, v - 2n \cdot vn) & \text{otherwise.} \end{cases}
$$

These two conditions are easily expressed in terms of the variables $b$, $\tilde{n}$ and $y$, to fit in the (2+2)d representation of the transport equation.

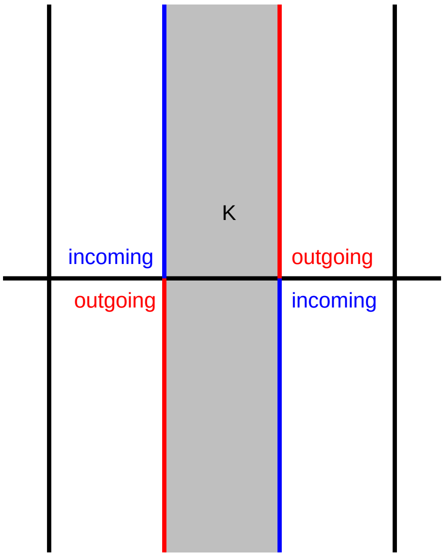


Figure 4.1.1: Sketch for a $1d \times 1d$ situation, where $\Omega \subset \mathbb{R}$ and the momentum space is also restricted to $\mathbb{R}$. The grey shaded domain is the actual element of interest $K \times \mathbb{R}$. In the upper half of the plane are the positive velocities, the negative ones in the lower part. The incoming boundary part consists of those $(x,v) : n(x) \cdot v < 0$, while the points $(x,v)$ on the outgoing part satisfy $n(x) \cdot v \geq 0,\ \forall\ (x,v) \in \partial K$. Consequently a particle on the incoming part of the boundary is transported to the inside of the element $K \times \mathbb{R}$ and vice versa for particles on the outgoing parts, justifying the notation "incoming" and "outgoing". In context of the upwind value, this means $f_h^{\text{up}} = f_h|_K$ on the outgoing and $f_h^{\text{up}} = f_h|_{\tilde{K}}$ on the incoming parts.

The last two equations show the simplicity in implementing the boundary conditions. We let the outgoing particles leave, corresponding to $v \cdot n \geq 0$ and have the incoming particles defined by the boundary condition. This simple treatment of boundary conditions is a result of the nature of the boundary conditions, since they are only imposed for velocities in $\mathbb{R}^2_{\text{in}}$ and thus are easy to incorporate in a DG method with upwind fluxes.

In order to arrive at a semi discrete equation we chose a basis $\{f_j, j = 0 \ldots \text{ndof} - 1\}$ of the trial space $\widetilde{V}_{h,N}$, as well as a basis $\{\phi_j, j = 0 \ldots \text{ndof} - 1\}$ of the test space $V_{h,N}$. We expand $f_h(t,x,v) = \sum_{j=0}^{\text{ndof}-1} c_j(t) f_j(x,v)$, plug everything into the variational formulation (4.1.9) and obtain a system of ODEs for the coefficient vector $c(t)$

$$M_h \frac{\partial c}{\partial t}(t) + A_h c(t) = Q_h(c(t)). \tag{4.1.11}$$

$M_h$ is the mass matrix, i.e. $(M_h)_{ij} = \int_{\tilde{\Omega}} f_i \phi_j$. $A_h$ denotes the discretization of the transport term and $Q_h$ is the application of the collision operator. To solve this system by a forward Euler scheme with time step $\tau_n = t_{n+1} - t_n$, we denote by $c^n \approx c(t_n), n = 0 \dots N_t$ and assume $c^0$ given. This results in

$$c^{n+1} = c^n + \tau_n M_h^{-1}(Q_h(c^n) - A_h c^n). \tag{4.1.12}$$

The situation is quite similar when using higher order Runge Kutta methods instead. Using these schemes, typically the inverse of the mass matrix arises in the calculation of $c^{n+1}$. Thus, a basis of $V_{h,N}$ respectively $\widetilde{V}_{h,N}$, generating a sparse mass matrix is preferred.

### 4.1.3 Polynomial basis in $\mathbf{V_{h,N}}$

We denote by $(x_{\text{ip}}, \omega_{\text{ip}}),\ i = 0 \dots N$ the nodes and weights of a Gauss-Hermite quadrature rule satisfying

$$\int_{\mathbb{R}} e^{-v^2} p(v) = \sum_{\text{ip}} \omega_{\text{ip}} p(x_{\text{ip}}),\ \forall\, p \in P^{2N+1}(\mathbb{R}). \tag{4.1.13}$$

The nodes are the roots of the $(N+1)$-st orthogonal polynomial w.r.t. the weighted inner product $\langle f, g\rangle := \int_{\mathbb{R}} e^{-x^2} f(x) g(x)\, dx$, the Hermite polynomials. The $j-$th weight is obtained as integral over the $j-$th Lagrange polynomial defined to the above mentioned roots, but is a non suitable representation for the computation of the weights, since the numerical evaluation of the integrals over the Lagrange polynomials already requires the knowledge of the weights.
Actual computation of the nodes is done via an eigenvalue problem, resulting from the three term recurrence relation satisfied by the orthogonal polynomials. The weights are properly scaled entries of the corresponding eigenvectors [Pla10, STW11].

The basis polynomials are defined as the Lagrange collocation polynomials to the Gauss-Hermite quadrature nodes and are denoted by $l$:

$$l_j(v) := \prod_{\substack{i=0 \\ i \neq j}}^{N} \frac{v - x_i}{x_j - x_i}. \tag{4.1.14}$$

The two dimensional basis is constructed as the tensor product of the 1d polynomials, its elements are denoted by $L_j, j = 0 \dots \mathrm{ndof}_v - 1$:

$$L_j(v) = l_u(v_x) l_v(v_y), \tag{4.1.15}$$

with $j = (N+1)u + v$. This gives a total number of $\mathrm{ndof}_v = (N+1)^2$ basis functions, $N$ is the maximum partial polynomial degree in the space $V^N$. Using the properties of the multivariate Lagrange polynomials $L_j$ and of the Gauss Hermite quadrature rule there holds

$$\begin{aligned}\int\limits_{\mathbb{R}^2} e^{-\left|\frac{v-\overline{V}_K}{\sqrt{\overline{T}_K}}\right|^2} L_m\left(\frac{v-\overline{V}_K}{\sqrt{\overline{T}_K}}\right) L_n\left(\frac{v-\overline{V}_K}{\sqrt{\overline{T}_K}}\right) dv &= \overline{T}_K \int\limits_{\mathbb{R}^2} e^{-|v|^2} L_m(v) L_n(v)\, dv \\ = \overline{T}_K \sum_{\mathrm{ip}=0}^{\mathrm{ndof}_v - 1} \omega_{\mathrm{ip}}^{(2)} L_m(x_{\mathrm{ip}}^{(2)}) L_n(x_{\mathrm{ip}}^{(2)}) = \overline{T}_K \delta_{m,n} \omega_n^{(2)} =: (M_V)_{n,m}, &\qquad n, m = 0 \dots \mathrm{ndof}_v - 1.\end{aligned} \tag{4.1.16}$$

Moreover, there also holds

$$\int_{\mathbb{R}^2} v e^{-\left|\frac{v-\overline{V}_K}{\sqrt{\overline{T}_K}}\right|^2} L_m\left(\frac{v-\overline{V}_K}{\sqrt{\overline{T}_K}}\right) L_n\left(\frac{v-\overline{V}_K}{\sqrt{\overline{T}_K}}\right) dv = x_n^{(2)} \overline{T}_K \delta_{m,n} \omega_n^{(2)}, \qquad n, m = 0 \dots \mathrm{ndof}_v - 1. \tag{4.1.17}$$

Note that in (4.1.16) and (4.1.17) the nodes $x_{\mathrm{ip}}^{(2)}$ and the weights $\omega_{\mathrm{ip}}^{(2)}$ correspond to the 2d Gauss Hermite formula with nodes $x_{\mathrm{ip}}^{(2)} = (x_i, x_j)$ and $\omega_{\mathrm{ip}}^{(2)} = \omega_i \omega_j$ with $\mathrm{ip} = (N+1)i + j$.

For given $\overline{V}_K$ and $\overline{T}_K$, the polynomials are scaled and shifted in the argument to incorporate the above orthogonality relations in the basis. Thus, we have

$$V_N = \mathrm{span}\{L_j\left(\frac{v-\overline{V}_K}{\sqrt{\overline{T}_K}}\right), j = 0 \dots \mathrm{ndof}_v - 1\}. \tag{4.1.18}$$

Note that the same space is obtained by $V_N = \mathrm{span}\{L_j(v)\}$, but the above notation additionally emphasizes the scaling of the basis functions we use in practice. The approximation properties of the scaled and non scaled basis polynomials clearly differ.

Denoting the spatial basis polynomials in a single element by $u_r, r = 0 \dots \mathrm{ndof}_x - 1$ and testing

$\phi_{r,m} = L_m\left(\frac{v-\overline{V}_K}{\sqrt{T_K}}\right)u_r(x)$ with $\phi_{r',n} = L_n\left(\frac{v-\overline{V}_K}{\sqrt{T_K}}\right)u_{r'}(x)$ results in

$$\begin{aligned}
&\int\limits_{K\times\mathbb{R}^2} \phi_{r,m}(x,v)\phi_{r',n}(x,v)d(x,v)\\
=&\int\limits_K u_r(x)u_{r'}(x)\,dx \int\limits_{\mathbb{R}^2} e^{-\left|\frac{v-\overline{V}_K}{\sqrt{T_K}}\right|^2} L_m\left(\frac{v-\overline{V}_K}{\sqrt{T_K}}\right) L_n\left(\frac{v-\overline{V}_K}{\sqrt{T_K}}\right)\,dv\\
=&\int\limits_K u_r(x)u_{r'}(x)\,dx\,(M_V)_{m,n}.
\end{aligned}$$

Thus, if the degrees of freedom are enumerated lexicographically i.e. $\phi_{0,0},\ldots,\phi_{0,\mathrm{ndof}_v-1},\ldots,\phi_{\mathrm{ndof}_x-1,0},\ldots,\phi_{\mathrm{ndof}_x-1,\mathrm{ndof}_v-1}$, the total mass matrix $M_K$ for a single spatial element $K$ is the Kronecker product $M_K = M_x\otimes M_V \in \mathbb{R}^{\mathrm{ndof}_x\mathrm{ndof}_v\times\mathrm{ndof}_x\mathrm{ndof}_v}$. Here $M_x$ denotes the spatial mass matrix, i.e. $(M_x)_{i,j} := \int_K u_j(x)u_i(x)\,dx,\ i,j = 0\ldots\mathrm{ndof}_x - 1$. Note that the mass matrix in velocity $M_V$ is diagonal only.

The global matrix $M_h$ results in a block diagonal matrix and is cheap to invert, only the element matrices $M_K$ need to be inverted. Its inverse is again a block diagonal matrix and thus cheap to apply.

**Remark 4.1.2.** *As we have noticed, the element mass matrices $M_K$ are the Kronecker product of the spatial mass matrix with the mass matrix in velocity space. The inverse of such a matrix is given by the Kronecker product of the inverse mass matrix in space and the inverse mass matrix in velocity, i.e. $M_K^{-1} = M_x^{-1}\otimes M_V^{-1}$. Thus, only the spatial mass matrix $M_x$ needs actual inversion to calculate $M_K^{-1}$.*

## 4.2 Application of the collision integrals

A crucial part of the scheme is the application of the collision integrals. This is not due to the discretization scheme but to the collision operator itself. For a given distribution function $f_h \in \widetilde{V}_{h,N}$ and a fixed spatial point the evaluation of the operator takes $O(N^6)$ operations.
In the sequel we consider a fixed element $K$ in the mesh $\mathcal{T}_h$. The collision integrals inside the

element $K \in \mathcal{T}_h$ are calculated via

$$\int\limits_{K\times\mathbb{R}^2} Q(f_h)(t,x,v)\phi(v)\,dv\,dx = \int\limits_K \underbrace{\left(\int\limits_{\mathbb{R}^2} Q(f_h(t,x,.))(v)\phi(x,v)\,dv\right)}_{:=g(x)} dx$$
$$= \sum_{\text{ip}} \omega_{\text{ip}} g(x_{\text{ip}}),$$

with the pair $(x_{\text{ip}},\omega_{\text{ip}})$ being an integration rule on the spatial element $K$, such that $\int_K p(x)\,dx = \sum_{\text{ip}} \omega_{\text{ip}} p(x_{\text{ip}})\ \forall p \in V_h^{\text{DG}}\big|_K$. The values of $g$ shall be calculated independent of $\overline{T}_K$ and $\overline{V}_K$. Therefore we substitute

$$\frac{v-\overline{V}_K}{\sqrt{\overline{T}_K}} =: \tilde{v} \quad \text{and} \quad \frac{w-\overline{V}_K}{\sqrt{\overline{T}_K}} =: \tilde{w}. \tag{4.2.1}$$

The post collision velocities in terms of the new variables are given by

$$\begin{aligned} v' &= \frac{\sqrt{\overline{T}_K}\tilde{v}+\overline{V}_K+\sqrt{\overline{T}_K}\tilde{w}+\overline{V}_K}{2} + \sqrt{\overline{T}_K}e'\frac{|\tilde{v}-\tilde{w}|}{2} = \sqrt{\overline{T}_K}\tilde{v}'+\overline{V}_K \\ w' &= \frac{\sqrt{\overline{T}_K}\tilde{v}+\overline{V}_K+\sqrt{\overline{T}_K}\tilde{w}+\overline{V}_K}{2} - \sqrt{\overline{T}_K}e'\frac{|\tilde{v}-\tilde{w}|}{2} = \sqrt{\overline{T}_K}\tilde{w}'+\overline{V}_K, \end{aligned} \tag{4.2.2}$$

where $\tilde{v}' = \frac{\tilde{v}+\tilde{w}}{2}+e'\frac{|\tilde{v}-\tilde{w}|}{2}$ and $\tilde{w}' = \frac{\tilde{v}+\tilde{w}}{2}-e'\frac{|\tilde{v}-\tilde{w}|}{2}$ are the post collision velocities in the variables $\tilde{v}$ and $\tilde{w}$.
In the next equation we transform the collision integral according to the substitution (4.2.1) and use the representation (4.2.2) for the post collision velocities. In addition we introduce the normalized distribution $f_h^{0,1}(t,x,v) := f_h(t,x,\sqrt{\overline{T}_K}v+\overline{V}_K)$ to obtain for the collision integral

$$\int\limits_{\mathbb{R}^2} Q(f_h)\phi\,dv \tag{4.2.3}$$
$$= \int\limits_{\mathbb{R}^2}\int\limits_{\mathbb{R}^2}\int\limits_{S^1} B(v,w,e')[f_h(t,x,v')f_h(t,x,w') - f_h(t,x,v)f_h(t,x,w)]\phi(x,v)\,dv$$
$$= \overline{T}_K^{2+\frac{\beta}{2}} \int\limits_{\mathbb{R}^2}\int\limits_{\mathbb{R}^2}\int\limits_{S^1} b_r(|v-w|)b_\theta(\tfrac{(v-w)\cdot e'}{|v-w|})\times$$

$$[f_h^{0,1}(t,x,v')f_h^{0,1}(t,x,w') - f_h^{0,1}(t,x,v)f_h^{0,1}(t,x,w)]\phi^{0,1}(x,v)\,dv$$

$$= \overline{T}_K^{2+\frac{\beta}{2}} \int\limits_{\mathbb{R}^2} Q(f_h^{0,1})\phi^{0,1}\,dv. \tag{4.2.4}$$

The additional power $\frac{\beta}{2}$ of $\overline{T}_K$ results from $b_r(|v-w|) = (\overline{T}_K)^{\frac{\beta}{2}} b_r(|\tilde{v}-\tilde{w}|)$.
With the expansion of our discrete solution in terms of the basis functions

$$f_h(t,x_0,v) = e^{-\left|\frac{v-\overline{V}_K}{\sqrt{\overline{T}_K}}\right|^2} \sum_{m=0}^{\mathrm{ndof}_v-1} \underbrace{\sum_{r=0}^{\mathrm{ndof}_x-1} c_{r,m} u_r(x_0)}_{:=c_m^{x_0}} L_m\left(\frac{v-\overline{V}_K}{\sqrt{\overline{T}_K}}\right) \tag{4.2.5}$$

we obtain for the normalized distribution

$$f_h^{0,1}(t,x_0,v) = e^{-|v|^2} \sum_{m=0}^{\mathrm{ndof}_v-1} c_m^{x_0} L_m(v). \tag{4.2.6}$$

By conservation of energy during a binary collision (i.e. $|v|^2+|w|^2 = |v'|^2+|w'|^2$) we get

$$\begin{aligned} f_h^{0,1}(t,x_0,v')f_h^{0,1}(t,x_0,w') &= e^{-|v|^2-|w|^2} \sum_{m,n=0}^{\mathrm{ndof}_v-1} c_m^{x_0} c_n^{x_0} L_m(v')L_n(w') \\ f_h^{0,1}(t,x_0,v)f_h^{0,1}(t,x_0,w) &= e^{-|v|^2-|w|^2} \sum_{m,n=0}^{\mathrm{ndof}_v-1} c_m^{x_0} c_n^{x_0} L_m(v)L_n(w). \end{aligned} \tag{4.2.7}$$

The post collision velocities enter only into the polynomial part of the product $f_h^{0,1}(t,x_0,v')f_h^{0,1}(t,x_0,w')$.
With the expansion in (4.2.5) we can evaluate $g$ (using $\phi(x,v) = u_r(x)L_k\left(\frac{v-\overline{V}_K}{\sqrt{\overline{T}_K}}\right)$ as test function ) via

$$\begin{aligned} g(x_{\mathrm{ip}}) = u_r(x_{\mathrm{ip}})\overline{T}_K^{2+\frac{\beta}{2}} \sum_{m,n} c_m^{x_{\mathrm{ip}}} c_n^{x_{\mathrm{ip}}} \times \int\limits_{\mathbb{R}^2}\int\limits_{\mathbb{R}^2}\int\limits_{S^1} b_r b_\theta e^{-|v|^2-|w|^2} \times \\ \left[L_m(v')L_n(w') - L_m(v)L_n(w)\right] L_k(v)\,de'\,dv\,dw. \end{aligned} \tag{4.2.8}$$

For the sake of simplicity in the above formula, the dependency of $b_r$ and $b_\theta$ on their arguments has been omitted. The evaluation of the collision integrals on the discrete level is independent from the current macroscopic velocity $\overline{V}_K$ of the element Maxwellian, the dependency on the temperature reduces to a simple multiplication with $\overline{T}_K^{2+\beta/2}$.

From (4.2.8) we see that the numerical work needed to evaluate $g$ is bounded by $\mathcal{O}(N^6)$ operations. To this end we note that the double sum is of length $N^4$ and has to be evaluated for $N^2$ different values of $k$. In addition we see that the evaluation of $g$ corresponds to the evaluation of a third order tensor.

## 4.3 Efficient algorithm for the collision operator

To calculate the collision integrals in an efficient way we now present the techniques introduced in [KS13, KS15]. The collision operator acts local in position and time, but is a global operator in velocity. Thus, the $t$ and $x$ dependency of $f_h$ are omit for the rest of the section since they only act as parameters in the collision integral. Moreover, we also drop the subscript $h$ from the discrete solution. Due to (4.2.4) the following considerations can be restricted to the case $\overline{V}_K = 0$, $\overline{T}_K = 1$, the general case is then obtained by multiplying the collision result with $\overline{T}_K^{2+\beta/2}$.

### 4.3.1 Approximation on shifted grids

At the moment we have the approximation of the distribution function associated with the grid defined by the Gauss Hermite nodes. In a first step we now shift this grid according to the mean velocity of the collision partners and re approximate the distribution on the shifted grid. To come to that end, we start with the first representation of the collision operator in (2.2.3) and substitute $\bar{v} := \frac{v+w}{2}$, $\hat{v} := \frac{v-w}{2}$.

$$\begin{aligned}&\int\limits_{\mathbb{R}^2} Q(f(v))\phi(v)\,dv\\&= 4\int\limits_{\mathbb{R}^2}\int\limits_{\mathbb{R}^2}\int\limits_{S^1} b_r(|\hat{v}|)b_\theta(\tfrac{\hat{v}\cdot e'}{|\hat{v}|})f(\bar{v}+\hat{v})f(\bar{v}-\hat{v})[\phi(\bar{v}+e'|\hat{v}|)-\phi(\bar{v}+\hat{v})]\,de'd\hat{v}\,d\bar{v}.\end{aligned} \tag{4.3.1}$$

Throughout the integrand, the argument of $f$ and also of the test functions is of the form $\bar{v}$ plus or minus something. We interpret this as a shift of the coordinate origin. Thus, we approximate $f$ not on the initial grid associated with the Lagrange polynomials, but on the grid shifted by the mean velocity $\bar{v}$. We define approximations on the shifted grid, denoted by $f^{\bar{v}}(\hat{v}) := f(\bar{v}+\hat{v})$. For the collision integral we obtain

$$\begin{aligned}&\int\limits_{\mathbb{R}^2} Q(f(v))\phi(v)\,dv\\ &= 4\int\limits_{\mathbb{R}^2}\int\limits_{\mathbb{R}^2}\int\limits_{S^1} b_r(|\hat{v}|)b_\theta(\tfrac{\hat{v}\cdot e'}{|\hat{v}|})f^{\bar{v}}(\hat{v})f^{\bar{v}}(-\hat{v})[\phi^{\bar{v}}(e'|\hat{v}|)-\phi^{\bar{v}}(\hat{v})]\,de'd\hat{v}\,d\bar{v}.\end{aligned} \tag{4.3.2}$$

Letting $f_2^{\bar{v}}(\hat{v}) := f^{\bar{v}}(\hat{v})f^{\bar{v}}(-\hat{v})$, we arrive at

$$\int\limits_{\mathbb{R}^2} Q(f(v))\phi(v)\,dv = 4\int\limits_{\mathbb{R}^2}\underbrace{\int\limits_{\mathbb{R}^2}\int\limits_{S^1} b_r(|\hat{v}|)b_\theta(\tfrac{\hat{v}\cdot e'}{|\hat{v}|})f_2^{\bar{v}}(\hat{v})[\phi^{\bar{v}}(e'|\hat{v}|)-\phi^{\bar{v}}(\hat{v})]\,de'd\hat{v}}_{:=Q^I(b_r f_2^{\bar{v}},\phi^{\bar{v}})(\bar{v})}\,d\bar{v}. \tag{4.3.3}$$

For a given function $f \in \widetilde{V}_{h,N}$, the approximation on the shifted grid $f^{\bar{v}}(\hat{v})$ can be calculated within $O(N^3)$ operations. This is achieved by using the tensor product structure in the trial space. We denote the expansion coefficients of $f$ by $c$, require

$$f(\bar{v}+\hat{v}) = e^{-|\bar{v}+\hat{v}|^2}\sum_{i=0}^{\mathrm{ndof}_v-1} c_i L_i(\bar{v}+\hat{v}) = e^{-|\bar{v}+\hat{v}|^2}\sum_{i=0}^{\mathrm{ndof}_v-1} d_i L_i(\hat{v}) = f^{\bar{v}}(\hat{v}) \tag{4.3.4}$$

and solve for the unknown coefficients $d$. By the use of the Gauss Hermite quadrature nodes, i.e. $\hat{v} = x_j^{(2)} = (x_r, x_s)$, $j = (N+1)r+s$, the previous equation turns into

$$d_j = \sum_{i=0}^{\mathrm{ndof}_v-1} c_i L_i(\bar{v}+x_j^{(2)}) \qquad j = 0\ldots\mathrm{ndof}_v-1, \tag{4.3.5}$$

or in more compact form

$$d = S^{\bar{v}}c,$$

with $S^{\bar{v}} \in \mathbb{R}^{\mathrm{ndof}_v\times\mathrm{ndof}_v}$, $S^{\bar{v}}_{j,i} = L_i(\bar{v}+x_j^{(2)})$. The calculation of the sum in (4.3.5) requires $\mathrm{ndof}_v$ operations, and needs to be evaluated for $\mathrm{ndof}_v$ different values of $j$, such that the calculation of the coefficient vector $d$ requires $\mathrm{ndof}_v^2 = \mathcal{O}(N^4)$ operations. Using the tensor product structure in the trial space and the Cartesian structure in the 2d quadrature nodes, $d$ can be calculated more efficient via

$$d^m_{r,s} = \sum_{u=0}^{N} l_u(v_x + x_r) \sum_{v=0}^{N} c^m_{u,v} l_v(v_y + x_s). \tag{4.3.6}$$

Here, the coefficient vectors $c$ and $d$ have been reshaped to matrices $c^m, d^m$ in $\mathbb{R}^{(N+1)\times(N+1)}$ with $c^m_{u,v} = c_{u(N+1)+v}$ and $d^m_{r,s} = d_{r(N+1)+s}$. In addition the same splitting has been applied to the multivariate Lagrange polynomials $L_i(v_x, v_y) = l_u(v_x) l_v(v_y)$, $(N+1)u + v = i$, as well as to the nodes $x^{(2)}_j$.
The inner sum is required for $N+1$ different values of $u$, $N+1$ different values of $s$ and the length of the sum is $N+1$, resulting in $\mathcal{O}(N^3)$ operations for evaluation. The outer sum is required for $N+1$ values of $r$ and $v$ and is also of length $N+1$, resulting again in $\mathcal{O}(N^3)$ operations for evaluation. In order to use optimized LAPACK [ABB⁺99] routines, we rewrite the above sums as matrix products. For this purpose we define the 1d shift matrices

$$S^{\bar{v}_x}_{i,j} = l_j\left(\bar{v}_x + (x_i)_x\right), \qquad i, j = 0 \dots N, \tag{4.3.7}$$

analogue for the $y$ direction. The coefficients $d^m$ result in

$$S^{\bar{v}_y} c^m (S^{\bar{v}_x})^T = d^m. \tag{4.3.8}$$

Here $A^T$ denotes the transpose of the matrix $A \in \mathbb{R}^{(N+1)\times(N+1)}$. The above factorization of the sum in (4.3.5) can be interpreted as a shift of the grid in $y$ direction first and a shift in the $x$ direction afterwards.

In terms of the trial space, the function $f_2^{\bar{v}}$ is a polynomial of degree $2N$, multiplied with a Maxwellian with temperature $\frac{1}{4}$:

$$\begin{aligned} f_2^{\bar{v}}(\hat{v}) &= e^{-|\bar{v}+\hat{v}|^2} e^{-|\bar{v}-\hat{v}|^2} \sum_{i=0}^{\mathrm{ndof}_v - 1} d_i L_i(\hat{v}) \sum_{i=0}^{\mathrm{ndof}_v - 1} d_i L_i(-\hat{v}) \\ &= e^{-2|\bar{v}|^2 - 2|\hat{v}|^2} P(\hat{v}), \qquad P \in P^{2N}(\mathbb{R}^2). \end{aligned} \tag{4.3.9}$$

We want to represent $P$ in the above equation with Lagrange collocation polynomials $\tilde{L} \in P^{2N}(\mathbb{R}^2)$ of degree $2N$. The polynomials $\tilde{L}$ are – as the polynomials $L$ – defined as the product of 1d Lagrange polynomials $\tilde{l}$, i.e. $\tilde{L}_m = \tilde{l}_u \tilde{l}_v, m = u(2N+1) + v$. The collocation nodes for the 1d polynomials $\tilde{l}$ are chosen as $\frac{\tilde{x}_i}{\sqrt{2}}, i = 0 \dots 2N$, where $(\tilde{\omega}_i, \tilde{x}_i)$, $i = 0 \dots 2N$ is the Gauss Hermite quadrature rule of length $2N+1$. For $f_2^{\bar{v}}$ we now write

$$f_2^{\bar{v}}(\hat{v}) = e^{-2|\hat{v}|^2 - 2|\bar{v}|^2} \sum_{i=0}^{\tilde{N}} e_i \tilde{L}_i(\hat{v}), \quad \tilde{N} = (2N+1)^2 - 1. \tag{4.3.10}$$

The coefficients can be obtained by evaluating the sums in (4.3.9) at the nodes $\frac{\tilde{x}_i}{\sqrt{2}}$. This would require $\mathcal{O}(N^4)$ operations.

To calculate the coefficients $e_i, i = 0 \dots \tilde{N}$ efficiently, the representation of the shifted function $f^{\bar{v}}$ requires a small modification. Instead of representing it by basis polynomials $L_j$, $j = 0 \dots \mathrm{ndof}_v - 1$ of degree $N$, it is better to approximate it already by the polynomials $\tilde{L}_j \in P^{2N}$, $j = 0 \dots \tilde{N}$. Thus, the requirement in (4.3.4) is replaced by

$$f(\bar{v} + \hat{v}) = e^{-|\bar{v}+\hat{v}|^2} \sum_{i=0}^{\mathrm{ndof}_v - 1} c_i L_i(\bar{v} + \hat{v}) = e^{-|\bar{v}+\hat{v}|^2} \sum_{i=0}^{\tilde{N}} d_i \tilde{L}_i(\hat{v}) = f^{\bar{v}}(\hat{v}). \tag{4.3.11}$$

The coefficients $d$ can be computed analogue to the previous presentation. We just exchange the collocation nodes $x_j$ from the quadrature rule of length $N+1$ with the nodes $\frac{\tilde{x}_j}{\sqrt{2}}$, where $\tilde{x}_j$ is from the quadrature rule $(\tilde{\omega}_j, \tilde{x}_j), j = 0 \dots 2N$ of length $2N+1$. The adapted 1d shift matrices result in

$$S_{i,j}^{\bar{v}_x} = l_j \left( \bar{v}_x + \frac{(\tilde{x}_i)_x}{\sqrt{2}} \right) \quad j = 0 \dots N, i = 0 \dots 2N. \tag{4.3.12}$$

The coefficients are calculated via

$$S^{\bar{v}_y} c^m (S^{\bar{v}_x})^T = d^m \in \mathbb{R}^{(2N+1)\times(2N+1)}. \tag{4.3.13}$$

With this modification the calculation of the shifted function is more expensive, but is still bounded by $\mathcal{O}(N^3)$ operations. The benefit of the modification is in the calculation of $f_2^{\bar{v}}$. Replacing the polynomials of order $N$ in (4.3.9) by the polynomials of order $2N$ and using the expansion (4.3.10) for $f_2^{\bar{v}}$ yields

$$\sum_{i=0}^{\tilde{N}} e_i \tilde{L}_i(\hat{v}) = \sum_{i=0}^{\tilde{N}} d_i \tilde{L}_i(\hat{v}) \sum_{i=0}^{\tilde{N}} d_i \tilde{L}_i(-\hat{v}). \tag{4.3.14}$$

Using $\hat{v} = \left(\frac{\tilde{x}_u}{\sqrt{2}}, \frac{\tilde{x}_v}{\sqrt{2}}\right)$ with $j = u(2N+1) + v$ gives

$$e_j = \underbrace{\sum_{i=0}^{\tilde{N}} d_i \tilde{L}_i(\left(\frac{\tilde{x}_u}{\sqrt{2}}, \frac{\tilde{x}_v}{\sqrt{2}}\right))}_{=d_j} \underbrace{\sum_{i=0}^{\tilde{N}} d_i \tilde{L}_i(-\left(\frac{\tilde{x}_u}{\sqrt{2}}, \frac{\tilde{x}_v}{\sqrt{2}}\right))}_{=d_{\tilde{N}-j}} = d_j d_{\tilde{N}-j}. \tag{4.3.15}$$

(4.3.15) shows that the coefficients $e$ representing the function $f_2^{\bar{v}}$ result in the multiplication of 2 coefficients of the shifted function. The second sum in the above equation simplifies to $d_{\tilde{N}-j}$ due to the symmetry of the Gauss Hermite quadrature nodes. From (4.3.15), the benefit of representing $f^{\bar{v}}$ by $\tilde{L}$ is obvious: The evaluation of $e_j$ requires the values of the shifted functions at the nodes $\frac{\tilde{x}_j}{\sqrt{2}}$ and $-\frac{\tilde{x}_j}{\sqrt{2}}$. These values simply result in the coefficients $d_j$ and $d_{\tilde{N}-j}$ when using the polynomials $\tilde{L}$ to approximate $f^{\bar{v}}$.

So far we arrived at

$$f_2^{\bar{v}}(\hat{v}) = e^{-2|\bar{v}|^2 - 2|\hat{v}|^2} \sum_{i=0}^{\tilde{N}} e_i \tilde{L}_i(\hat{v}), \qquad e_i = d_i d_{\tilde{N}-i} \tag{4.3.16}$$

within $\mathcal{O}(N^3)$ floating point operations.

**Shifting the test functions**

Our initial variational formulation consists of the Lagrange polynomials on the non shifted grid. Thus, after calculating the collision integral w.r.t. the test functions on the shifted grid $\phi^{\bar{v}}$ as in (4.3.3), we have to transfer the result back to the Lagrange polynomials on the non shifted grid as test functions. As we will see, this transformation corresponds – roughly speaking – to the transpose of the forward shift of $f$. On the shifted grid we define the Lagrange polynomials $L^{\bar{v}}$ via the nodes $\frac{x_j^{(2)}}{\sqrt{2}}, j = 0 \ldots \mathrm{ndof}_v - 1$, where the nodes $x_j^{(2)} = (x_u, x_v)$, $j = u(N+1) + v$ are obtained from the 1d Gauss Hermite formula of length $N+1$. Since we do not need to increase the polynomial order as in the calculation of $f_2^{(\bar{v})}$, order $N$ polynomials are sufficient at this point. We require

$$L_k(\bar{v} + \hat{v}) = \sum_{i=0}^{\mathrm{ndof}_v - 1} \varphi_i L_i^{\bar{v}}(\hat{v}), \tag{4.3.17}$$

and look for the unknown coefficients $\varphi_i$.
Using $\hat{v} = \frac{x_j^{(2)}}{\sqrt{2}} = \frac{(x_r, x_s)}{\sqrt{2}}$ gives

$$\varphi_j = L_k \left( \bar{v} + \frac{x_j^{(2)}}{\sqrt{2}} \right).$$

In compact form and simultaneously for all $k = 0 \dots \mathrm{ndof}_v$ we obtain

$$\begin{pmatrix} L_0(\bar{v} + \hat{v}) \\ \vdots \\ L_{\mathrm{ndof}_v - 1}(\bar{v} + \hat{v}) \end{pmatrix} = S^{\bar{v},\phi} \begin{pmatrix} L_0^{\bar{v}}(\hat{v}) \\ \vdots \\ L_{\mathrm{ndof}_v - 1}^{\bar{v}}(\hat{v}) \end{pmatrix}, \tag{4.3.18}$$

with $S^{\bar{v},\phi} \in \mathbb{R}^{\mathrm{ndof}_v \times \mathrm{ndof}_v}$ and $S_{i,j}^{\bar{v},\phi} = L_i(\bar{v} + \frac{x_j^{(2)}}{\sqrt{2}}), i, j = 0 \dots \mathrm{ndof}_v - 1$. To save operations, the sum respectively matrix vector multiplication can be factorized similar to the forward shifting of the solution function as was done in (4.3.6). If we use the factorization, the required number of operations to shift the test functions is bounded by $\mathcal{O}(N^3)$.

**Remark 4.3.1.** *The evaluation of $\tilde{f}_2^{\bar{v}} := b_r f_2^{\bar{v}}$ is done at this point of the calculations and is realized as a point-wise multiplication of the coefficients of $f_2^{\bar{v}}$ with the values of $b_r(|\hat{v}|)$ at the collocation nodes for $f_2^{\bar{v}}$. In the sequel, the tilde sign is removed and we denote by $f_2^{\bar{v}}$ the above mentioned function $\tilde{f}_2^{\bar{v}}$. Note that the calculation of $b_r f_2^{\bar{v}}$ is not exact, even if $b_r$ is a polynomial.*

### 4.3.2 The integral w.r.t. the mean velocity

Now we aim to evaluate the integration with respect to $\bar{v}$ by a 2d Gauss Hermite quadrature rule

$$\begin{aligned} \int\limits_{\mathbb{R}^2} Q(f)(v)\phi(v)\,dv &= 4 \int\limits_{\mathbb{R}^2} e^{-2|\bar{v}|^2} Q^I \left( e^{-2|\hat{v}|^2} \sum_{i=0}^{\tilde{N}} e_i \tilde{L}_i, S^{\bar{v},\phi} L(\hat{v}) \right) d\bar{v} \\ &= 2 \sum_{\mathrm{ip}=0}^{n_{\mathrm{ip}}-1} \omega_{\mathrm{ip}} S^{\frac{x_{\mathrm{ip}}}{\sqrt{2}},\phi} Q^I(f_2^{\frac{x_{\mathrm{ip}}}{\sqrt{2}}}, L)(\frac{x_{\mathrm{ip}}}{\sqrt{2}}). \end{aligned} \tag{4.3.19}$$

Here $(\omega_i, x_i)$, $i = 0 \dots n_{\mathrm{ip}} - 1$ denote the weights and nodes of the quadrature rule, for the sake of readability we have omit here the superscript $(2)$, which was used to distinguish between the one and two dimensional nodes and weights respectively. The additional scaling of the quadrature

nodes by $1/\sqrt{2}$ is to obtain a Maxwellian with temperature $1/2$ in the integral with respect to $\bar{v}$, since the Gauss-Hermite quadrature rule is well suited for such integrands.
Note that in actual computations the matrix $S^{\frac{x_{\mathrm{ip}}}{\sqrt{2}},\phi}$ used to shift the test functions is not the transposed of the matrix $S^{\frac{x_{\mathrm{ip}}}{\sqrt{2}}}$ used to shift the solution function due to different polynomial degrees. Thus, both matrices have to be calculated.

At this point, the computational effort is bounded by the number $n_{\mathrm{ip}}$ of integration points w.r.t. $\bar{v}$, multiplied with the computational effort for evaluating $Q^I(f)$. Since $Q^I$ is linear in $f_2^{\bar{v}}$, its evaluation can be written as a matrix-vector product $Ab$, where $A \in \mathbb{R}^{(N+1)^2 \times (2N+1)^2}$ and $b \in \mathbb{R}^{(2N+1)^2}$, resulting in $\mathcal{O}(N^4)$ operations. This gives a total complexity of $\mathcal{O}(N^4) n_{\mathrm{ip}}$ for the moment. In actual computations we use a 1d Gauss Hermite rule of length $N$ to construct the 2d formula. This yields $\mathcal{O}(N^2)$ integration points and thus, we still have costs of $\mathcal{O}(N^6)$ for the collision integrals. Therefore we now investigate the application of the inner collision operator $Q^I$ to reduce the effort by one power of $N$.

### 4.3.3 Hermite and Polar-Laguerre polynomial bases

The efficiency considerations for $Q^I$ are based on a basis transformation in the momentum space for $f_2^{\bar{v}}$. $Q^I$ shall be applied in a basis which is given in Polar coordinates. As we will see, in this basis $Q^I$ is diagonal and thus, cheap to apply. We are in the sequel going to define the basis functions. In order to transform from the nodal to the Polar basis efficiently, we introduce an additional basis spanned by the Hermite polynomials to reduce computational effort. The transformation will be of the form

$$\text{Lagrange} \to \text{Hermite} \to \text{Polar}. \tag{4.3.20}$$

By $H_n(v)$ we denote the $n$-th (scaled) $1d$- Hermite polynomial. These are orthonormal w.r.t. $\langle f, g\rangle = \int_{\mathbb{R}} e^{-v^2} f g \, dv$ [STW11]. In addition by $\mathcal{L}_n^\alpha$ we denote the $n$-th (scaled) generalized Laguerre polynomial. These are orthonormal w.r.t. $\langle f, g\rangle = \int_{\mathbb{R}^+} v^\alpha e^{-v} f g \, dv$ [STW11].
From the Hermite polynomials we construct a 2-dimensional hierarchical basis

$$\mathcal{H}_{n,m}(v) := H_n(v_x) H_{m-n}(v_y) \qquad n \leq m, \; m = 0 \dots 2N. \tag{4.3.21}$$

The Polar polynomials are defined via the Laguerre polynomials and trigonometric functions and are given by

$$\begin{array}{l}
\Psi^{\cos}_{j,k}(v) := \begin{cases} s_{j,k}\cos(2j\varphi)r^{2j}\mathcal{L}^{(2j)}_{\frac{k}{2}-j}(r^2), & k \in 2\mathbb{N},\ j = 0\dots{}^{k}\!/_{2} \\ s_{j,k}\cos((2j+1)\varphi)r^{2j+1}\mathcal{L}^{(2j+1)}_{\frac{k-1}{2}-j}(r^2), & k \in 2\mathbb{N}+1,\ j = 0\dots\lfloor{}^{k}\!/_{2}\rfloor \end{cases} \\
\quad\text{and} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad . \\
\Psi^{\sin}_{j,k}(v) := \begin{cases} s_{j,k}\sin(2j\varphi)r^{2j}\mathcal{L}^{(2j)}_{\frac{k}{2}-j}(r^2), & k \in 2\mathbb{N},\ j = 1\dots{}^{k}\!/_{2} \\ s_{j,k}\sin((2j+1)\varphi)r^{2j+1}\mathcal{L}^{(2j+1)}_{\frac{k-1}{2}-j}(r^2), & k \in 2\mathbb{N}+1,\ j = 0\dots\lfloor{}^{k}\!/_{2}\rfloor \end{cases}
\end{array} \tag{4.3.22}$$

Here, $(r, \varphi)$ are the Polar coordinates of the velocity $v$, $s_{j,k}$ is a normalization constant for the angular part, i.e. $s_{0,2k} = \sqrt{\frac{2}{\pi}}$, $s_{j,k} = \sqrt{\frac{1}{\pi}}$ in all other cases.
In the sequel we use the notations

$$\begin{aligned}
\mathbb{H}_N &:= \{\mathcal{H}_{n,m} : n \leq m, m \leq N\} \\
\mathbb{L}_N &:= \{\Psi^{\cos}_{j,k} : k \leq N, j \leq \lfloor\tfrac{k}{2}\rfloor\} \cup \{\Psi^{\sin}_{j,k} : k \leq N, I_{2\mathbb{N}}(k) \leq j \leq \lfloor\tfrac{k}{2}\rfloor\}
\end{aligned}$$

to denote the sets of the Hermite and Polar basis polynomials respectively. In the above equations, $I_{2\mathbb{N}}$ denotes the indicator function of the even numbers. As a result of the ongoing calculations we obtain that both, $\mathbb{L}_N$ and $\mathbb{H}_N$ form bases of $\operatorname{span}\{x^i y^j : i + j \leq N\}$. Therefore we will already use the term Hermite basis and Polar Laguerre basis.

### Properties of the Polar Laguerre basis functions

In the next lemmata we collect some useful properties of the recently introduced polynomials. The first lemma ensures that we are indeed transforming to a polynomial basis w.r.t. Cartesian coordinates.

**Lemma 4.3.2.** *The Polar Laguerre functions* $\Psi^{\cos/\sin}_{j,k} \in \mathbb{L}_k \setminus \mathbb{L}_{k-1}$ *are polynomials in Cartesian coordinates of total degree* $k$.

*Proof.* For the proof we use the expansion of the trigonometric functions

$$\begin{aligned}\cos(n\varphi) &= \sum_{i=0}^{\lfloor\frac{n}{2}\rfloor} \binom{n}{2i} \sin(\varphi)^{2i}\cos(\varphi)^{n-2i} \\ \sin(n\varphi) &= \sum_{i=0}^{\lfloor\frac{n-1}{2}\rfloor} \binom{n}{2i+1} \sin(\varphi)^{2i+1}\cos(\varphi)^{n-2i-1}.\end{aligned} \tag{4.3.23}$$

With the power series expansion for the trigonometric part we get for even $k$:

$$\begin{aligned}\Psi^{\cos}_{j,k} &= \sum_{i=0}^{\lfloor\frac{2j}{2}\rfloor} \binom{2j}{2i} \sin(\varphi)^{2i}\cos(\varphi)^{2j-2i}\mathcal{L}^{2j}_{\frac{k}{2}-j}(r^2)r^{2j} \\ &= \sum_{i=0}^{j} \binom{2j}{2i} \underbrace{\sin(\varphi)^{2i}r^{2i}}_{y^{2i}}\underbrace{\cos(\varphi)^{2j-2i}r^{2j-2i}}_{x^{2j-2i}}\mathcal{L}^{2j}_{\frac{k}{2}-j}(r^2) \\ &= \sum_{i=0}^{j} \binom{2j}{2i} y^{2i}x^{2j-2i}\mathcal{L}^{2j}_{\frac{k}{2}-j}(x^2+y^2) \\ \Psi^{\sin}_{j,k} &= \sum_{i=0}^{\lfloor\frac{2j-1}{2}\rfloor} \binom{2j}{2i+1} \sin(\varphi)^{2i+1}\cos(\varphi)^{2j-2i-1}\mathcal{L}^{2j}_{\frac{k}{2}-j}(r^2)r^{2j} \\ &= \sum_{i=0}^{j} \binom{2j}{2i+1} \underbrace{\sin(\varphi)^{2i+1}r^{2i+1}}_{y^{2i+1}}\underbrace{\cos(\varphi)^{2j-2i-1}r^{2j-2i-1}}_{x^{2j-2i-1}}\mathcal{L}^{2j}_{\frac{k}{2}-j}(r^2) \\ &= \sum_{i=0}^{j} \binom{2j}{2i+1} y^{2i+1}x^{2j-2i-1}\mathcal{L}^{2j}_{\frac{k}{2}-j}(x^2+y^2)\end{aligned} \tag{4.3.24}$$

The monomials $y^{2i}x^{2j-2i}$ and $y^{2i+1}x^{2j-2i-1}$ are of total degree $2j$. Multiplying with the Laguerre polynomial $\mathcal{L}^{(2j)}_{\frac{k}{2}-j}(x^2+y^2)$ which is a polynomial in Cartesian coordinates of total degree $2(\frac{k}{2}-j) = k-2j$ results in a polynomial of total degree $k$. For odd $k$ the proof is analogue. □

Now we want to state the essential benefit of the Polar Laguerre basis for our calculations. That is, the inner collision operator $Q^I$ is diagonal in the Polar Laguerre basis. This is stated in the next lemma.

**Lemma 4.3.3.** *Let* $S, S_0 \in \{\sin, \cos\}$ *and* $\Psi^{S}_{j,k}, \Psi^{S_0}_{j_0,k_0} \in \mathbb{L}_N$. *Then there holds* $Q^I(e^{-|.|^2}\Psi^{S}_{j,k}, \Psi^{S_0}_{j_0,k_0}) = \frac{1}{2} b_{k_0,j_0}\delta_{j,j_0}\delta_{k,k_0}\delta_{S,S_0}$. *The value of the constant* $b_{k_0,j_0}$ *is given by*

$$b_{k_0,j_0} = \begin{cases} \int\limits_0^{2\pi} b_\theta(\cos(\alpha))(\cos(2j_0\alpha) - 1)\, d\alpha & k_0 \in 2\mathbb{N} \\ \int\limits_0^{2\pi} b_\theta(\cos(\alpha))(\cos((2j_0+1)\alpha) - 1)\, d\alpha & k_0 \in 2\mathbb{N}+1 \end{cases}. \tag{4.3.25}$$

*Proof.* We start with the case $k,\ k_0 \in 2\mathbb{N}$. We transform the integration w.r.t. the relative velocity $\hat{v}$ to Polar coordinates yielding

$$\begin{aligned} Q^I(e^{-|.|^2}\Psi^{S}_{j,k}, \Psi^{S_0}_{j_0,k_0}) &\overset{\hat{v}=r\,e}{=} \int\limits_{\mathbb{R}^+}\int\limits_{S^1}\int\limits_{S^1} b_\theta(e\cdot e')e^{-r^2} r \Psi^{S}_{j,k}(r\,e) \\ &\quad \times \left[\Psi^{S_0}_{j_0,k_0}(r\,e') - \Psi^{S_0}_{j_0,k_0}(r\,e)\right] de\, de'\, dr. \end{aligned} \tag{4.3.26}$$

On the unit sphere we use the usual parametrization $e' = (\cos(\alpha'), \sin(\alpha'))^T$ and $e = (\cos(\alpha), \sin(\alpha))^T$ respectively. Thus, the inner product $e \cdot e'$ results in $e \cdot e' = \cos(\alpha - \alpha')$. Using the definition of the the Polar Laguerre polynomials, the integral for the inner collision operator results in

$$\begin{aligned} &Q^I(e^{-|.|^2}\Psi^{S}_{j,k}, \Psi^{S_0}_{j_0,k_0}) \\ &= \int\limits_{\mathbb{R}^+} e^{-r^2} r r^{2j_0} r^{2j} \mathcal{L}^{(2j)}_{\frac{k}{2}-j}(r^2)\mathcal{L}^{(2j_0)}_{\frac{k_0}{2}-j_0}(r^2)\, dr \times \\ &s_{j_0,k_0}s_{j,k}\int\limits_0^{2\pi}\int\limits_0^{2\pi} b_\theta(\cos(\alpha-\alpha'))S(2j\alpha)\left[S_0(2j_0\alpha') - S_0(2j_0\alpha)\right]\, d\alpha\, d\alpha'. \end{aligned} \tag{4.3.27}$$

In the next step we substitute $\alpha = \alpha^\Delta + \alpha'$ and consider the two innermost integrals. This yields

$$\begin{aligned} &\int\limits_0^{2\pi}\int\limits_0^{2\pi} b_\theta(\cos(\alpha-\alpha'))S(2j\alpha)\left[S_0(2j_0\alpha') - S_0(2j_0\alpha)\right]\, d\alpha\, d\alpha' \\ &= \int\limits_0^{2\pi}\int\limits_0^{2\pi} b_\theta(\cos(\alpha^\Delta))S(2j(\alpha^\Delta+\alpha'))\left[S_0(2j_0\alpha') - S_0(2j_0(\alpha^\Delta+\alpha'))\right]\, d\alpha^\Delta\, d\alpha'. \end{aligned} \tag{4.3.28}$$

Now we interchange the order of integration and investigate the $d\alpha'$ integral in (4.3.28). In the case $S = S_0$ we obtain

$$\int_0^{2\pi} S_0(2j(\alpha^\Delta + \alpha')) \left[S_0(2j_0\alpha') - S_0(2j_0(\alpha^\Delta + \alpha'))\right] \, d\alpha' = \delta_{j,j_0}\pi(\cos(2j_0\alpha^\Delta) - 1). \quad (4.3.29)$$

For different trigonometric functions there holds

$$\int_0^{2\pi} S(2j(\alpha^\Delta + \alpha')) \left[S_0(2j_0\alpha') - S_0(2j_0(\alpha^\Delta + \alpha'))\right] \, d\alpha' = \pm\delta_{j,j_0}\pi \sin(2j_0\alpha^\Delta). \quad (4.3.30)$$

In fact, if $S = \sin$ and $S_0 = \cos$, the integral in (4.3.30) evaluates to $\pi\delta_{j,j_0}\sin(2j_0\alpha^\Delta)$. In the vice versa case it evaluates to $-\pi\delta_{j,j_0}\sin(2j_0\alpha^\Delta)$. Both, (4.3.29) and (4.3.30) consist of a $\delta$ relation for the angular frequencies. As a consequence, $j$ and $j_0$ have to coincide for the $d\alpha$ integral to be non zero.

Next we carry out the integration w.r.t. $\alpha^\Delta$ for the case of different types of angular functions to obtain

$$\begin{aligned} &\int_0^{2\pi}\int_0^{2\pi} b_\theta(\cos(\alpha - \alpha'))S(2j\alpha) \left[S_0(2j_0\alpha') - S_0(2j_0\alpha)\right] \, d\alpha \, d\alpha' \\ &= \pm\delta_{j,j_0}\pi \int_0^{2\pi} b_\theta(\cos(\alpha^\Delta)) \sin(2j_0\alpha^\Delta) \, d\alpha^\Delta = 0. \end{aligned} \quad (4.3.31)$$

The last equal sign in (4.3.31) is due to symmetries of the trigonometric functions, the integrand $f$ satisfies $f(\pi + \alpha^\Delta) = -f(\pi - \alpha^\Delta)$. By the last equality we obtain for the angular integrals

$$\begin{aligned} &\int_0^{2\pi}\int_0^{2\pi} b_\theta(\cos(\alpha - \alpha'))S(2j\alpha) \left[S_0(2j_0\alpha') - S_0(2j_0\alpha)\right] \, d\alpha \, d\alpha' \\ &= \delta_{j,j_0}\delta_{S,S_0}\pi \underbrace{\int_0^{2\pi} b_\theta(\cos(\alpha^\Delta))(\cos(2j_0\alpha^\Delta) - 1) \, d\alpha^\Delta}_{:=b_{k_0,j_0}}. \end{aligned} \quad (4.3.32)$$

Plugging (4.3.32) into (4.3.27), the inner collision operator results in

$$\begin{aligned}
Q^I(e^{-|.|^2\Psi^S_{j,k},\Psi^{S_0}_{j_0,k_0}}) &= \delta_{j,j_0}\delta_{S,S_0}b_{k_0,j_0}\int\limits_{\mathbb{R}^+} e^{-r^2} r r^{4j_0}\mathcal{L}^{(2j_0)}_{\frac{k}{2}-j_0}(r^2)\mathcal{L}^{(2j_0)}_{\frac{k_0}{2}-j_0}(r^2)\,dr \\
&\overset{r=\sqrt{\tilde{r}}}{=} \frac{1}{2}\delta_{j,j_0}\delta_{S,S_0}b_{k_0,j_0}\underbrace{\int\limits_{\mathbb{R}^+} e^{-r} r^{2j_0}\mathcal{L}^{(2j_0)}_{\frac{k}{2}-j_0}(r)\mathcal{L}^{(2j_0)}_{\frac{k_0}{2}-j_0}(r)\,dr}_{=\delta_{k,k_0}} \\
&= \frac{1}{2}\delta_{j,j_0}\delta_{S,S_0}\delta_{k,k_0}b_{k_0,j_0}.
\end{aligned} \tag{4.3.33}$$

We have presented the proof for the case $k, k_0 \in 2\mathbb{N}$, the case $k, k_0 \in 2\mathbb{N}+1$ is obtained in the same way and differs only in the indices of the Laguerre polynomials and the order of the monomial in the radial integral. In addition, the above proof shows that for $k \in 2\mathbb{N}$ and $k_0 \in 2\mathbb{N}+1$ or vice versa, one obtains a $\delta-$relation between an even and an odd number for the $d\alpha$ integral. Thus, in this case we also obtain 0 for $Q^I(e^{-|.|^2}\Psi^S_{j,k}, \Psi^{S_0}_{j_0,k_0})$. $\square$

Lemma 4.3.3 provides us with a polynomial basis in which the inner collision operator is very cheap to apply. Unfortunately, the distribution function $f_2^{\bar{v}}$ and the test functions are represented by nodal functions. To use the sparse representation of the inner collision integrals, $f_2^{\bar{v}}$ needs to be transferred to the Polar Laguerre basis. A direct transformation would introduce a new bottle neck. To have the transformations efficiently, the Hermite basis will be of great advantage since it has properties of both spaces, the nodal space $V_{2N}$ and also of the hierarchical space spanned by the Polar Laguerre polynomials. The tensor product structure on one hand provides again the ability to factorise the transformations from the nodal to the Hermite space. The second important property is stated in the next lemma and deals with the orthogonality of the Hermite and Polar Laguerre polynomials w.r.t. the Maxwellian weighted $L_2$ inner product on $\mathbb{R}^2$. This property will be a key element when transforming from the Hermite space to the Polar Laguerre space, since it gives structured and sparse transformation matrices.

**Lemma 4.3.4.** *The above introduced polynomial bases $\mathbb{H}_{2N}$ and $\mathbb{L}_{2N}$ are orthogonal w.r.t. the weighted $L_2$-inner product $\langle f, g\rangle = \int_{\mathbb{R}^2} e^{-|v|^2} fg\,dv$.*

*Proof.* The calculations are straightforward for the Hermite basis.

$$\begin{aligned}
&\int\limits_{\mathbb{R}^2} e^{-|v|^2}\mathcal{H}_{n,m}(v)\mathcal{H}_{n_0,m_0}(v)\,dv \\
&= \int\limits_{\mathbb{R}} e^{-v_x^2}H_n(v_x)H_{n_0}(v_x)\,dv_x \int\limits_{\mathbb{R}} e^{-v_y^2}H_{m-n}(v_y)H_{m_0-n_0}(v_y)\,dv_y = \delta_{m,m_0}\delta_{n,n_0}.
\end{aligned} \tag{4.3.34}$$

For the Polar Laguerre basis the situation is a little more complex. Transforming the integral to Polar coordinates, the calculations are quite similar to the proof of lemma 4.3.3. Again, we present it exemplary for $k, k_0 \in 2\mathbb{N}$.

$$\begin{aligned}
&\int\limits_{\mathbb{R}^2} e^{-|v|^2} \Psi^{\alpha}_{j,k}(v) \Psi^{\beta}_{j_0,k_0}(v)\, dv \\
&= \int\limits_{\mathbb{R}^+} \int\limits_{S^1} r e^{-r^2} \Psi^{\alpha}_{j,k}(re) \Psi^{\beta}_{j_0,k_0}(re)\, de\, dr \\
&= \delta_{\alpha,\beta} \delta_{j,j_0} \int\limits_{\mathbb{R}^+} r e^{-r^2} r^{4j_0} \mathcal{L}^{(2j_0)}_{\frac{k}{2}-j_0}(r^2) \mathcal{L}^{(2j_0)}_{\frac{k_0}{2}-j_0}(r^2)\, dr \\
&= \delta_{\alpha,\beta} \delta_{j,j_0} \delta_{k,k_0}.
\end{aligned} \tag{4.3.35}$$

If $k, k_0$ are both odd the proof is analogue. If $k$ is even and $k_0$ is odd or vice versa, the angular integrals already evaluate to zero. ☐

From the orthogonality of the Hermite and the Polar Laguerre polynomials we additionally obtain their linear independence. Counting the functions in $\mathbb{L}_{2N}$ and $\mathbb{H}_{2N}$ we see that both sets form a basis of the space $\mathrm{span}\{x^i y^j : i + j \leq 2N\}$.

**Remark 4.3.5.** *Obviously the polynomial spaces for $f_2^{\bar{v}}$, $V_{2N}$=$\mathrm{span}\{x^i y^j,\ i, j = 0 \ldots 2N\}$ and $\mathbb{L}_{2N}$ do not coincide (The first one is of partial order $2N$, the latter is of total order $2N$). At least total order $4N$ is necessary to represent $f_2^{\bar{v}}$ exact in the Polar Laguerre basis, $2N$ is sufficient for the test functions. Thus, we choose a Polar Laguerre test space of order $2N$. On the other hand, the Polar Laguerre polynomials are hierarchical and orthogonal. Thus, if we project $f_2^{\bar{v}}$ in its nodal representation onto $\mathbb{L}_{2N}$ and $\mathbb{L}_{4N}$ we obtain the same coefficients up to total order $2N$. Using lemma 4.3.3, the collision integral vanishes for all Polar Laguerre trial functions of total order greater than $2N$. As a consequence, it is sufficient to project $f_2^{\bar{v}}$ onto $\mathbb{L}_{2N}$ to obtain the exact contribution of the inner collision operator applied to $f_2^{\bar{v}}$. Figure 4.3.1 presents a sketch of the different representations of $f_2^{\bar{v}}$.*

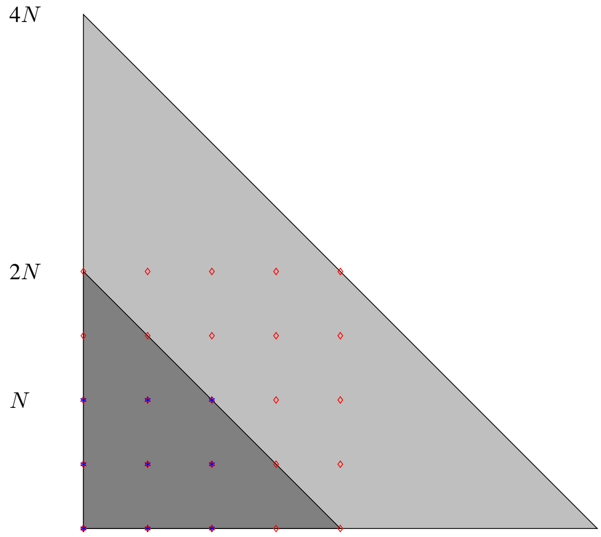


Figure 4.3.1: The markers are the polynomial coefficients in $v_x$ respectively $v_y$ direction. The blue ones represent the monomials in $V_N$, the red ones the monomials in $V_{2N}$. The gray shaded domains represent the hierarchical bases. In the lighter one we have an exact representation of $f_2^{\bar{v}}$ in the spaces $\mathbb{L}_{4N}$ and $\mathbb{H}_{4N}$ respectively. The darker one represents $P_{2N}(f_2^{\bar{v}})$, with $P_{2N}$ denoting the projection onto the space $\mathbb{L}_{2N}$.

### 4.3.4 Efficient transformation between polynomial bases

As has already been stated, the Hermite basis is of great use when transforming from our nodal representation to the Polar Laguerre polynomials. The transformation we execute consists of two steps. First we project $f_2^{\bar{v}}$ onto the Hermite space $\mathbb{H}_{2N}$. Then we transform the result to the Polar Laguerre basis $\mathbb{L}_{2N}$. In the sequel we show how to execute this transformations efficiently.

#### Transformation to Hermite

We start with the transformation from the Lagrange basis to the Hermite basis. Let $f_2^{\bar{v}} = e^{-2|v|^2} \sum_{m=0}^{\tilde{N}} c_m \tilde{L}_m(v)$, $\tilde{m} = (2N+1)^2 - 1$. $\tilde{L}_m = \tilde{l}_u \tilde{l}_v, m = u(2N+1) + v$ is the Lagrange polynomial of degree $2N$. The 1d polynomials $\tilde{l}$ are defined via the scaled Gauss Hermite quadrature nodes $\frac{\tilde{x}_j}{\sqrt{2}}, j = 0 \ldots 2N$ of the quadrature formula of length $2N+1$. We perform an $L_2$ orthogonal projection of $f_2^{\bar{v}}$ on the Hermite space $\mathbb{H}_{2N}$. This results in the following requirement for the Hermite coefficients.

$$\sum_{m=0}^{\tilde{N}} c_m \int\limits_{\mathbb{R}^2} e^{-2|v|^2} \tilde{L}_m(v) \mathcal{H}_{j_0,k_0}(\sqrt{2}v)\, dv = \sum_{\substack{k=0 \\ j\leq k}}^{2N} h_{j,k} \int\limits_{\mathbb{R}^2} e^{-2|v|^2} \mathcal{H}_{j,k}(\sqrt{2}v) \mathcal{H}_{j_0,k_0}(\sqrt{2}v)\, dv. \tag{4.3.36}$$

The additional scaling of the Hermite polynomials argument is due to the temperature $\frac{1}{4}$ of the Maxwellian weighting factor for $f_2^{\bar{v}}$ and is necessary to keep the orthogonality properties of the Hermite polynomials. The right hand side of the above equation turns due to orthogonality into

$$\sum_{\substack{k=0 \\ j\leq k}}^{2N} h_{j,k} \int\limits_{\mathbb{R}^2} e^{-2|v|^2} \mathcal{H}_{j,k}(\sqrt{2}v) \mathcal{H}_{j_0,k_0}(\sqrt{2}v)\, dv = \frac{1}{2} h_{j_0,k_0}. \tag{4.3.37}$$

A straight forward evaluation of the left hand side of (4.3.36) for each $0 \leq j \leq k$ and $0 \leq k \leq 2N$ requires $\mathcal{O}(N^4)$ operations. To achieve the optimal floating point operations, a similar factorization as for the shifting can be applied. As for the shifting, the factorization enables us to use efficient LAPACK [ABB+99] routines to calculate the resulting matrix products. The factorization – corresponding to a transformation in $v_x$ and $v_y$ direction separately – can be written as

$$\begin{aligned} \sum_{m=0}^{\tilde{N}} c_m \int\limits_{\mathbb{R}^2} e^{-2|v|^2} \tilde{L}_m(v) \mathcal{H}_{j_0,k_0}(\sqrt{2}v)\, dv &= \sum_{u=0}^{2N} \sum_{v=0}^{2N} c_{u(2N+1)+v} \\ \times \int\limits_{\mathbb{R}} e^{-2|v_x|^2} \tilde{l}_u(v_x) H_{j_0}(\sqrt{2}v_x)\, dv_x &\int\limits_{\mathbb{R}} e^{-2|v_y|^2} \tilde{l}_v(v_y) H_{k_0-j_0}(\sqrt{2}v_y)\, dv_y. \end{aligned} \tag{4.3.38}$$

Again as for the shifting we store the coefficient vector as a $(2N+1) \times (2N+1)$ matrix denoted by $c^m$, with $c^m_{u,v} = c_{u(2N+1)+v}$. In addition we define the 1d projection matrix $\mathrm{N2H} \in \mathbb{R}^{(2N+1)\times(2N+1)}$ via $\mathrm{N2H}_{u,v} := \int\limits_{\mathbb{R}} e^{-2|v|^2} \tilde{l}_u(v) H_v(\sqrt{2}v)\, dv,\ u,v = 0 \ldots 2N$ to obtain

$$\begin{aligned} \frac{1}{2} h_{j_0,k_0} = \sum_{m=0}^{\tilde{N}} c_m \int\limits_{\mathbb{R}^2} & e^{-2|v|^2} \tilde{L}_m(v) \mathcal{H}_{j_0,k_0}(\sqrt{2}v)\, dv \\ &= \sum_{u=0}^{2N} \sum_{v=0}^{2N} c^m_{u,v} \mathrm{N2H}_{u,j_0} \mathrm{N2H}_{v,k_0-j_0} \\ &= \left(\mathrm{N2H}^T c^m\, \mathrm{N2H}\right)_{j_0,k_0-j_0}. \end{aligned} \tag{4.3.39}$$

The calculation of the matrix products in (4.3.39) is bounded by $\mathcal{O}(N^3)$ operations in total. The integrals defining the entries of the matrix N2H can be evaluated by the Gauss Hermite quadrature rule of length $2N+1$ resulting in

$$\begin{aligned} \mathrm{N2H}_{i,j} &= \int\limits_{\mathbb{R}} e^{-2|v|^2} \tilde{l}_i(v) H_j(\sqrt{2}v)\, dv = \frac{1}{\sqrt{2}} \int\limits_{\mathbb{R}} e^{-|v|^2} \tilde{l}_i(\frac{v}{\sqrt{2}}) H_j(v)\, dv \\ &= \frac{1}{\sqrt{2}} \sum_{\mathrm{ip}=0}^{2N} \tilde{\omega}_{\mathrm{ip}} \tilde{l}_i(\frac{\tilde{x}_{\mathrm{ip}}}{\sqrt{2}}) H_j(\tilde{x}_{\mathrm{ip}}) = \frac{1}{\sqrt{2}} \tilde{\omega}_i H_j(\tilde{x}_i). \end{aligned} \tag{4.3.40}$$

### Transformation to Polar Laguerre

The transformation from the Hermite to the Polar Laguerre basis is also applied in the sense of an orthogonal projection. Assume

$$f_2^{\bar{v}}(v) = e^{-2|v|^2} \sum_{\substack{m \le 2N \\ n \le m}} h_{n,m} \mathcal{H}_{n,m}(\sqrt{2}v) \tag{4.3.41}$$

is given in the Hermite basis. With the Ansatz

$$f_2^{\bar{v}}(v) = \sum_{\substack{k \le 2N \\ j \le \lfloor k/2 \rfloor \\ a \in \{\sin, \cos\}}} \psi_{j,k,a} e^{-2|v|^2} \Psi^a_{j,k}(\sqrt{2}v) \tag{4.3.42}$$

we require

$$\begin{aligned} &\sum_{\substack{m \le 2N \\ n \le m}} h_{n,m} \int\limits_{\mathbb{R}^2} e^{-2|v|^2} \mathcal{H}_{n,m}(\sqrt{2}v) \Psi^b_{j_0,k_0}(\sqrt{2}v)\, dv \\ =&\sum_{\substack{k \le 2N \\ j \le \lfloor k/2 \rfloor \\ a \in \{\sin, \cos\}}} \psi_{j,k,a} \int\limits_{\mathbb{R}^2} e^{-2|v|^2} \Psi^a_{j,k}(\sqrt{2}v) \Psi^b_{j_0,k_0}(\sqrt{2}v)\, dv, \end{aligned} \tag{4.3.43}$$

to be satisfied for all Polar Laguerre test functions $\Psi^b_{j_0,k_0} \in \mathbb{L}_{2N}$. We note that in the above expansion to the Polar polynomials for even $k$ and $a = \sin$ the value $j = 0$ does not arise in the sum. We keep this in mind, but omit this in the notation. Due to the orthogonality of the Polar

Laguerre polynomials, the right hand side of (4.3.43) turns into $\frac{1}{2}\psi_{j_0,k_0,b}$. On the left hand side we split the sum into

$$\begin{aligned}\sum_{\substack{m\le 2N\\ n\le m}} h_{n,m}\int\limits_{\mathbb{R}^2} e^{-2|v|^2}\mathcal{H}_{n,m}(\sqrt{2}v)\Psi^b_{j_0,k_0}(\sqrt{2}v)\,dv\\ =\underbrace{\sum_{\substack{m< k_0\\ n\le m}} h_{n,m}\int\limits_{\mathbb{R}^2} e^{-2|v|^2}\mathcal{H}_{n,m}(\sqrt{2}v)\Psi^b_{j_0,k_0}(\sqrt{2}v)\,dv}_{:=A}\\ +\underbrace{\sum_{\substack{m= k_0\\ n\le m}} h_{n,m}\int\limits_{\mathbb{R}^2} e^{-2|v|^2}\mathcal{H}_{n,m}(\sqrt{2}v)\Psi^b_{j_0,k_0}(\sqrt{2}v)\,dv}_{:=B}\\ +\underbrace{\sum_{\substack{m> k_0\\ n\le m}} h_{n,m}\int\limits_{\mathbb{R}^2} e^{-2|v|^2}\mathcal{H}_{n,m}(\sqrt{2}v)\Psi^b_{j_0,k_0}(\sqrt{2}v)\,dv\,.}_{:=C}\end{aligned}\tag{4.3.44}$$

For the investigation of $A$ we expand $\mathcal{H}_{n,m}$ to Polar Laguerre polynomials.

$$\begin{aligned}\sum_{\substack{m< k_0\\ n\le m}} h_{n,m}\int\limits_{\mathbb{R}^2} e^{-2|v|^2}\mathcal{H}_{n,m}(\sqrt{2}v)\Psi^b_{j_0,k_0}(\sqrt{2}v)\,dv =\\ \sum_{\substack{m< k_0\\ n\le m}} h_{n,m}\int\limits_{\mathbb{R}^2} e^{-2|v|^2}\sum_{\substack{k\le m\\ j\le \lfloor k/2\rfloor\\ a\in\{\sin,\cos\}}}\psi^{n,m}_{j,k,a}\Psi^a_{j,k}(\sqrt{2}v)\Psi^b_{j_0,k_0}(\sqrt{2}v)\,dv.\end{aligned}\tag{4.3.45}$$

By the orthogonality of the Polar Laguerre polynomials we conclude $A = 0$ from the above equation. For $C$ we expand $\Psi^b_{j_0,k_0}$ to Hermite polynomials. Again by orthogonality we obtain $C = 0$. Thus, the coefficients in the Polar Laguerre basis for a fixed total order $k$ depend only on those coefficients in the Hermite basis of the same total polynomial order. Summarizing the above calculations, the coefficients of order $k$ in the Polar Laguerre basis are given by

$$\frac{1}{2}\psi_{j,k,b} = \sum_{n\le k} h_{n,k}\int\limits_{\mathbb{R}^2} e^{-2|v|^2}\mathcal{H}_{n,k}(\sqrt{2}v)\Psi^b_{j,k}(\sqrt{2}v)\,dv \quad j\le\lfloor\frac{k}{2}\rfloor, b\in\{\sin,\cos\}.\tag{4.3.46}$$

For each total polynomial degree $k$, $k+1$ coefficients have to be calculated. Each of them requires the evaluation of the above sum which is of length $k+1$. This gives a total computational effort of

$\sum_{k=0}^{2N}(k+1)^2 = \mathcal{O}(N^3)$ for the calculation of $f_2^{\bar{v}}$ in the Polar Laguerre basis. Since the transformation is linear in the Hermite coefficients, it can be represented by a matrix-vector multiplication, with a block diagonal matrix H2P. The structure of H2P is depicted in Figure 4.3.2.

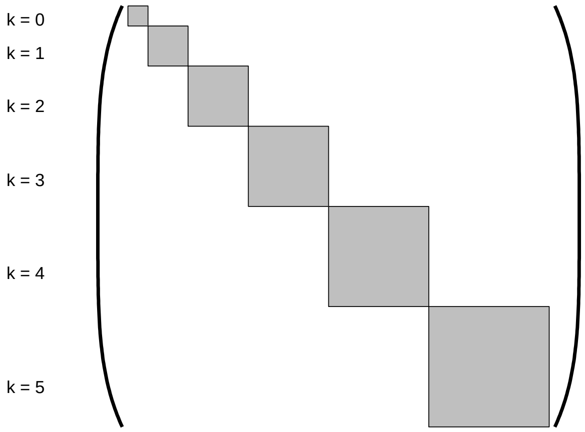


Figure 4.3.2: The structure of the transformation matrix from the Hermite to the Polar Laguerre basis, when sorting both bases hierarchical as in (4.3.48). The gray shaded blocks are the only non zero entries in the matrix. The $k$-th block is of size $(k+1)\times(k+1)$. The structure highlights the fact that coefficients in the Polar Laguerre basis of total order $k$ depend only on those coefficients in the Hermite basis of the same total order.

The $(i,j)$ entry of the $k$-th block of the matrix N2H is given by the weighted $L_2$ inner product of the $i$-th Polar Laguerre polynomial of total order $k$ and the $j$-th Hermite polynomial of total order $k$. With an enumeration as in (4.3.48), such a block results for even $k$ in

$$\mathrm{H2P}_k = \begin{pmatrix} \int e^{-2|v|^2}\mathcal{H}_{0,k}(\sqrt{2}v)\Psi^{\cos}_{0,k}(\sqrt{2}v) & \dots & \int e^{-2|v|^2}\mathcal{H}_{k,k}(\sqrt{2}v)\Psi^{\cos}_{0,k}(\sqrt{2}v) \\ \int e^{-2|v|^2}\mathcal{H}_{0,k}(\sqrt{2}v)\Psi^{\cos}_{1,k}(\sqrt{2}v) & \dots & \int e^{-2|v|^2}\mathcal{H}_{k,k}(\sqrt{2}v)\Psi^{\cos}_{1,k}(\sqrt{2}v) \\ \int e^{-2|v|^2}\mathcal{H}_{0,k}(\sqrt{2}v)\Psi^{\sin}_{1,k}(\sqrt{2}v) & \dots & \int e^{-2|v|^2}\mathcal{H}_{k,k}(\sqrt{2}v)\Psi^{\sin}_{1,k}(\sqrt{2}v) \\ \vdots & & \\ \int e^{-2|v|^2}\mathcal{H}_{0,k}(\sqrt{2}v)\Psi^{\cos}_{\frac{k}{2},k}(\sqrt{2}v) & \dots & \int e^{-2|v|^2}\mathcal{H}_{k,k}(\sqrt{2}v)\Psi^{\cos}_{\frac{k}{2},k}(\sqrt{2}v) \\ \int e^{-2|v|^2}\mathcal{H}_{0,k}(\sqrt{2}v)\Psi^{\sin}_{\frac{k}{2},k}(\sqrt{2}v) & \dots & \int e^{-2|v|^2}\mathcal{H}_{k,k}(\sqrt{2}v)\Psi^{\sin}_{\frac{k}{2},k}(\sqrt{2}v) \end{pmatrix}. \quad (4.3.47)$$

To arrive at the matrix structure depicted in Figure 4.3.2, a hierarchical enumeration of degrees of freedom is used in the Hermite and the Polar Laguerre basis

$$\mathcal{H} := \begin{pmatrix} \mathcal{H}_{0,0} \\ \mathcal{H}_{0,1} \\ \mathcal{H}_{1,1} \\ \vdots \\ \mathcal{H}_{0,2N} \\ \mathcal{H}_{1,2N} \\ \vdots \\ \mathcal{H}_{2N-1,2N} \\ \mathcal{H}_{2N,2N} \end{pmatrix} \qquad \Psi := \begin{pmatrix} \Psi^{\cos}_{0,0} \\ \Psi^{\cos}_{0,1} \\ \Psi^{\sin}_{0,1} \\ \vdots \\ \Psi^{\cos}_{0,2N} \\ \Psi^{\cos}_{1,2N} \\ \Psi^{\sin}_{1,2N} \\ \vdots \\ \Psi^{\cos}_{N,2N} \\ \Psi^{\sin}_{N,2N} \end{pmatrix}. \tag{4.3.48}$$

Note that the Hermite coefficients obtained in 4.3.39 need to be rearranged to fit the above enumeration.

### 4.3.5 Transformations of the test functions

So far, $f_2^{\bar{v}}$ is transferred to the Polar Laguerre basis efficiently and the collision integrals can be evaluated for the Polar Laguerre test polynomials. In the original variational formulation the test polynomials are the Lagrange polynomials of course. Thus, the shifting, Hermite transformation and finally the Polar Laguerre transformation have to be applied also to the test functions vice versa to end up with the Lagrange polynomials as test functions. Our strategy will therefore be to transform $f_2^{\bar{v}}$ to $\mathbb{L}_{2N}$, test with the Polar Laguerre basis polynomials, transform the result back to the Hermite polynomials as test functions, to the shifted Lagrange polynomials and finally invert the shifting of the test functions. These transformations of the test functions shall be discussed in the following. It turns out that the transformations acting on the test space are the transposed transformations as those for the function $f_2^{\bar{v}}$.

**Polar to Hermite**

We begin with transforming the result w.r.t. the Polar Laguerre test polynomials to the Hermite polynomials. Due to the linearity of the collision operator w.r.t. the test functions, our goal is simply to compute the expansion coefficients of a fixed Hermite Polynomial in the Polar polynomial expansion. By orthogonality one concludes that the Hermite polynomial $\mathcal{H}_{n,k}$ is a linear combination of the Polar Laguerre polynomials $\Psi^S_{j,k}, S \in \{\cos, \sin\}, j \leq \lfloor \frac{k}{2} \rfloor$ only. Thus, we are looking for coefficients $\psi^{n,k}_{j,S}$ such that

$$\mathcal{H}_{n,k}(\sqrt{2}v) = \sum_{\substack{j \leq \lfloor k/2 \rfloor \\ S \in \{\sin, \cos\}}} \psi^{n,k}_{j,S} \Psi^S_{j,k}(\sqrt{2}v) \tag{4.3.49}$$

holds. Now we test the above Ansatz with the Polar Laguerre polynomial $\Psi^{S_0}_{j_0,k}, j_0 \leq \lfloor k/2 \rfloor,\ S_0 \in \{\cos, \sin\}$ in $L_{2,\omega}(\mathbb{R}^2)$, where $\omega = e^{-2|\cdot|^2}$. Due to orthogonality we obtain from (4.3.49)

$$\int_{\mathbb{R}^2} e^{-2|v|^2} \mathcal{H}_{n,k}(\sqrt{2}v) \Psi^{S_0}_{j_0,k}(\sqrt{2}v)\, dv = \frac{1}{2} \psi^{n,k}_{j_0,S_0}. \tag{4.3.50}$$

The integral on the left hand side is an entry of the matrix $\mathrm{H2P}_k$, with column number $n$. The row number coincides with the number of the Polar Laguerre polynomial $\Psi^{S_0}_{j_0,k}$ in the hierarchical enumeration. If we denote this number by $r(k, j_0, S_0)$, we can write for the sum in (4.3.49)

$$\mathcal{H}_{n,k}(\sqrt{2}v) = 2 \sum_{\substack{j \leq \lfloor k/2 \rfloor \\ S \in \{\sin, \cos\}}} (\mathrm{H2P}_k)_{r(k,j,S),n} \Psi^S_{j,k}(\sqrt{2}v). \tag{4.3.51}$$

We note that this sum is the inner product in $\mathbb{R}^{k+1}$ of the $n$-th column of $\mathrm{H2P}_k$ and the vector consisting of the Polar polynomials of degree $k$. Thus, we calculate all Hermite polynomials simultaneously via

$$\begin{pmatrix} \mathcal{H}_{0,k} \\ \vdots \\ \mathcal{H}_{k,k} \end{pmatrix} = 2\, \mathrm{H2P}_k^T \begin{pmatrix} \Psi^{\cos}_{0,k} \\ \vdots \\ \Psi^{\sin}_{k/2,k} \end{pmatrix}. \tag{4.3.52}$$

Finally, by linearity of the inner collision operator w.r.t. the test polynomials we can write

$$Q^I(f_2^{\bar{v}}, \mathcal{H}) = 2\, \mathrm{H2P}^T Q^I(f_2^{\bar{v}}, \Psi). \tag{4.3.53}$$

The quantities $\mathcal{H}$ and $\Psi$ denote the complete set of Hermite and Polar Laguerre basis polynomials in the hierarchical enumeration in (4.3.48).

### Hermite to Lagrange

For the transformation from the Hermite to the Lagrange basis of the test polynomials the strategy is the same as before. We require

$$L_m(v) = \sum_{\substack{k=0\\ j\leq k}}^{2N} h^m_{j,k}\mathcal{H}_{j,k}(\sqrt{2}v), \quad m = 0\ldots \mathrm{ndof}_v - 1 \tag{4.3.54}$$

and look again for the coefficients $h^m_{j,k}$. In the above equation, the polynomials $L$ denote the Lagrange polynomials on the shifted grid. For the sake of readability we suppress the superscript $\bar{v}$ in the ongoing calculations. Similar as before, we test the Ansatz with the Hermite polynomial $\mathcal{H}_{j_0,k_0}(\sqrt{2}v)$ in $L_{2,\omega}(\mathbb{R}^2), \omega = e^{-2|.|^2}$. Using the orthogonality of the Hermite polynomials, (4.3.54) turns into

$$\int\limits_{\mathbb{R}^2} e^{-2|v|^2} L_m(v)\mathcal{H}_{j_0,k_0}(\sqrt{2}v)\, dv = \frac{1}{2}h^m_{j_0,k_0}. \tag{4.3.55}$$

The integral on the left hand side can be factorized as in the forward transformation, yielding

$$\int\limits_{\mathbb{R}} e^{-2v_x^2} l_u(v_x) H_{j_0}(\sqrt{2}v_x)\, dv_x \int\limits_{\mathbb{R}} e^{-2v_y^2} l_v(v_y) H_{k_0-j_0}(\sqrt{2}v_y)\, dv_y = \frac{1}{2}h^m_{j_0,k_0}, \tag{4.3.56}$$

with $m = u(N+1)+v$. Both integrals in the above equation are entries of a matrix similar to N2H, differing in the order of the Lagrange polynomial $L$ only. Therefore we define the matrix H2N $\in \mathbb{R}^{(N+1)\times(2N+1)}$ via

$$\mathrm{H2N}_{i,j} := \int\limits_{\mathbb{R}} e^{-2v^2} l_i(v) H_j(\sqrt{2}v)\, dv,$$

and rewrite the sum in (4.3.54) in terms of the matrix entries H2N as

$$L_m(v) = 2\sum_{k=0}^{2N}\sum_{j=0}^{k} \mathrm{H2N}_{u,j}\mathcal{H}_{j,k}(\sqrt{2}v)\mathrm{H2N}_{v,k-j} \quad \text{with } m = u(2N+1)+v. \tag{4.3.57}$$

For the further manipulation of the above sum it is convenient to sort the Hermite polynomials similar to a tensor product. To that end we define the matrix $H \in \mathbb{R}^{(2N+1)\times(2N+1)}$ via

$$H_{j,k} := \begin{cases} \mathcal{H}_{j,k+j}(\sqrt{2}v) & j+k \leq 2N \\ 0 & j+k > 2N \end{cases}.$$

In the matrix $H$ the Hermite Polynomials are sorted in the same structure as we obtained for the coefficients of $f_2^{\bar{v}}$ in the Hermite basis in (4.3.39). The entries $H_{i,j}$ with $i+j=k$ correspond to the $i-$th Hermite polynomial of total degree $k$ in the enumeration in (4.3.48). Note that by definition $H$ is an upper left triangular matrix.

We now replace $\mathcal{H}_{j,k}$ in (4.3.57) by the matrix entries $H_{j,k-j}$ and interchange the order of the summations to arrive at

$$\begin{aligned} L_m(v) &= 2\sum_{k=0}^{2N}\sum_{j=0}^{k} \mathrm{H2N}_{u,j} H_{j,k-j} \mathrm{H2N}_{v,k-j} \\ &= 2\sum_{j=0}^{2N}\sum_{k=j}^{2N} \mathrm{H2N}_{u,j} H_{j,k-j} \mathrm{H2N}_{v,k-j}. \end{aligned} \tag{4.3.58}$$

Note that the inner sum in the second line covers all values of $k-j \in \{0\dots 2N-j\}$ for which $H_{j,k-j}$ is non zero, such that we can write

$$\begin{aligned} L_m(v) &= 2\sum_{j=0}^{2N} \mathrm{H2N}_{u,j} (H\,\mathrm{H2N}^T)_{j,v} \\ &= 2\,(\mathrm{H2N}\,H\,\mathrm{H2N}^T)_{u,v}. \end{aligned} \tag{4.3.59}$$

Thus, the complete set of Lagrange polynomials is obtained via

$$L = 2\,\mathrm{H2N}\,H\,\mathrm{H2N}^T, \tag{4.3.60}$$

where $L$ denotes the set of all Lagrange basis polynomials sorted matrix wise, i.e. the $u, v$ entry of the matrix $L$ is the $m$-th Lagrange polynomial, where $m = u(N+1)+v$ and $u, v = 0\dots N$.

The transformations we obtained for the test functions show that the number of operations required to perform them is the same as the forward transformations of $f_2^{\bar{v}}$. Thus, we are now able to calculate the inner collision operator $Q^I(f_2^{\bar{v}}, L)$ within $\mathcal{O}(N^3)$ operations in total. Combining this result with the integration formular w.r.t. $\bar{v}$, we obtain a collision algorithm with $\mathcal{O}(N^5)$ complexity.

In order to illustrate our approach for the application of the collision integrals we present the following pseudo code. Due to a better readability we do not use a representation of the operations in terms of matrix products, but in terms of functions.

```
input : $(c_0 \dots c_{\text{ndof}_v - 1})$ representing $f(v)$
output: $q_j = \int_{\mathbb{R}^2} Q(f) L_j \, dv, \ j = 0 \dots \text{ndof}_v - 1$
$N = \sqrt{\text{ndof}_v} - 1$;
$(x, \omega) = \texttt{GaussHermiteRuleTensored}\,(N+1)$;
$q = 0$;
foreach $(x_i, \omega_i)$ do
    $d = \texttt{Shift}_{y_i}\,(\texttt{Shift}_{x_i}\,(c))$;
    for $j = 0 : \tilde{N} := (2N+1)^2 - 1$ do $e_j = d_j d_{\tilde{N}-j}$;
    $h = \texttt{Nodal2Hermite}\,(e)$;
    $p = \texttt{Hermite2Polar}\,(h)$;
    $p_{\text{coll}} = \texttt{DiagCollision}\,(p)$;
    $h_{\text{coll}} = \texttt{Hermite2PolarT}\,(p_{\text{coll}})$;
    $n_{\text{coll}} = \texttt{Hermite2Nodal}\,(h_{\text{coll}})$;
    $q += \omega_i \, \texttt{ShiftT}_{y_i}\,(\texttt{ShiftT}_{x_i}\,(n_{\text{coll}}))$;
end
```

**Algorithm 4:** a pseudo code for the collision integrals

**Remark 4.3.6.** *The presented expansion in terms of a Maxwellian multiplied with a polynomial has a close connection to the approach investigated in [FGH14]. In contrast to our approach, the solution is directly expanded to generalized Laguerre polynomials with a Maxwellian weighting factor. The efficiency considerations are based on the orthogonality of the trigonometric functions. The paper generalizes the approach from [EE99] to radially non symmetric solutions.*
*For the transport operator, similar approaches are presented in [DDCS12, HGMM12], where a Discontinuous Galerkin projection is applied to the Vlasov-Poisson System. In contrast to our method, local polynomials in space as well as in velocity are used.*

## 4.4 Adaptive choice of element Maxwellians

A crucial part of the scheme is the choice of the quantities $\overline{V}_K$ and $\overline{T}_K$ describing the element Maxwellian. For convenience we restate the macroscopic quantities density $\rho$, mean velocity $V$

and temperature $T$ for a 2d setting below.

$$\begin{aligned} \rho(t,x) &:= \int_{\mathbb{R}^2} f(t,x,v)\,dv \qquad V(t,x) := \frac{1}{\rho(t,x)} \int_{\mathbb{R}^2} v f(t,x,v)\,dv \\ T(t,x) &:= \frac{1}{2\rho(t,x)} \int_{\mathbb{R}^2} (v - V(t,x))^2 f(t,x,v)\,dv. \end{aligned} \tag{4.4.1}$$

We expect good approximation properties if the parameters $\overline{V}_K$ and $\overline{T}_K$ of the element Maxwellians are close to the macroscopic velocity $V(t,x)$ and temperature $T(t,x)$. We choose element wise constant $\overline{T} \in P^0(\mathcal{T}_h) := \{u \in L_2(\Omega) \,:\, u\big|_K \in P^0(K),\ \forall\, K \in \mathcal{T}_h\}$ and $\overline{V} \in [P^0(\mathcal{T}_h)]^2$. By $\{u\}_K$ we denote the mean value of a function $u \in L_1$, i.e. $\{u\}_K = \frac{1}{|K|}\int_K u(x)\,dx$, where $|K|$ is the volume of the element $K \in \mathcal{T}_h$. A simple requirement for the parameters of the element Maxwellian is

$$\overline{V}_K \equiv \{V(t,.)\}_K \quad \text{and} \quad \overline{T}_K \equiv \{T(t,.)\}_K \qquad \forall\, K \in \mathcal{T}_h. \tag{4.4.2}$$

In practice this choice is not very useful, since it turns out to be quite unstable, a heuristic reason can be obtained by the following considerations.

The definition of the upwind function incorporates $f\big|_K$ as well as $f\big|_{\tilde{K}}$, with $\tilde{K}$ being a neighbour element to $K$. In order to calculate the skeleton integrals by Gauss Hermite rules, the solution from the neighbouring element has to be projected to the element Maxwellian of the element $K$. This projection is not well defined for all pairs of Ansatz temperatures. To illustrate the problem let $L_{2,T} := \{f \in L_2(\mathbb{R}^2) \,:\, \int_{\mathbb{R}^2} e^{\frac{|v|^2}{T}} f(v)^2\,dv < \infty\}$ and consider a discrete solution $f(v) = e^{-\frac{|v|^2}{T_0}} P(v) \in L_{2,T_0}$, with $P \in P^N(\mathbb{R}^2)$ which shall be projected to $\tilde{f}(v) = e^{-\frac{|v|^2}{T_1}} \tilde{P}(v) \in L_{2,T_1}$, with $\tilde{P} \in P^N(\mathbb{R}^2)$. To have the orthogonal projection well defined, $f$ has to be in $L_{2,T_1}$. This yields

$$\int_{\mathbb{R}^2} e^{\frac{v^2}{T_1}} f(v)^2\,dv = \int_{\mathbb{R}^2} e^{v^2\left(\frac{1}{T_1} - \frac{2}{T_0}\right)} P(v)^2\,dv < \infty \Leftrightarrow T_0 < 2T_1. \tag{4.4.3}$$

Thus, the projection is well defined if $T_0 < T_1$. The other way round, a condition on the temperatures is necessary to have the $L_{2,T_1}$-projection well defined. In order to avoid this problem – which effects stability in actual computations – we bound $\frac{T_0}{T_1}$ by a constant $c \leq 2$. In addition we require $\overline{T}_K$ to be greater or equal than the mean of the macroscopic temperature in (4.4.1). Thus, we are looking for $\overline{T}$ as the minimizer of

$$\overline{T} := \operatorname*{argmin}_{s_K \in P^0(\mathcal{T}_h)} \sum_{K \in \mathcal{T}_h} |s_K - \{T\}_K|\,|K|, \tag{4.4.4}$$

under the constraints

$$\begin{aligned} &(1)\ s_K \geq \tfrac{1}{2} s_{K'},\ \forall K' \text{s.t.:} K' \text{ is neighbour of } K \\ &(2)\ s_K \geq \{T\}_K. \end{aligned}$$

In actual computations we initially set $s_K = \{T\}_K$. Now we loop over the elements of the mesh and update $s_K$ to satisfy the constraints. We iterate this loop until no update of any $s_K$ is necessary any more.

The choice of the velocity parameter $\overline{V}_K$ is motivated by the behaviour at the boundary. The boundary conditions presented in (1.2.4a) and (1.2.4b) yield a vanishing normal component of the macroscopic velocity. This is the natural boundary condition for functions in $H(\text{div}, \Omega) := \{u \in [L_2(\Omega)]^2 : \text{div}(u) \in L_2(\Omega)\}$, which is equipped with the norm $\|u\|^2_{H(\text{div},\Omega)} = \|u\|^2_{L_2(\Omega)} + \|\text{div}(u)\|^2_{L_2(\Omega)}$ [BF91].

In order to incorporate the behaviour at the boundary in the Ansatz velocity $\overline{V}$, we can interpolate $V(t,x)$ with a function $\overline{V}$ in $H(\text{div}, \Omega)$, such that $u \cdot n = 0$ if $x \in (\partial\Omega)_{\text{ref}}$ by using the Raviart Thomas interpolation operator $\mathcal{I}_K^{\text{RT}}$ (of lowest order). By $(\partial\Omega)_{\text{ref}}$ we denote the part of the boundary of $\Omega$ where reflection conditions, i.e. (1.2.4a) or (1.2.4b) are prescribed. However, by such an interpolation the condition $u \cdot n = 0$ affects the interpolant only in the elements which have at least one edge in $(\partial\Omega)_{\text{ref}}$. This can lead to large jumps in the Ansatz velocity, causing instabilities in actual computations.
In order to obtain a boundary layer of a prescribed thickness and therefore smaller jumps, we use the solution $u$ of the following coercive $H(\text{div}, \Omega)$ problem for the Ansatz velocity $\overline{V}$.

Find $u \in H(\text{div}, \Omega)$ with $u \cdot n = 0,\ x \in (\partial\Omega)_{\text{ref}}$ s.t.:

$$\begin{aligned} \sum_{K \in \mathcal{T}_h} \int_K u(x)\varphi(x)\,dx + \alpha^2 \sum_{K \in \mathcal{T}_h} \int_K \text{div}(u)(x)\text{div}(\varphi)(x)\,dx& \\ = \sum_{K \in \mathcal{T}_h} \int_K V(t,x)\varphi(x)\,dx \qquad \forall \varphi \in H(\text{div}, \Omega).& \end{aligned} \tag{4.4.5}$$

The value of the constant $\alpha$ is related to the thickness of the layer in which the boundary condition affects the solution. Thus, letting $\alpha^2 = Ch$, with $h$ being the mesh size, we can – roughly speaking – control on how many element layers the solution ”sees” the boundary condition. In other words we can use $\alpha^2$ to balance between ”$u = V$” and ”$u$ is sufficiently smooth to end up with a stable simulation”. This is quite similar to the boundary layers obtained when solving

Find $u \in H_0^1(0,1)$ s.t.:

$$\int_0^1 u(x)v(x)\,dx + \epsilon^2 \int_0^1 u'(x)v'(x)\,dx = \int_0^1 1v(x)\,dx \quad \forall v \in H_0^1(0,1). \tag{4.4.6}$$

The solution of the above problem has the constant value 1 corresponding to the right hand side. To satisfy the boundary condition, it drops on $(0,\epsilon)$ and $(1-\epsilon,1)$ towards 0. For small $\epsilon$ this yields large gradients.

In order to obtain $\overline{V}$, we solve (4.4.5) using the lowest order Raviart Thomas [RT77] element. Since this gives a solution $u_h$ which is piecewise linear, we project $u_h$ onto $[P^0(\mathcal{T}_h)]^2$ to arrive finally at $\overline{V}$.

# 5 Numerical results

In this section we present numerical results as a validation for our method. First we discuss a space homogeneous problem with a known analytic solution. The 2-dimensional examples involve two model problems of the flow around a wedge and the flow through a tube with a cylindrical hole. Then we apply the method to a NACA 7410 air foil with either specular and diffuse reflecting boundary conditions. Additionally we show a simulation result for the Mach 3 wind tunnel with backward facing step. All the above mentioned examples are simulated with a rather small Knudsen number. In order to demonstrate the performance of the method for larger Knudsen numbers we conclude the section with a simulation result for a Knudsen pump, where the Knudsen numbers $0.1$ and $0.7$ have been tested.

All calculations were performed on a machine with 2 Intel(R) Xeon(R) CPU E5-2670 v3 @ 2.30GHz processors with 12 cores each. The method is implemented within the Finite Element Library NgSolve [Sch, Sch14], featuring shared memory parallelization [DM98].

## 5.1 Space homogeneous problems

### 5.1.1 BKW Solution

The spatially homogeneous example we present is well known as the BKW solution [Kru67, Bob88, Ern84]. This is a non-stationary analytic solution of the spatially homogeneous Boltzmann equation $\frac{\partial f}{\partial t} = Q(f)$ for a Maxwellian gas with $B(v, w, e') = \frac{1}{2\pi}$. In 2 dimensions it is given by

$$f(t, v) = \frac{1}{2\pi s(t)} e^{-\frac{|v|^2}{2s(t)}} \left(1 - \frac{1-s(t)}{2s(t)} \left(2 - \frac{|v|^2}{s(t)}\right)\right), \tag{5.1.1}$$

with $s(t) = 1 - e^{-\frac{(t+t_0)}{8}}$.
We chose the starting time $t_0$ such that $s(0) = \frac{1}{2}$ and consequently $f(0, v) = \frac{1}{\pi}|v|^2 e^{-|v|^2}$. Since

$s \to 1$ if $t \to \infty$, the stationary solution is given by

$$f_\infty(v) = (2\pi)^{-1} e^{-\frac{|v|^2}{2}},$$

which is a Maxwellian with temperature 1, velocity 0 and density 1. Due to the conservation laws, $f_\infty$ can also be obtained by calculating density, momentum and energy of $f(t, .)$ for any arbitrary $t$ such that $f(t, .) \geq 0$, and then forming the Maxwellian corresponding to these macroscopic quantities.
For the simulation we have chosen the global element Maxwellian in accordance with the equilibrium solution. This yields an expansion in terms of

$$f_N(t, v) = e^{-\frac{|v|^2}{2}} \sum_{m=0}^{\mathrm{ndof}_v - 1} c_m L_m(\frac{v}{\sqrt{2}}). \tag{5.1.2}$$

In Figure 5.1.1 we show cross sections of the solution at different points in time. In Figure 5.1.2 we present the $L_\infty(0, T_{\text{end}})$ norm of the quantities $\|f_h(t, .) - f(t, .)\|_{L_j, \omega}, j = 1, 2, \infty$ with weight functions $\omega(v) = e^{0.5|v|^2}$ and $\omega(v) \equiv 1$ respectively. We see exponential convergence of the method in the momentum domain. The computation times for different polynomial orders are presented in table 5.1 and Figure 5.1.3. We note the expected $N^5$ asymptotic.

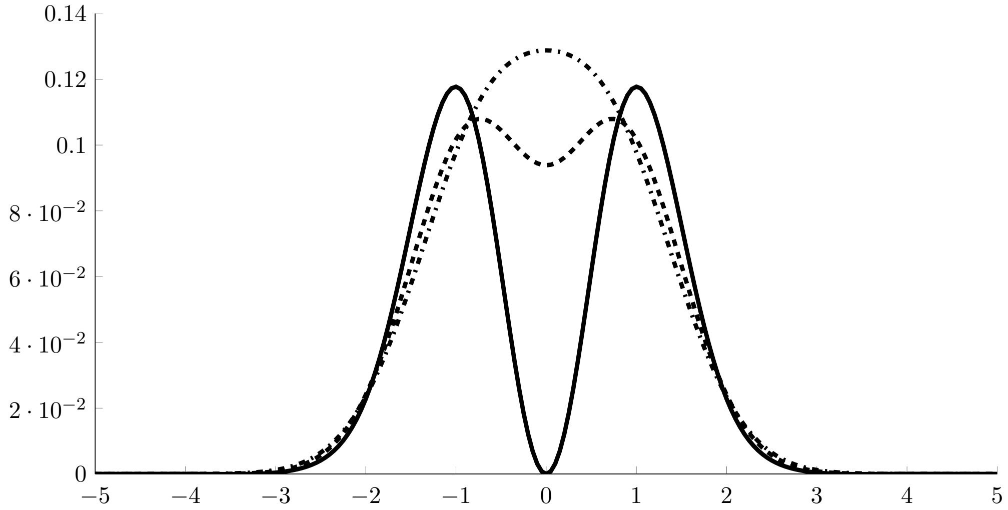


Figure 5.1.1: Snapshots of the distribution function for the BKW solution, obtained with an order 16 simulation. The solid line is at $t = 0$, the dashed line is at $t = 2$ and the dash-dotted line is at $t = 4$. Note that the solution is radially symmetric at any time $t$.

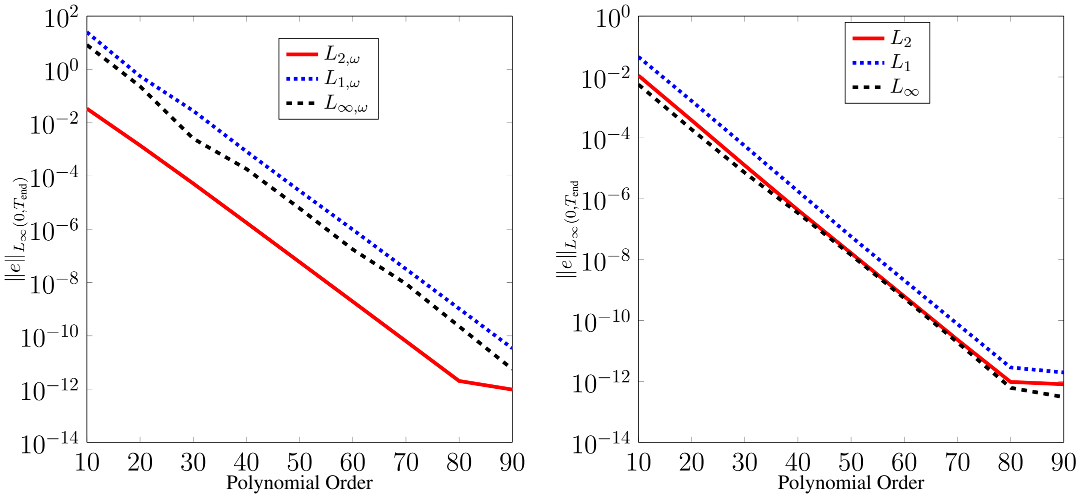


Figure 5.1.2: We present the $L_\infty$-error on $[0, T_{\text{end}}]$ of the quantity $e_N(t) := \|f_N(t) - f(t)\|_{L_j(\mathbb{R}^2),\omega}$, $j = 1, 2, \infty$ as a function of the expansion order $N$. By $\|.\|_{L_j(\mathbb{R}^2),\omega}$ we denote the $\omega$ weighted $L_j$ norms. The left Figure corresponds to $\omega(v) = e^{0.5|v|^2}$, on the right side the weight function is $\omega(v) \equiv 1$.

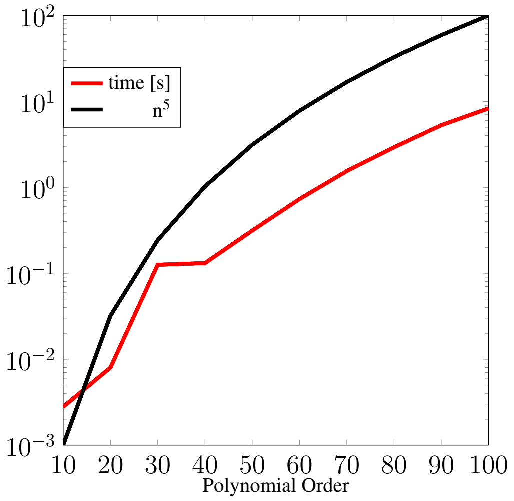


Figure 5.1.3: The computation times for the BKW solution. We present the time consumed for a single time step over the different polynomial orders. The reference line is the monomial $n^5$. The times are also listed in table 5.1.

| $n$ | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |
|---|---|---|---|---|---|---|---|---|---|---|
| $t$ | 0.003 | 0.008 | 0.126 | 0.131 | 0.314 | 0.731 | 1.548 | 2.926 | 5.288 | 8.289 |

Table 5.1: Computation times for the BKW solution. The numbers in the first line are the polynomial orders, the number below is the computation time per step in seconds.

## 5.2 Space dependent problems

For the space inhomogeneous problems with small Knudsen number we used $B(v, w, e') \equiv \frac{1}{2\pi}$. Due to the small Knudsen number we expect that the choice of the Kernel does not significantly influence the solution. For the Knudsen pump where the Knudsen numbers are chosen as $0.1$ and $0.7$, the hard sphere model, i.e. $B(v, w, e') = |v - w|$ was used.

We emphasize that quantitative accurate results require a three dimensional velocity space, what is not the case in our approach. As a consequence we mainly demonstrate what kind of problems we are able to solve by our method. The examples verify empirically that the method performs very well near the fluid dynamic limits.

### 5.2.1 Model problem 1 - Flow around a wedge

The first 2d problem is a model problem in which we consider the flow around a deltoid. The geometry and the mesh for the computations are depicted in Figure 5.2.1a. We have a supersonic free flow, resulting in a compression shock at the forefront of the deltoid. Due to the simple geometry, the angle of the shock as well as the macroscopic behaviour after the shock can be pre calculated. The relation between the angle of the compression $\beta$ and the deviation angle $\theta$ is given via

$$\tan(\theta) = 2\cot(\beta)\frac{M_\infty^2 \sin(\beta)^2 - 1}{M_\infty^2(\gamma + \cos(2\beta)) + 2}. \tag{5.2.1}$$

In the above equation, $M_\infty$ denotes the free flow Mach number and $\gamma$ is related to the degrees of freedom $d$ of the gas via $\gamma = \frac{d+2}{d}$. In our simulations $d = 2$ and consequently $\gamma = 2$. The angles $\theta$ and $\beta$ are both w.r.t. the direction of the free flow. The macroscopic properties $\rho$, $p$ and $T$ after the compression shock are expressed by

$$\frac{\rho}{\rho_\infty} = \frac{(\gamma+1)M_{n,\infty}^2}{2 + (\gamma - 1)M_{n,\infty}^2}, \qquad \frac{p}{p_\infty} = 1 + \frac{2\gamma}{\gamma+1}(M_{n,\infty}^2 - 1) \quad \text{and} \quad \frac{T}{T_\infty} = \frac{p\rho_\infty}{p_\infty\rho}, \tag{5.2.2}$$

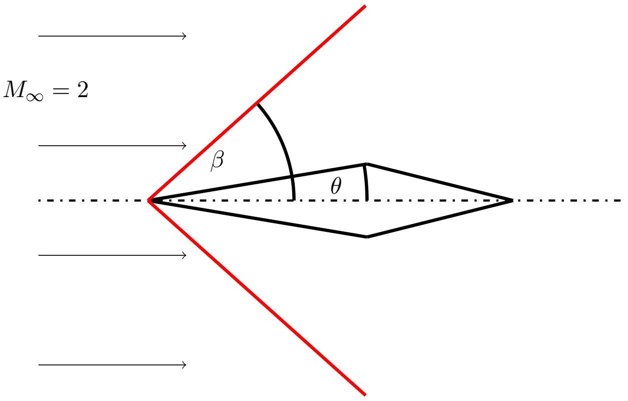
where $\rho_\infty$, $p_\infty$ and $T_\infty$ are the macroscopic properties of the free flow. The Mach numbers before and after the compression are related via the free flow Mach number in normal direction of the compression $M_{n,\infty} = \sin(\beta) M_\infty$ via

$$M = \frac{M_n}{\sin(\beta - \theta)} \quad \text{with} \quad M_n = \sqrt{\frac{1 + \frac{\gamma-1}{2} M_{n,\infty}^2}{\gamma M_{n,\infty}^2 - \frac{\gamma-1}{2}}}. \tag{5.2.3}$$

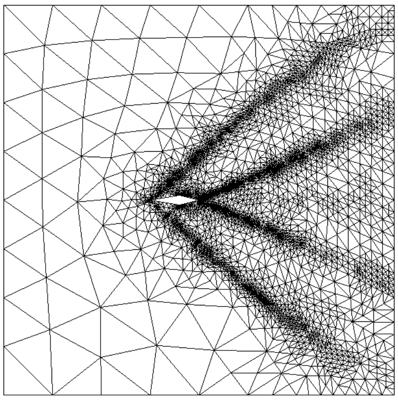
The simulation is performed for a free stream Mach 2 flow with $\rho_\infty = 1$, $T_\infty = 0.5$ and $V_\infty = (2, 0)^T$. The behaviour after the compression is pre-calculated by the above equations.

$$\rho = 1.4122, \qquad p = 1.0191, \qquad T = 0.7217, \qquad M = 1.4686. \tag{5.2.4}$$

The deviation angle is $\theta = 9.4623°$, resulting in a shock angle of $\beta = 41.8238°$.



(a) The geometry of the deltoid with deviation angle $\theta = 9.4623°$.



(b) The mesh at the 6th refinement level.

The simulation was performed with order 2 spatial elements and order 6 momentum elements. The mesh for the presented results consists of 8874 elements and was obtained by adaptive refinement at the shocks. The estimation of the error was based on the macroscopic Mach number at time $t = 3$ and was done as proposed by Zienkiewicz and Zhu [ZZ92a, ZZ92b].

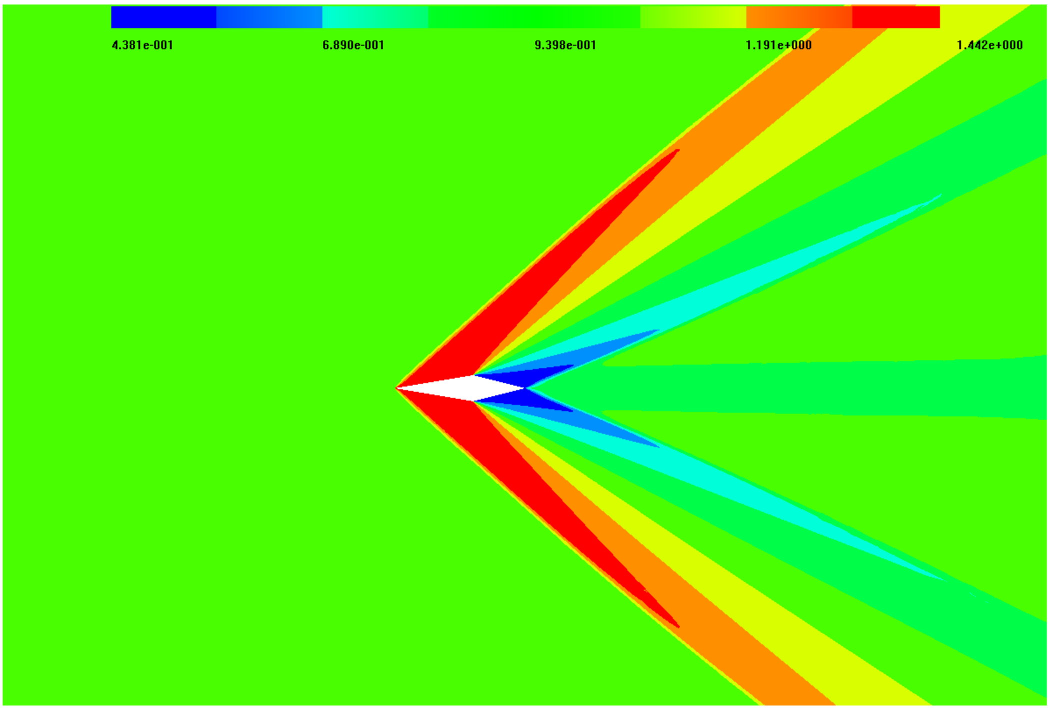
As a time stepping scheme we use the improved Euler method with time step $\tau = 0.125e{-}3$. The Knudsen number of the flow is $5e{-}3$.

We model the inflow and the initial distribution in terms of Maxwellian distributions with the desired macroscopic properties.

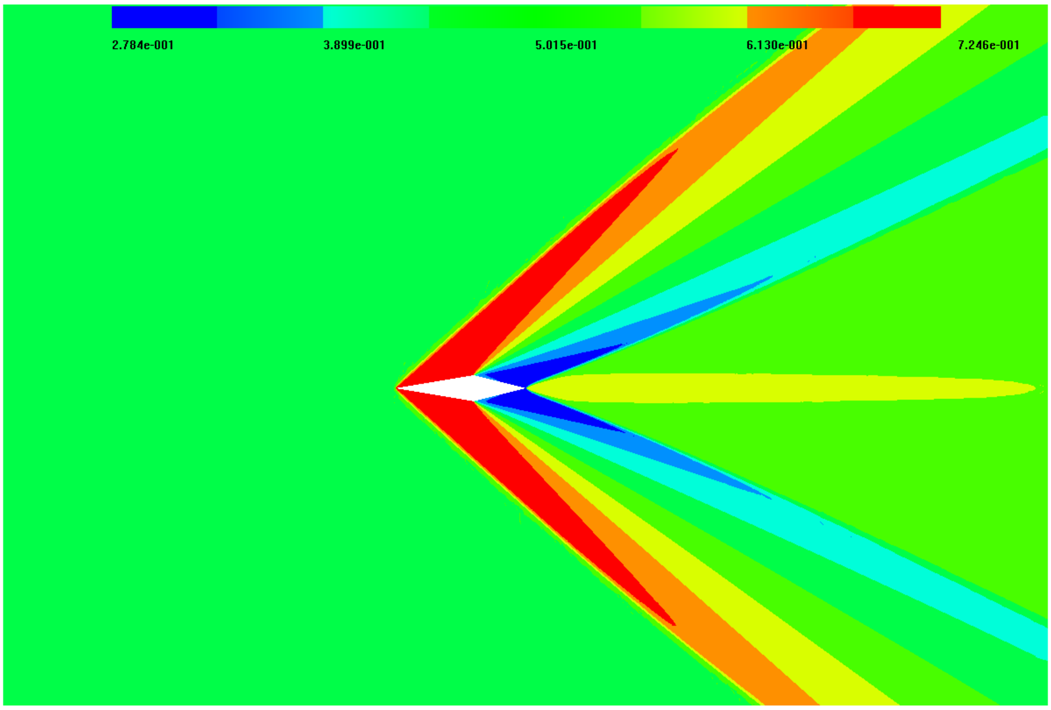
$$f_{\text{in}}(t,x,v) = f(0,x,v) = \frac{\rho_\infty}{2\pi T_\infty} e^{-\left|\frac{v-V_\infty}{\sqrt{2T_\infty}}\right|^2}. \tag{5.2.5}$$

Figure 5.2.2 shows the macroscopic properties of the flow at time $t = 3$. Comparing the solution after the shock with the theoretically obtained values is in good agreement. In the density, the temperature, the Mach number and the pressure the expected values are obtained. The measured angle of the compression is $41.82$ and is in perfect agreement with the predicted one.

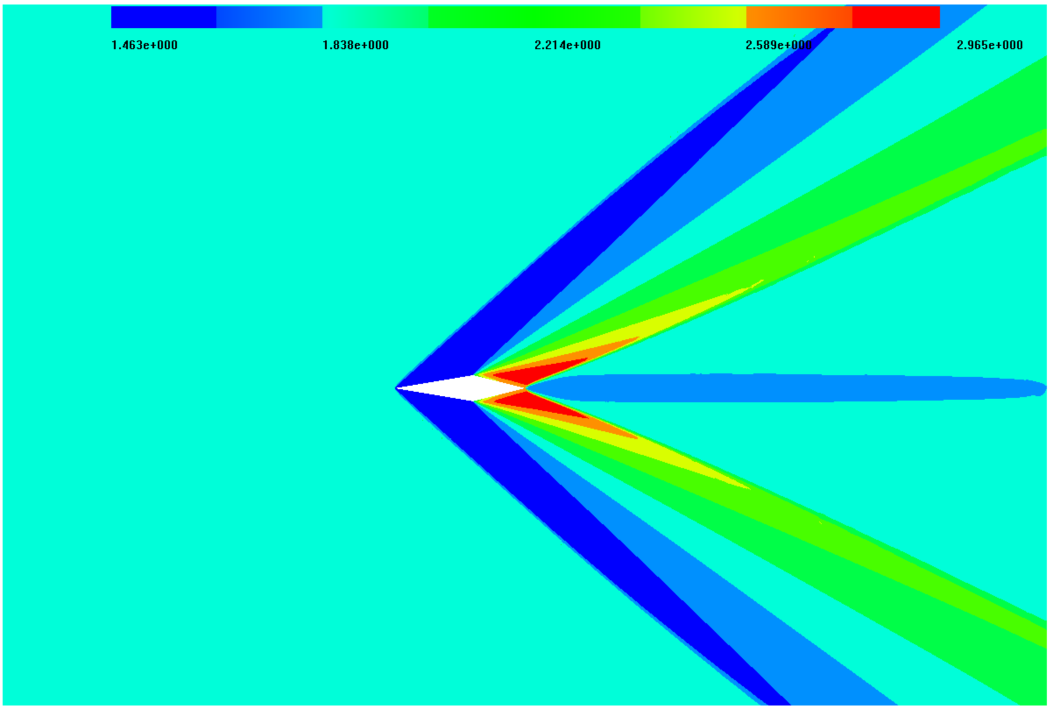
In Figure 5.2.3 we have depicted the distribution functions after the trailing edge along the symmetry axis of the deltoid at time $t = 3$. The distances to the edge are $0.02\%$, $1\%$ and $2\%$ of the length of the deltoid. Due to the symmetry of the deltoid, the distributions can be expected to be symmetric w.r.t. the $v_x$-axis what is quite well satisfied. Very close to the deltoid we obtain 2 separated peaks, symmetric w.r.t. $v_y$, resulting in a vanishing macroscopic velocity w.r.t. the $y$-direction. This separation decreases when moving further away from the wedge.



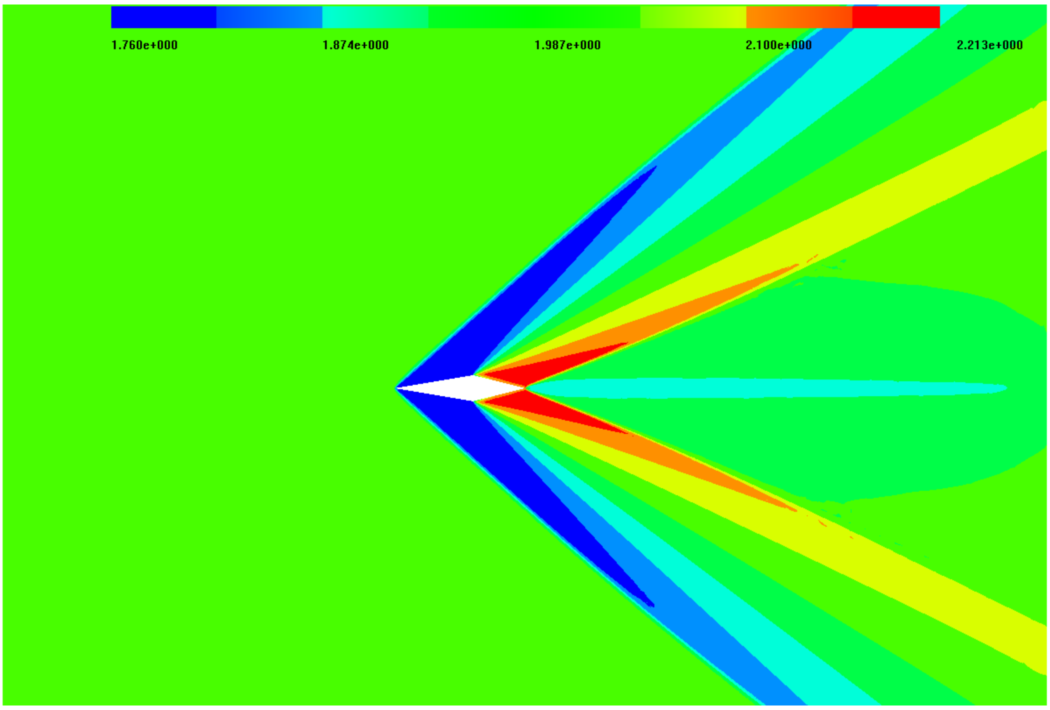
(a) Density



(b) Temperature

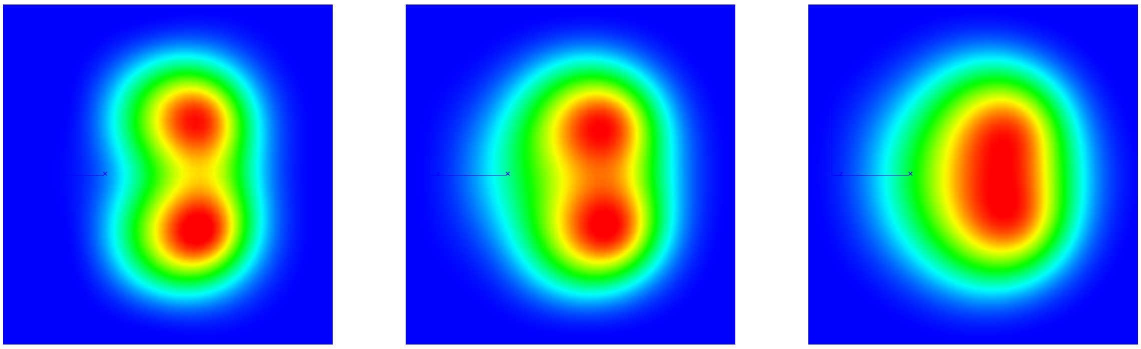


(c) Mach number



(d) Velocity

Figure 5.2.2: Macroscopic properties for the Mach 2 flow around a wedge.



(a) $0.02\%$ (b) $1\%$ (c) $2\%$

Figure 5.2.3: The microscopic behaviour at the trailing edge. From left to right the distance to the deltoid increases. The percentages denote the distance to the trailing edge w.r.t. the length of the deltoid.

### 5.2.2 Model problem 2 - Flow around a cylinder

In this example we consider the flow through a tube with a circular obstacle. Figure 5.2.4 shows the geometry and the mesh for the computations. The flow enters the tube from the left side and leaves it on the right side. The center of the circle is at $(0.5, 0)$, its radius is 0.1. The $y$ range of the tube is $[-0.25, 0.3]$, the $x$ range is $[0, 4]$. Note that the circle is not centered with respect to the $y$-direction.

For the simulation we used order 6 spatial basis functions on a mesh with 141 (partially) curved elements. The order of the momentum space is chosen quite low as 4.

For time stepping we used again the improved Euler method with time step $\tau = 0.5e-3$. The Knudsen number in this simulation was chosen equal to $5e-3$.

As in the previous example, we model the inflow as well as the initial distribution in terms of Maxwellian distributions with the desired macroscopic properties.

$$f_{\text{in}}(t, x, v) = f(0, x, v) = \frac{1}{2\pi T_\infty(x)} e^{-\left|\frac{v - V_\infty(x)}{\sqrt{2T_\infty(x)}}\right|^2}. \tag{5.2.6}$$

The boundary conditions on the cylinder and also on the upper and lower boundary of the computational domain are diffuse reflection (1.2.4b) with $T_{\text{bnd}} = 0.5$ and $V_{\text{bnd}} = (0, 0)^T$. The velocity $V_\infty$ of the initial distribution is chosen as a quadratic function with respect to $y$ such that

$V_\infty(x,-0.25) = V_\infty(x,0.3) = 0$. The temperature of the initial distribution is constant all over the domain and fits to the temperature of the wall.

$$V_\infty(x,y) = \left(0, \frac{-1200y^2 + 60y + 90}{121}\right), \qquad T_\infty(x) = 0.5. \tag{5.2.7}$$

The maximum inflow velocity is $\max\limits_{y\in[-0.25,0.3]} V_\infty(x,y) = 0.75$.

Figures 5.2.5a - 5.2.5d show the modulus of the macroscopic velocity $|V(t,x)|$ of the solution at different points in time. The first Figures capture the pressure wave that occurs in the beginning. Additionally we observe that a non stationary behaviour of the solution is obtained.

Table 5.2 presents the computation times for this example. To obtain the table, we performed the simulation multiple times with different polynomial orders in momentum space to find the fraction of the time step that was spent with the collision integrals. For low order the time consumed for the collision and the flux is more or less equally balanced. If the order is increased, we see a significant growth of the fraction spent for the collision.

The results indicate that the qualitative behaviour of the obtained solution is in good agreement with Navier-Stokes solutions for similar problems. Additionally they highlight the efficiency of the approximation by the shifted and scaled basis functions, only order 4 polynomials are used to approximate the whole velocity space $\mathbb{R}^2$. The timings in table 5.2 emphasize the necessity to keep the order as low as possible.

| order | time [s] | coll time | coll .% | flux time [s] | flux % | update time [s] | update % |
|---|---|---|---|---|---|---|---|
| 4 | 0.045 | 0.02 | 44 | 0.017 | 38 | 0.008 | 18 |
| 5 | 0.065 | 0.038 | 56 | 0.017 | 26 | 0.01 | 18 |
| 6 | 0.081 | 0.05 | 62 | 0.02 | 25 | 0.011 | 13 |
| 7 | 0.123 | 0.086 | 70 | 0.023 | 19 | 0.014 | 11 |
| 8 | 0.209 | 0.16 | 77 | 0.032 | 15 | 0.017 | 8 |
| 9 | 0.313 | 0.26 | 83 | 0.035 | 11 | 0.018 | 6 |
| 10 | 0.581 | 0.52 | 90 | 0.04 | 7 | 0.021 | 3 |

Table 5.2: Timings for the flow around a cylinder. The values in the table are obtained by running the simulation multiple times with different polynomial order in the momentum domain. All other parameters are kept unchanged. Already for order 10 polynomials 90 percent of the computational time is spent for the collision integrals.

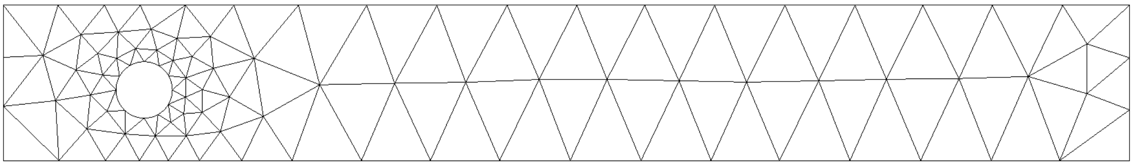
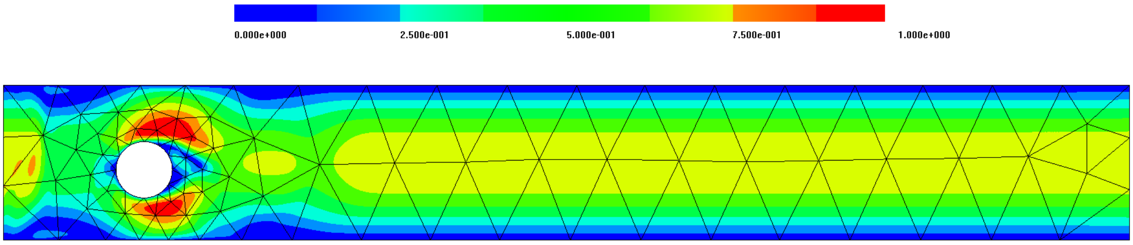
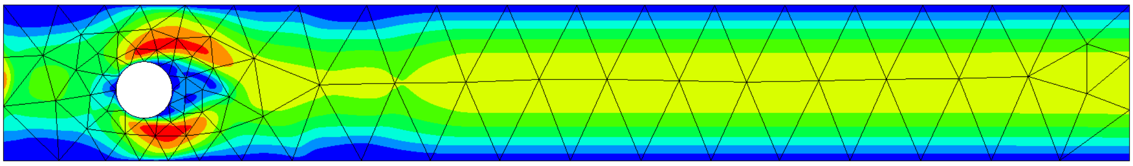


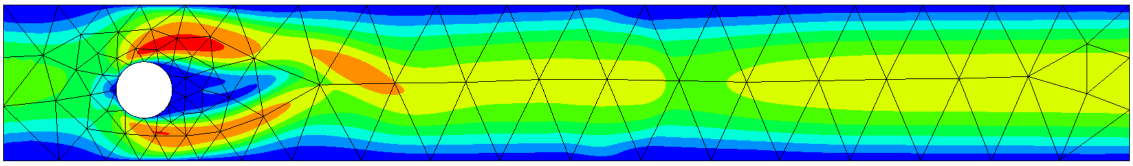
Figure 5.2.4: Geometry and mesh for the computations.



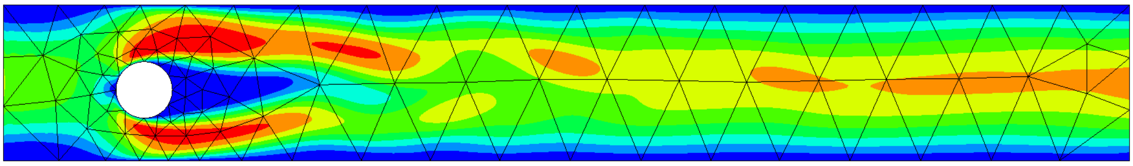
(a) $t = 0.45$.



(b) $t = 0.65$.



(c) $t = 1.5$.



(d) $t = 5.2$.

Figure 5.2.5: Modulus of the velocity at different points in time.

### 5.2.3 Flow around NACA 7410 air foil - specular reflection

Here we consider the flow around a NACA 7410 air foil. The surface of the air foil is assumed to be perfectly smooth w.r.t. the molecular diameter. Thus, the specular reflection boundary condition (1.2.4a) is applied along the air foil. At the outer boundary we apply the inflow boundary condition (1.2.4c) with constant inflow $f_{\text{in}}$ over time. The length of the air foil is 1, the length of the computational domain is 4 and its height is 6.

In the spatial space, the polynomial order is globally set to 2, the order of the velocity space is not constant over the mesh. There are 3 element layers with order 10 around the air foil, followed by 2 element layers with order 8. All other elements have velocity order 6. The higher $v$-order around the air foil is necessary to resolve the reflected particles.

To resolve the shocks, the computation is started on an initial mesh consisting of 749 elements, depicted in Figure 5.2.6a. At $t = 2$ simulation seconds, an error based refinement of the mesh is carried out and the simulation is restarted. The estimation of the error is done as presented by Zienkiewicz and Zhu [ZZ92a,ZZ92b], the quantity the estimation was based on is the macroscopic density. To end up with the mesh depicted in Figure 5.2.6c, 4 levels of refinement were done.

The initial distribution as well as the inflow distribution are chosen as Maxwellian distributions, resulting in

$$f_{\text{in}}(t,x,v) = f(0,x,v) = \frac{\rho_\infty}{2\pi T_\infty} e^{-\left|\frac{v-V_\infty}{\sqrt{2T_\infty}}\right|^2}, \tag{5.2.8}$$

with

$$\rho_\infty = 1 \qquad T_\infty = 0.5 \qquad V_\infty = (2,0)^T. \tag{5.2.9}$$

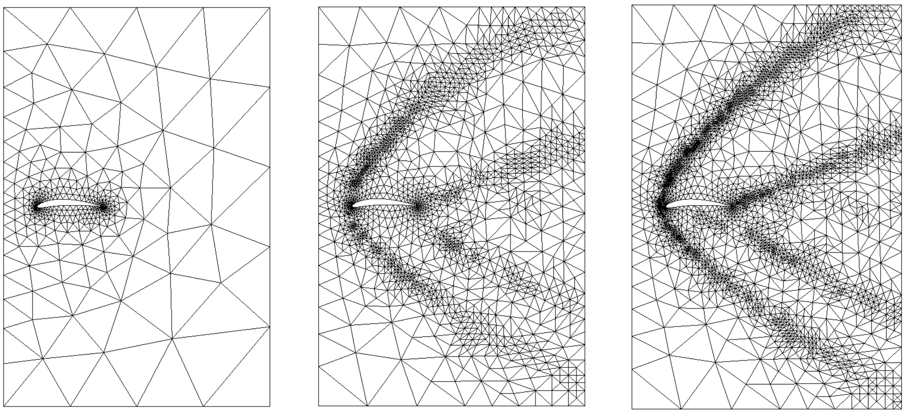
These values correspond to a free stream Mach number $M_\infty = 2$.

Time stepping is again carried out by the improved Euler method with time step $\tau = 0.25e{-}4$. The Knudsen number is set to $5e{-}3$.

Figure 5.2.7 depicts the macroscopic density, the temperature, the modulus of the macroscopic velocity and the Mach number at time $t = 3$ on the mesh 5.2.6c. The compression shocks are very well resolved as we see in each of the macroscopic quantities. The coloured results are hardly distinguishable w.r.t. the meshes 5.2.6b and 5.2.6c. Thus, only results for the finest mesh are presented.

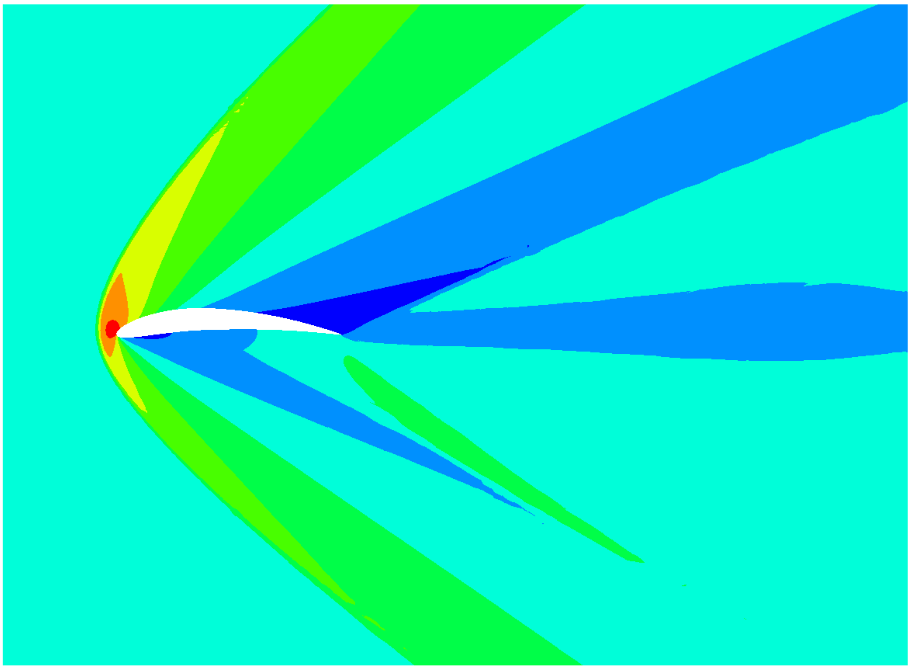
A comparison of solutions on these two meshes is shown with isolines in Figure 5.2.8. There are fewer overshoots in the approximation of the shocks on the finer mesh.

Figures 5.2.10 and 5.2.12 show the distribution function at different points in space through the bow shock and close to the trailing edge of the air foil respectively.
The position in 5.2.10a is not yet affected by the air foil, it is directly in front of the shock and represents the free flow distribution. Both, Figure 5.2.10b and 5.2.10c are placed in the transition, we see a slight derivation from a Maxwellian. Figure 5.2.10d presents the distribution behind the shock.
The distributions shown in Figure 5.2.12a-5.2.12d are on a horizontal line through the shock at the trailing edge. Since the air foil is not symmetric, there is no symmetry w.r.t. the $v_x$-axis any more, only a small number of particles travels upwards in 5.2.12a. When moving away from the edge, this effect decreases, in Figure 5.2.12c both, upwards and downwards moving particles are even spread. The distributions obviously not Maxwellian.
The Figures 5.2.12e-5.2.12g show the distribution along the trailing edge shock.

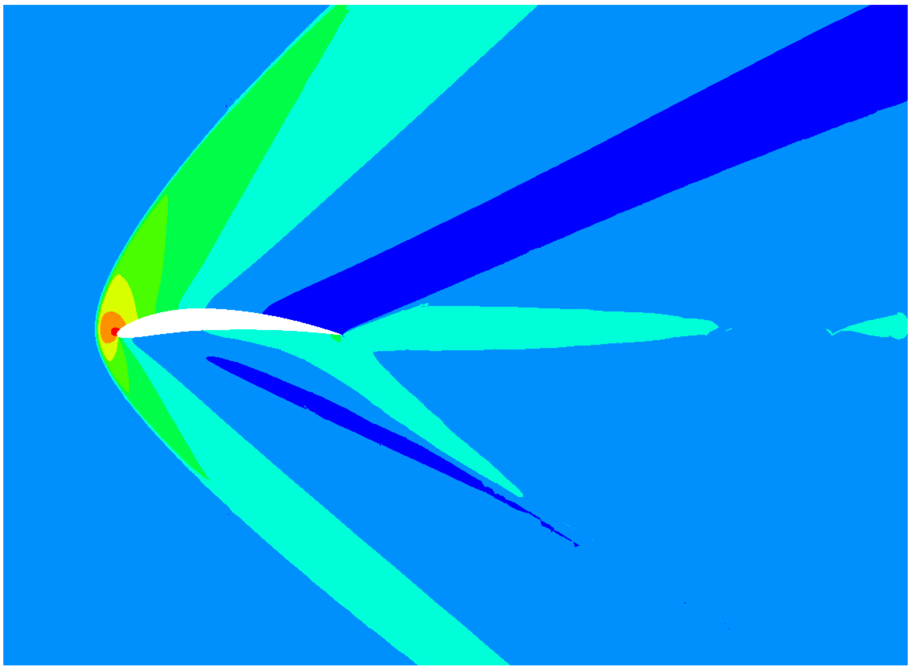


(a) Initial mesh. (b) level 3. (c) level 4.

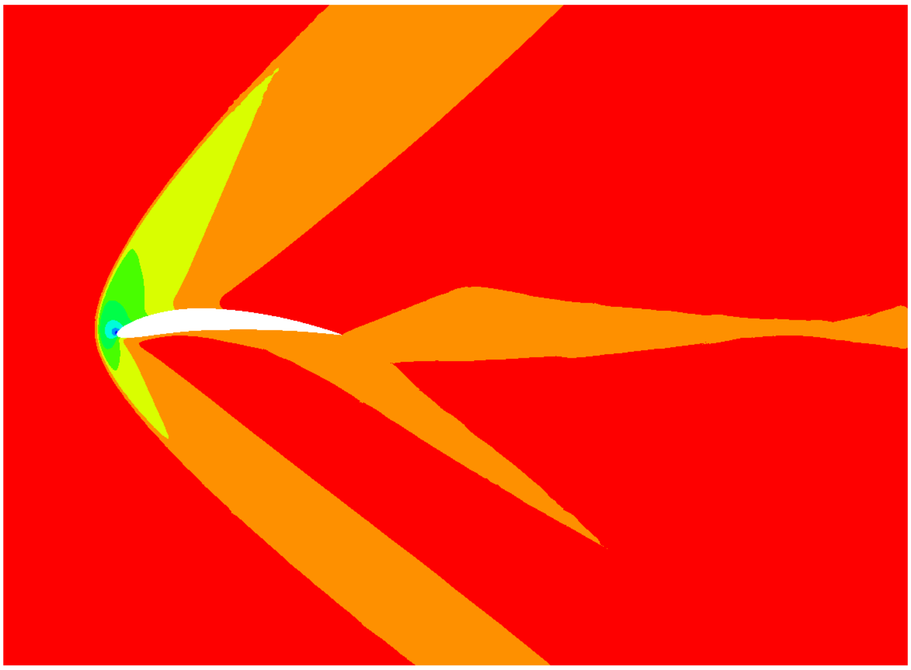
Figure 5.2.6: The initial mesh with 749 elements on the left, the refined mesh on the middle consists of 3560 elements, the final mesh on the right is made of 5875 elements.



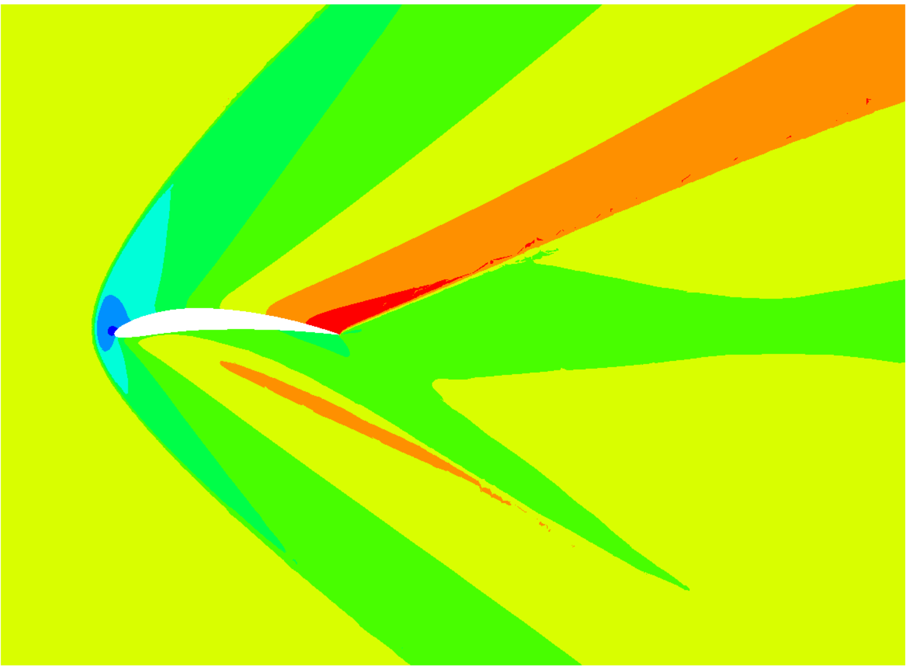
(a) Macroscopic density.



(b) Macroscopic temperature.

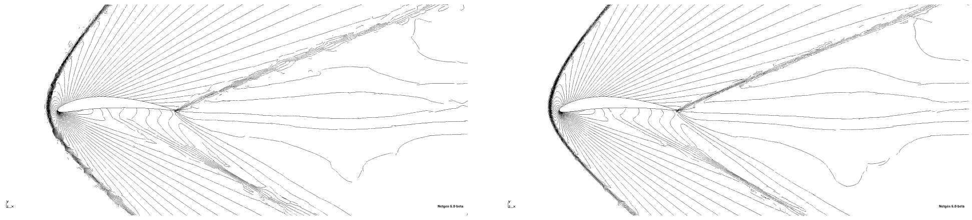


(c) Modulus of macroscopic velocity.

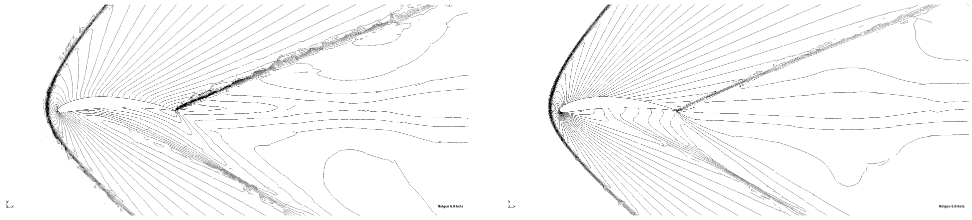


(d) Macroscopic Mach number.

Figure 5.2.7: Simulation results for NACA 7410 air foil on mesh 5.2.6c.



(a) Isolines for the macroscopic density.



(b) Isolines for the macroscopic Mach number.

Figure 5.2.8: Isolines for macroscopic density and Mach number on the mesh 5.2.6b (left) and 5.2.6c (right).

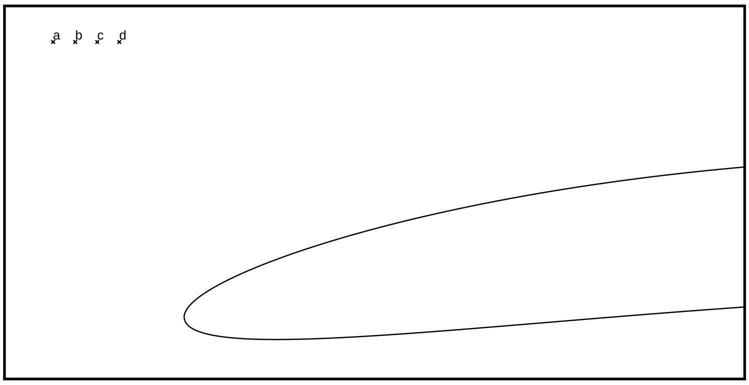
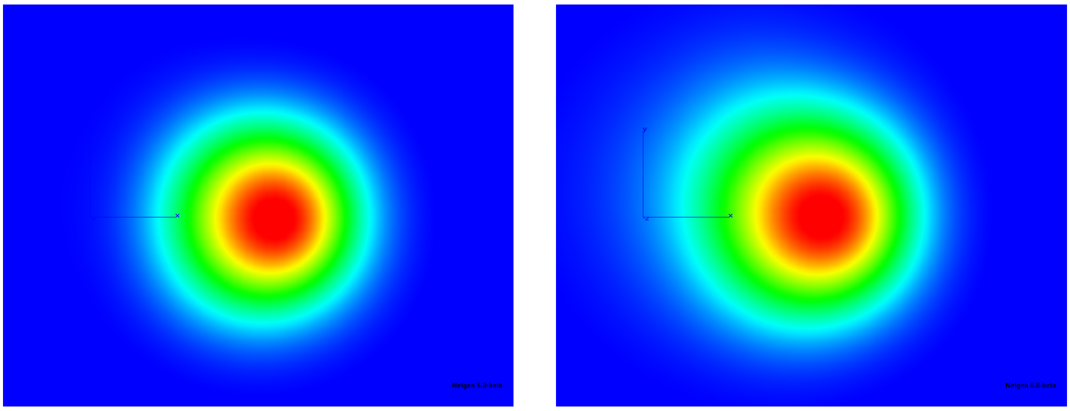
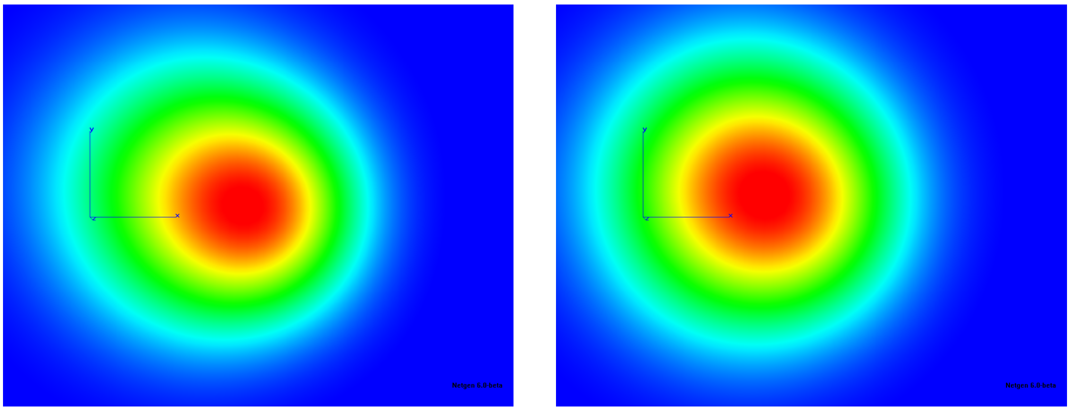


Figure 5.2.9: The spatial positions of the distribution functions presented in Figure 5.2.10.



(a) $(x, y) = (-0.56, 0.2)$.

(b) $(x, y) = (-0.55, 0.2)$.



(c) $(x, y) = (-0.54, 0.2)$.

(d) $(x, y) = (-0.53, 0.2)$.

Figure 5.2.10: Distribution functions on a straight line through the bow shock.

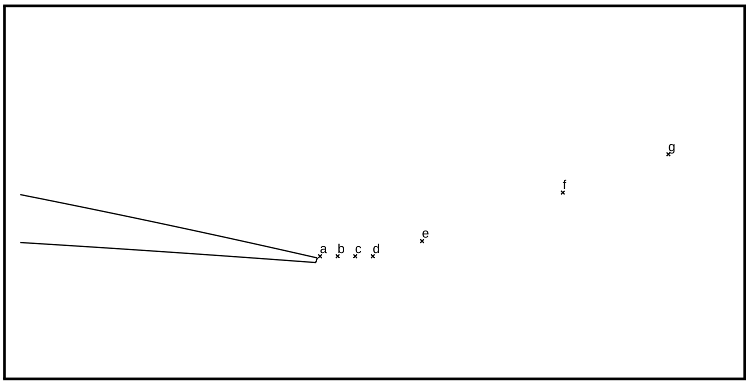
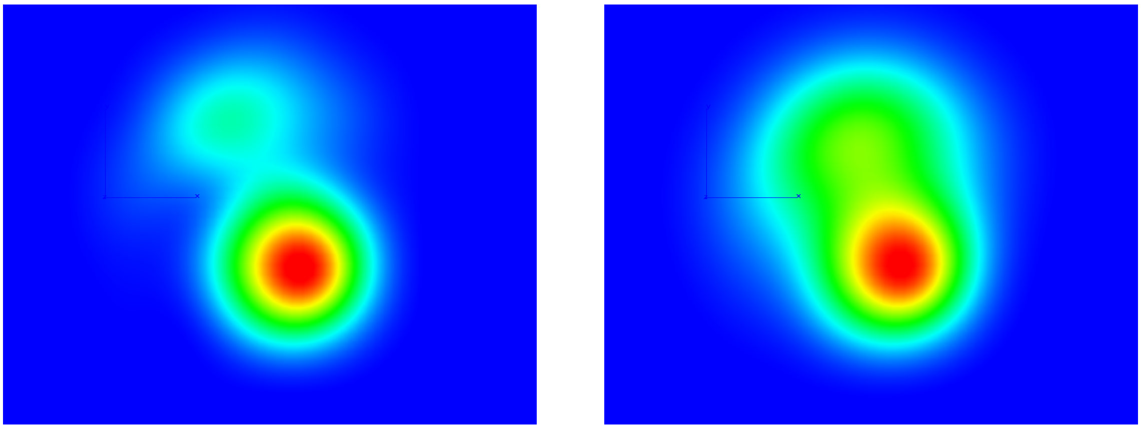
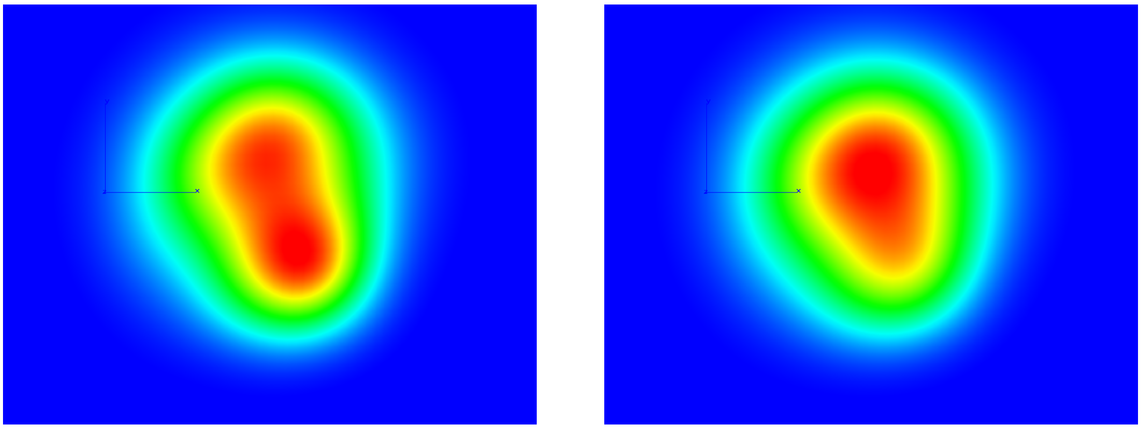


Figure 5.2.11: The spatial positions of the distribution functions presented in Figure 5.2.12.



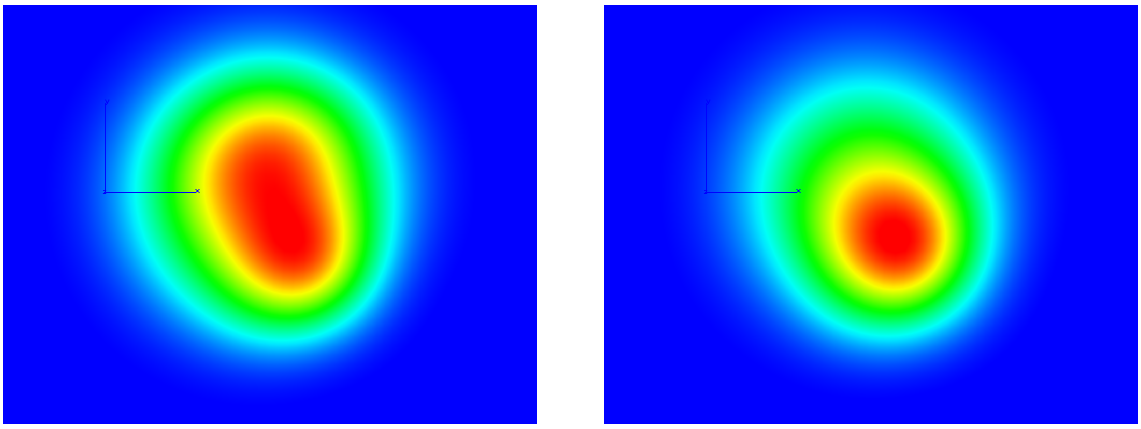
(a) $(x, y) = (0.501, 0.0018)$.

(b) $(x, y) = (0.506, 0.0018)$.



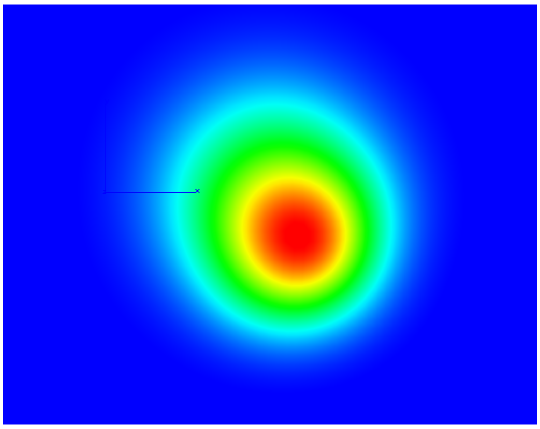
(c) $(x, y) = (0.511, 0.0018)$.

(d) $(x, y) = (0.516, 0.0018)$.



(e) $(x, y) = (0.53, 0.0085)$.

(f) $(x, y) = (0.57, 0.03)$.



(g) $(x, y) = (0.6, 0.047)$.

Figure 5.2.12: Distributions immediately after the trailing edge (5.2.12a-5.2.12d) and along the trailing edge shock (5.2.12e-5.2.12g).

### 5.2.4 Flow around NACA 7410 air foil - diffuse reflection

Now we consider the NACA 7410 air foil geometry in a second simulation, but in contrast to the previous example we use the diffuse reflecting boundary condition along the air foil. Particles are reflected according to (1.2.4b) when hitting the boundary.

The polynomial order in the spatial domain is 2. For the order in the momentum domain we have chosen two element layers around the air foil with polynomial order 4, all other elements consist of polynomial degree 3 in velocity direction. The mesh was not refined by error estimation, but is finer in general as in the Mach 2 example.

The time step $\tau$ was chosen as $0.25e$–$4$. As a time stepping scheme we used again the improved Euler method. The simulation was performed with Knudsen number $5e-3$.

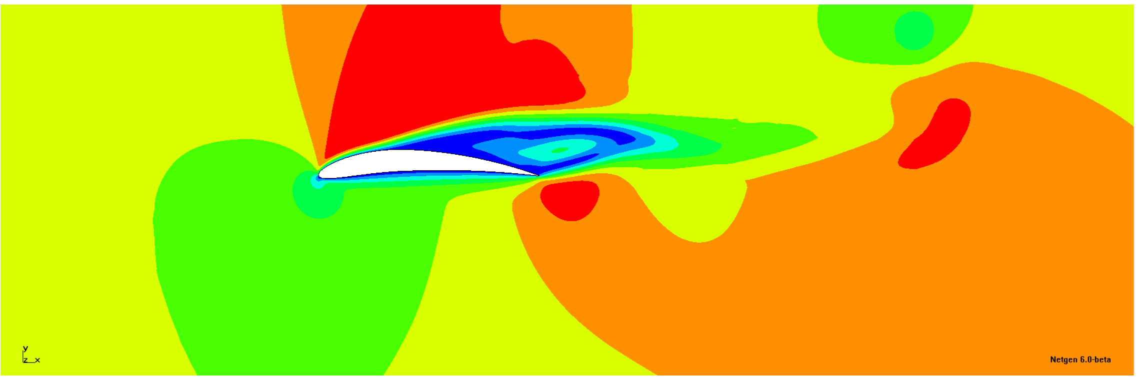
As before we model the initial distribution as well as the boundary distribution at the outer edges in terms of Maxwellian distributions

$$f_{\text{in}}(t, x, v) = f(0, x, v) = \frac{\rho_\infty}{2\pi T_\infty} e^{-\left|\frac{v-V_\infty}{\sqrt{2T_\infty}}\right|^2}. \tag{5.2.10}$$

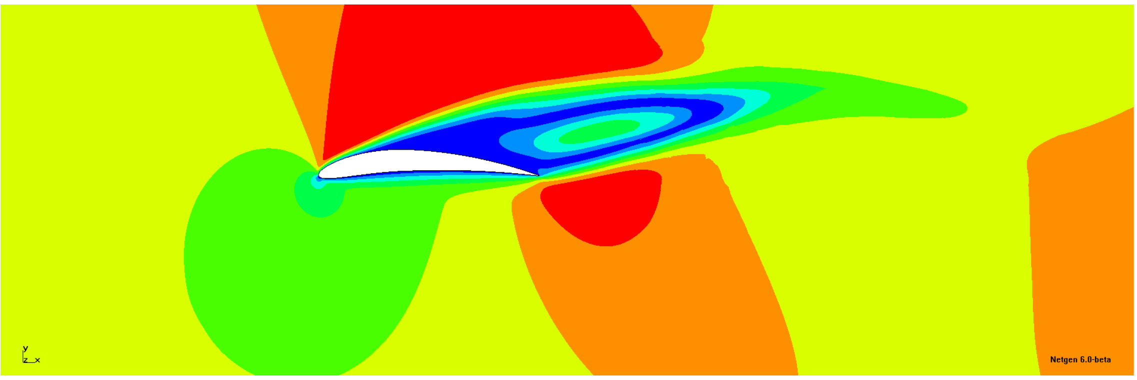
The macroscopic properties of the free flow are $\rho_\infty = 1.0, \quad T_\infty = 0.5, \quad V_\infty = (0.7, 0.2)^T$, resulting in a free stream Mach number of $M_\infty \approx 0.73$, the flow is subsonic. The temperature of the boundary is chosen in accordance with the free flow temperature, $T_{\text{bnd}} = 0.5$, its velocity is $0$. The choice of the free flow velocity yields an angle of attack of $15°$.

In Figure 5.2.13 we depict the modulus of the velocity at different points in time. As a consequence of the diffuse reflection condition we obtain a boundary layer around the air foil and the solution becomes non stationary.

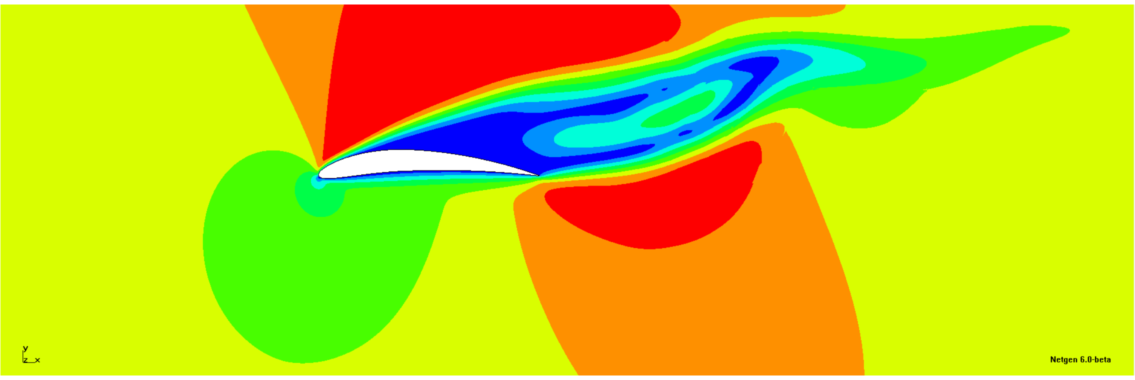
The example confirms once more the enhanced approximation properties of the shifted and scaled basis functions. Actually we have only used 16 basis functions in most of the elements to approximate the velocity space $\mathbb{R}^2$. The efficiency of our method therefore results not solely from the techniques to apply the collision operator, but also from the rather low expansion order we are able to chose in the velocity domain.



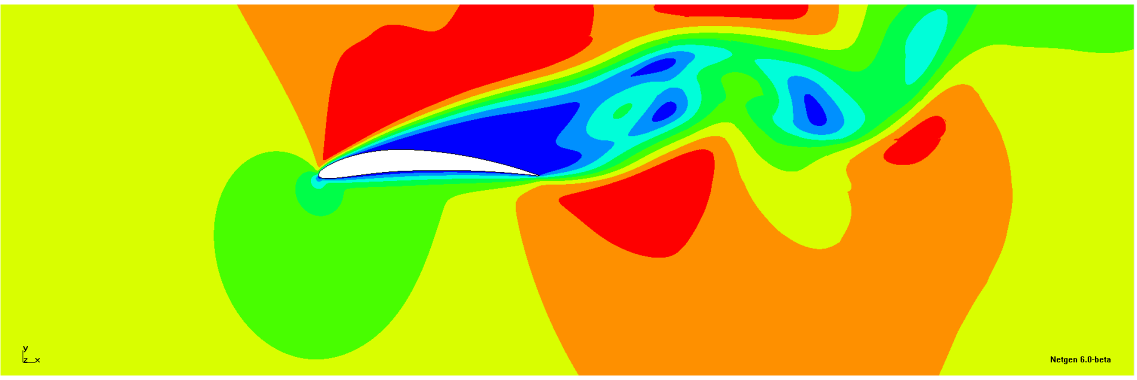
(a) Modulus of velocity at $t = 3$.



(b) Modulus of velocity at $t = 5$.



(c) Modulus of velocity at $t = 7$.



(d) Modulus of velocity at $t = 9$.

Figure 5.2.13: Macroscopic behaviour for NACA 7410 airfoil with diffuse reflection boundary condition.

### 5.2.5 Mach 3 wind tunnel with step

Here we consider the Mach 3 wind tunnel experiment. The geometry of the tunnel as well as the mesh for the computations are presented in Figure 5.2.14. The tunnel has a backward facing step at position $x = 0.6$ with height $0.2$. The total length of the tunnel is 3, the height on the left side is 1.

The mesh used for the calculation consists of 3772 spatial elements with order 2 trial and test polynomials in space. The order in momentum is 8 almost over the whole domain. A small fraction of elements close to the step has $v$-order 10.

As in the previous examples, we use the improved Euler method for time stepping with time step $\tau = 2.5e{-}5$. The Knudsen number in the current example is equal to $2.5e{-}3$.

The inflow occurs from the left side and is given by

$$f_{\text{in}}(t, x, v) = f(0, x, v) = \frac{\rho_\infty}{2\pi T_\infty} e^{-\left|\frac{v - V_\infty}{\sqrt{2T_\infty}}\right|^2} \tag{5.2.11}$$

with

$$\rho_\infty(x) = 1.4, \qquad V_\infty(x) = (3, 0)^T, \qquad T_\infty(x) = 0.5. \tag{5.2.12}$$

On the upper and lower walls of the tunnel we apply the specular reflection condition, on the left and the right side the inflow condition is used.

Figures 5.2.15a - 5.2.15c present the macroscopic density at different points in time. In the first Figure the compression shock starts to evolve from the step. At $t = 0.5$ it has already reached the top wall of the tunnel and gets reflected. At time $t = 1$ the shock has already detached from the upper wall. We note small oscillations in the region of the reflected shock at $t = 0.5$.

A qualitative comparison of our results with solutions obtained by the Euler equations shows good agreement. Typically, these results are for a gas with 5 degrees of freedom, yielding different positions and strengths of the shocks.

The computation time for the simulation per time step is 0.858 s, where 0.74 s (86%) are spent for the collision integrals, the flux takes 0.078 s (9%) and the application of the inverse mass matrix including the solution update takes 0.04 s (5%). Almost 90% of the time per step are spent in the calculation of the collision integrals. This shows the need for efficient evaluation of the collision integrals.

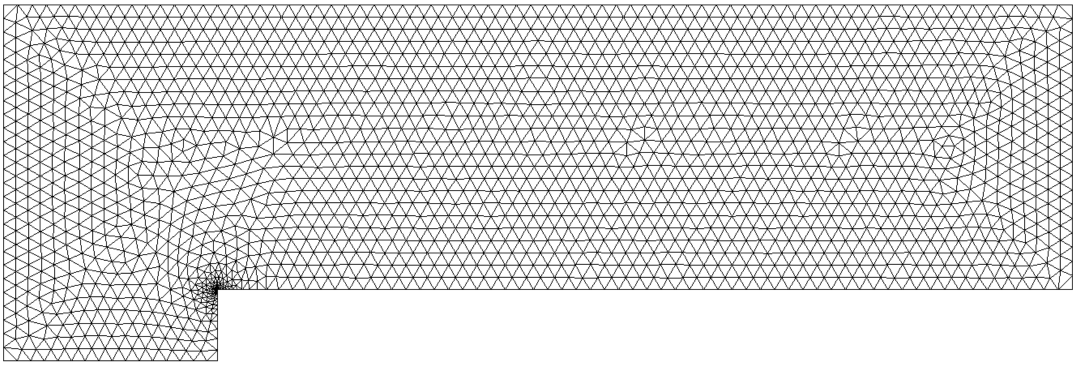
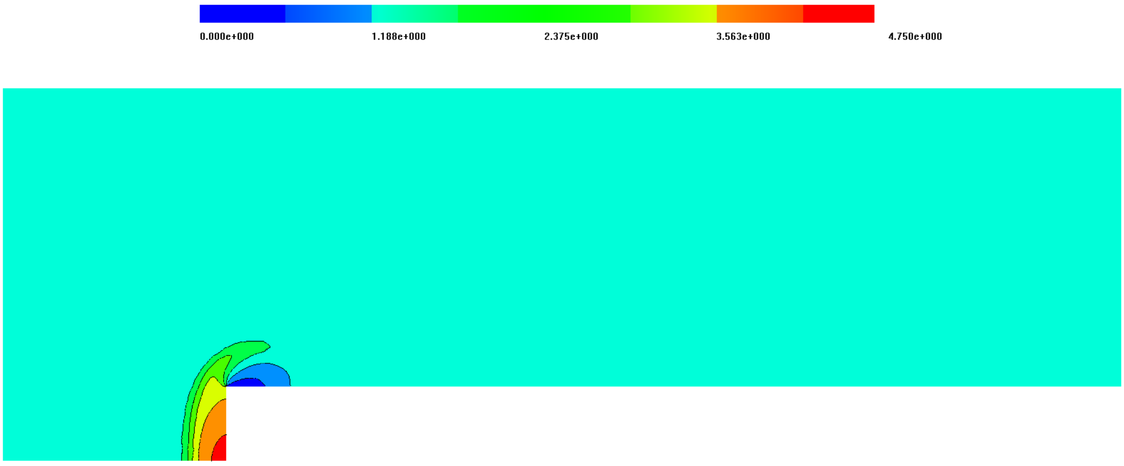
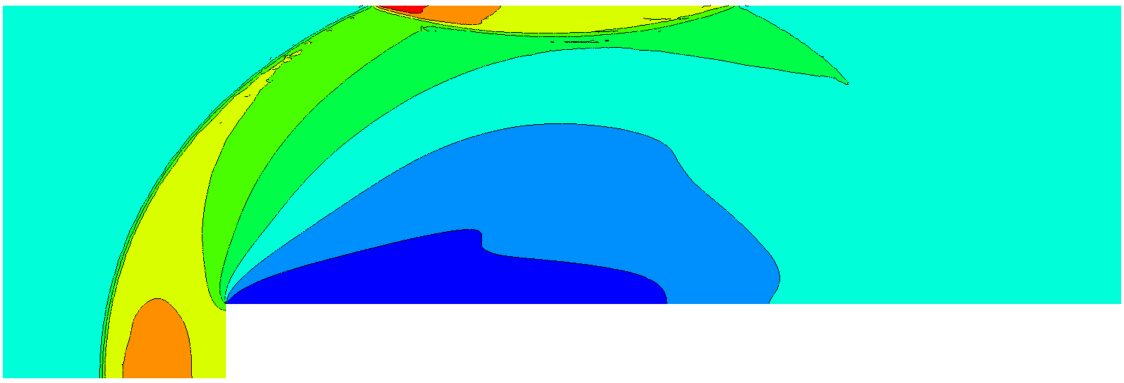


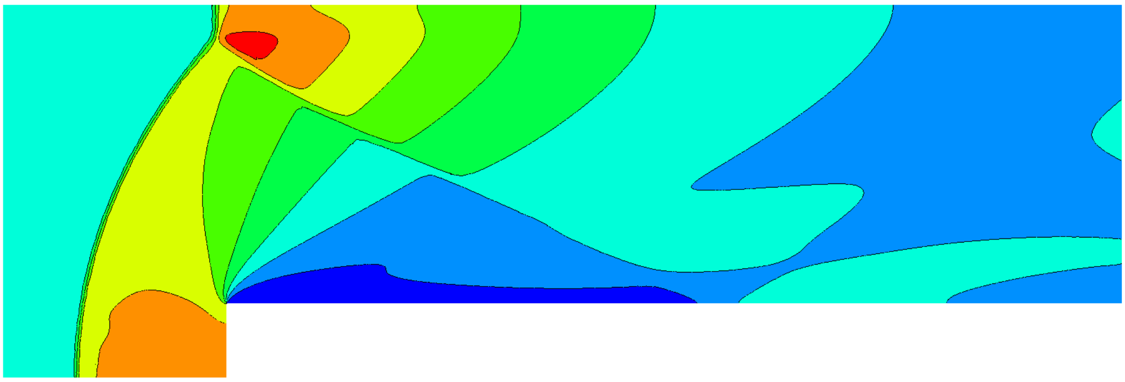
Figure 5.2.14: Wind tunnel geometry and computational mesh.



(a) Macroscopic Density $\rho(x)$ at time $t = 0.075$.



(b) Macroscopic Density $\rho(x)$ at time $t = 0.5$.



(c) Macroscopic Density $\rho(x)$ at time $t = 1.0$.

Figure 5.2.15: The simulation results for the Mach 3 wind tunnel experiment.

### 5.2.6 The Knudsen pump

In this example we consider the flow through a tube, the geometry is depicted in Figure 5.2.16. For a gas under sufficiently rarefied conditions, a unidirectional flow through such a tube can be obtained just by imposing a temperature gradient along the walls. Here this is done by keeping the temperature at $T_0$ at the points $B$ and $D$ and at temperature $T_1$ at the points $A$ and $C$. For the other points on the boundary it is assumed, that the temperature changes linearly with the distance along the walls. The geometry and the temperature distribution along the walls are taken from [ADM$^+$07]. The temperature $T_0$ is chosen as $0.5$, the larger temperature $T_1$ is chosen as $1.5$.

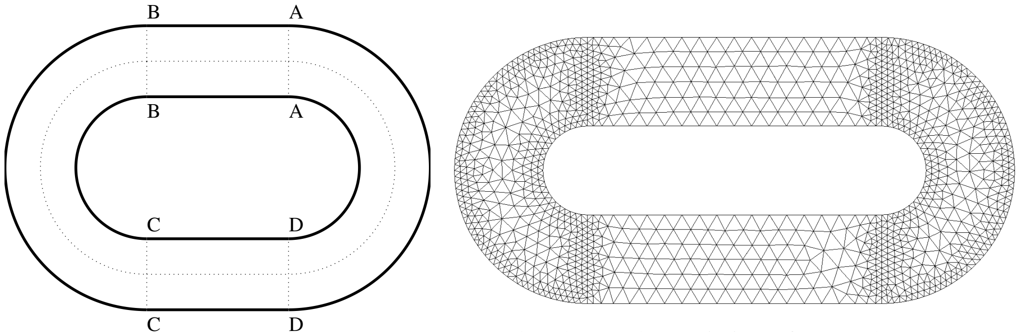
The mesh for the computations is depicted in Figure 5.2.16. The width of the tunnel is 1.5, the length of the straight segments is 5 and the radius of the center line of the curved segment is 1.5. The order in spatial space is 2. In the velocity space the polynomial orders are 7 (Kn $= 0.1$) and 9 (Kn $= 0.7$). We used the hard sphere interaction model, i.e. $B(v, w, e') = |v - w|$ in accordance with the collision kernel used in [ADM$^+$07].

The time step for the improved Euler method is $0.25e-2$.

The initial distribution is given by

$$f(0, x, v) = \frac{\rho_\infty}{2\pi T_\infty} e^{-\left|\frac{v}{\sqrt{2T_\infty}}\right|^2}, \tag{5.2.13}$$

where $T_\infty = T_0$ and $\rho_\infty = 1$.



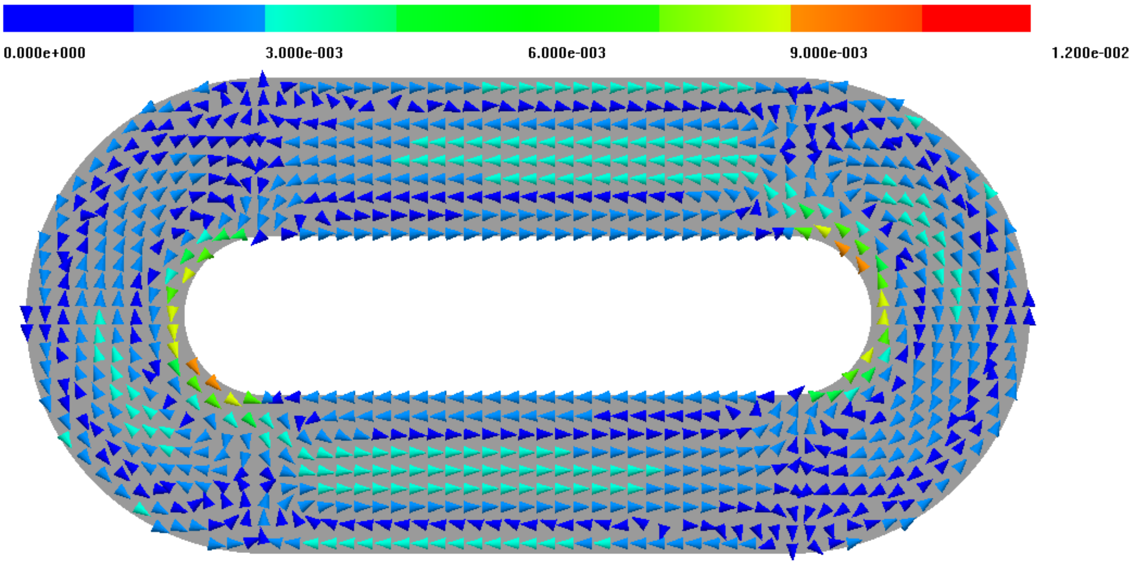
(a) The geometry.

(b) The mesh consisting of 2222 elements.

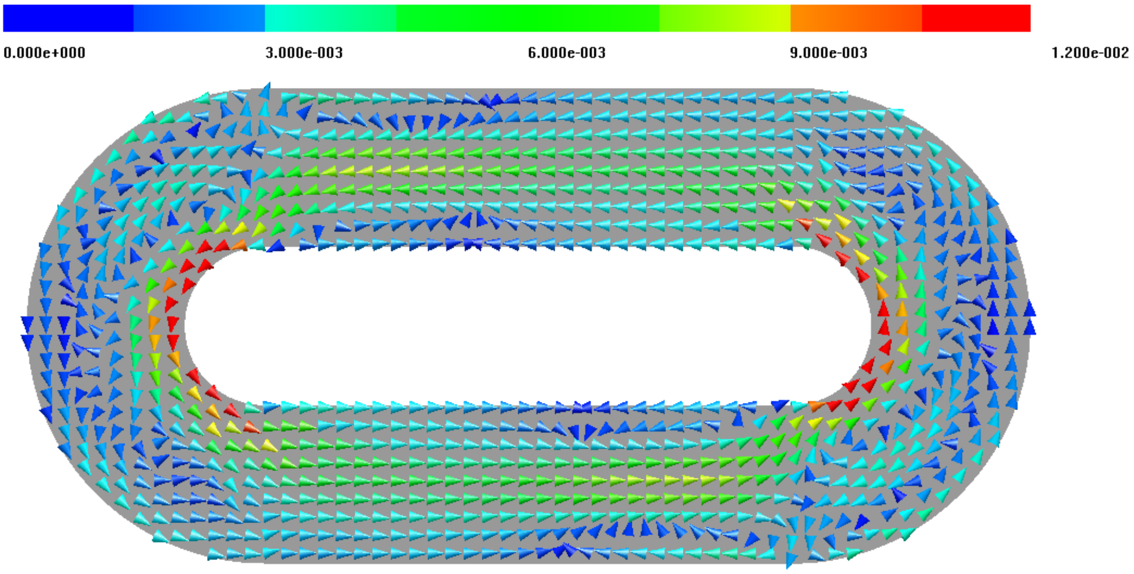
Figure 5.2.16: Geometry and mesh for the simulation of the Knudsen pump.

For both simulated Knudsen numbers we obtain a counter clock wise flow through the channel. For the low Knudsen number, shown in Figure 5.2.17a, the flow is rather limited to the inner wall of the channel in the curved segments and to the center of the channel in the straight segments. This is different in the result for the higher Knudsen number as shown in Figure 5.2.17b. Here the flow is not localized to the center of the straight segments any more, but spreads almost over the whole cross section. Additionally, the flow enters deeper into the outer half of the curved segments. The clock wise flow we observe in the straight segments for Kn $= 0.1$ is strongly reduced for Kn $= 0.7$. This tendency is also reported in [ADM$^+$07].

Figure 5.2.18 shows the expansion order that is necessary to approximate the $L_2$-norm of the stationary solution up to $99.5\%$. These results indicate the need for adaptivity w.r.t. the velocity variable. For the lower Knudsen number we can approximate the solution with a low order expansion almost in the whole channel, as is shown in Figure 5.2.18a. For the larger Knudsen number, the situation is different. The desired expansion order is larger in general. However, there is still a significant portion of domain where a lower order expansion would be sufficient to approximate the solution as can be seen in Figure 5.2.18b. By virtue of this result it should be possible to construct a criterion to adjust the expansion order based on this quantity.

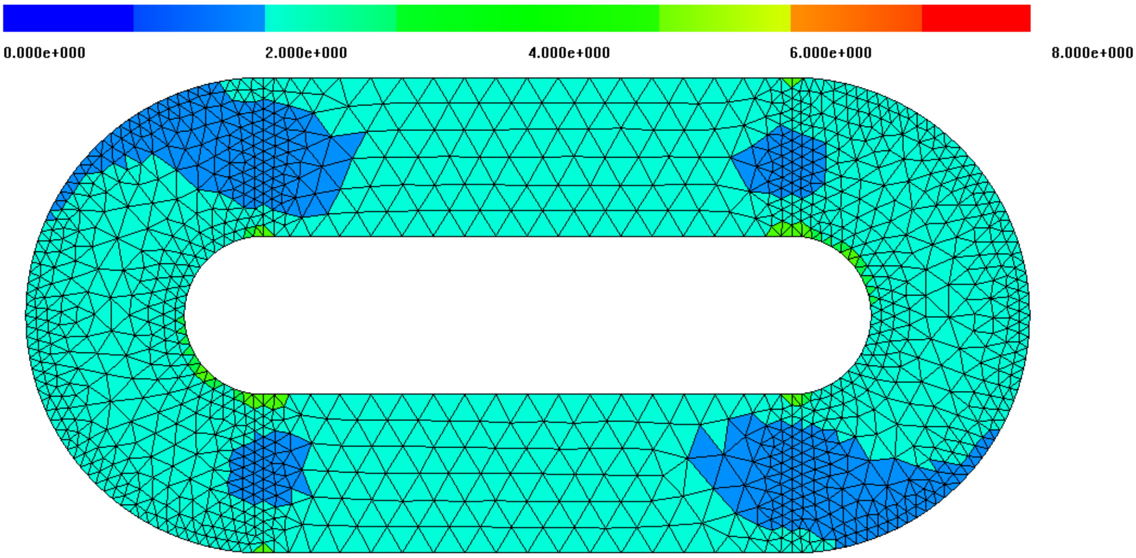


(a) Kn $= 0.1$

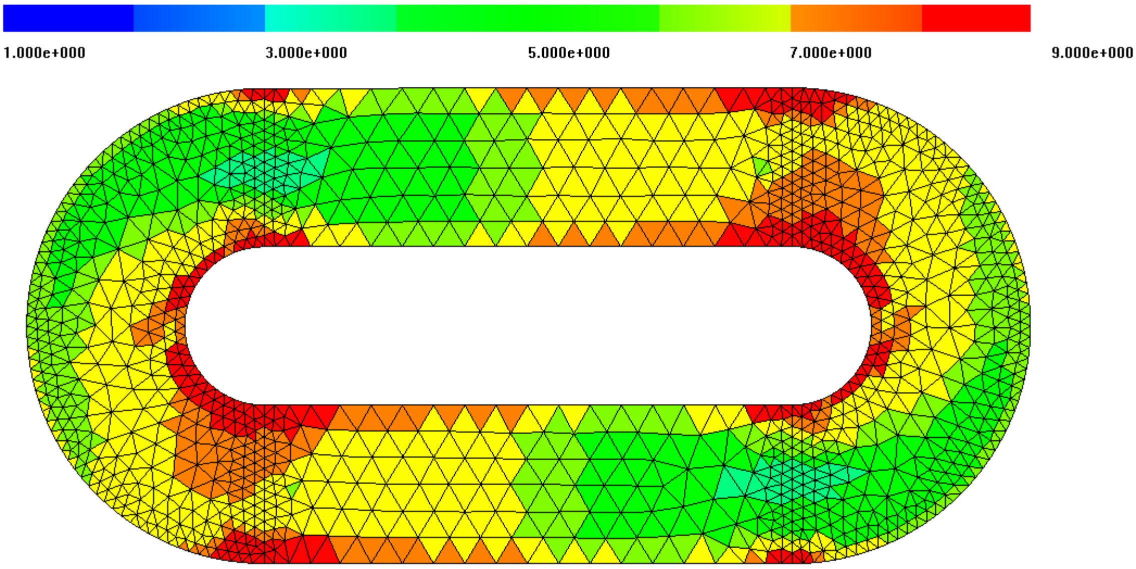


(b) Kn $= 0.7$

Figure 5.2.17: The simulation result for the Knudsen pump. Both Figures show the macroscopic velocity of the flow. While the flow is rather localized for Kn = 0.1, it spreads over almost the whole straight segments for Kn $= 0.7$.



(a) Kn $= 0.1$



(b) Kn $= 0.7$

Figure 5.2.18: Here we show the necessary polynomial orders to obtain the $L_2$ norm of the presented solution in Figure 5.2.17 with a relative error of $0.5\%$.

# 6 Summary and Outlook

## 6.1 Summary and Conclusion

In the thesis we developed a numerical method for the (2+2) dimensional Boltzmann equation based on a Discontinuous Galerkin projection. The approximation of the solution in the spatial and the velocity space was done via a tensor product.

We have used global polynomials w.r.t. the velocity variable multiplied with an appropriate Maxwellian to approximate the solution. As has been confirmed by numerical examples, this yields very good approximation properties close to equilibrium. On the other hand, by the use of polynomial test functions w.r.t. the velocity, the conservation equations satisfied by the solution were naturally satisfied on the discrete level.

In contrast to a lot of proposed deterministic methods there was no need to truncate the support of the solution function, neither the integration domains in the variational formulation. The arising integrals were evaluated by Gauss Hermite quadrature rules. In other words, we did not introduce additional modelling errors by truncation.

In order to enhance the approximation properties of the trial space we proposed an adaptation of the Maxwellian in terms of a variable shift and scale. By the use of a Discontinuous Galerkin method there is no coupling between the local basis functions on different elements. This allowed us to choose the parameters of the Maxwellian locally, i.e. element wise. Thus, we could adapt to the macroscopic velocity and temperature within each element, and only a low polynomial order in the velocity space was needed to approximate the solution. In section 4.4 we have noted that the choice of these parameters needs some additional treatment to avoid stability issues. We showed an appropriate smoothing of the temperature parameter and discussed how to incorporate the behaviour of the distribution function at the walls into the velocity parameter.

A huge part of the work was devoted to the application of the collision integrals. The collision operator takes $\mathcal{O}(N^6)$ operations for a straight forward application. The main ideas to arrive at reduced costs were presented in section 4.3. We transformed the collision integrals to mean and relative velocity and approximated the outer integral w.r.t. to the mean velocity by a Gauss Hermite

quadrature rule. The number of integration points in this quadrature rule is bounded by $\mathcal{O}(N^2)$. For the evaluation of the remaining integral w.r.t. the relative velocity we proposed a hierarchical polynomial basis in Polar coordinates. In lemma 4.3.3 we showed that the inner integral operator is diagonal in this basis and therefore easy and cheap to apply. For efficient transformation from the nodal to the hierarchical representation we introduced an intermediate hierarchical basis formulated in terms of Hermite polynomials. The transformation from the nodal to the Polar representation was then executed as a composition of the transforms from nodal to Hermite and from Hermite to Polar. The numerical work needed for both transforms is bounded by $\mathcal{O}(N^3)$ operations in total as was shown in sections 4.3.4 and 4.3.5. The first bound was obtained by exploiting the tensor product structure in both, the nodal and the Hermite basis. To arrive at the second bound we used the orthogonality of the Hermite and Polar basis w.r.t. to the Maxwellian weighted $L_2$-inner product on the one hand and their hierarchical structure on the other hand.
Summarizing, we obtain costs of $\mathcal{O}(N^3)$ times the number of integration points w.r.t. to the mean velocity. This results in $\mathcal{O}(N^5)$ operations for the application of the collision integrals, $N$ denoting the expansion order.

The numerical examples already showed the potential of the method. We obtained exponential convergence empirically for a space homogeneous problem. On the other hand, the space inhomogeneous problems demonstrate that Euler as well as Navier-Stokes solutions can be produced by our method. In addition, they demonstrate excellent approximation properties of the adapted trial spaces. Already very low expansion orders w.r.t. the velocity give reasonable results. Due to the immense costs of the collision operator, the importance of the ability to use low expansion orders is evident.

## 6.2 Outlook

As a next step we plan to extend our method to three dimensions, in space and velocity. In particular, a 3d velocity space is important to obtain quantitative accurate results. Our algorithm for the collision operator based on transformations between tensor product and hierarchical basis can be extended to 3d also. The trigonometric basis on the unit sphere therefore has to be replaced by the spherical harmonics. For the sparsity of the inner collision operator w.r.t. the Polar basis, the Funk Hecke theorem should be useful. We expect a total cost of $\mathcal{O}(N^7)$ operations. Choosing $N = 10$, this means 100 times more operations compared to 2d. Our actual computations are in the range of hours on a modern workstation. Thus, deterministic numerical simulation with the presented method will be feasible in 3d on clusters in near future.

Another point we plan to address is adaptivity w.r.t. the velocity variable. Typically, close to a wall a higher expansion order is necessary to resolve the reflected particles. On the other hand, in the free flow region the solution is already well approximated by a low expansion order.

The criterion we aim for is based on the hierarchical expansion of $f$. Therefore we compare the contributions of each total polynomial degree to the $L_2$ norm of $f$. By the decay of these contributions, we plan to construct a rule to increase or decrease the expansion order.

# 7 Bibliography

[ABB$^+$99] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen. *LAPACK Users' Guide*. Society for Industrial and Applied Mathematics, Philadelphia, PA, third edition, 1999. 4.3.1, 4.3.4

[ABCM01] D. N. Arnold, F. Brezzi, B. Cockburn, and L. D. Marini. UNIFIED ANALYSIS OF DISCONTINUOUS GALERKIN METHODS FOR ELLIPTIC PROBLEMS. *SIAM J. Numer. Anal.*, 39(5):1749–1779, May 2001. 4.1

[ADM$^+$07] K. Aoki, P. Degond, L. Mieussens, M. Nishioka, and S. Takata. Numerical Simulation of a Knudsen Pump Using the Effect of Curvature of the Channel. *Rarefied Gas Dynamics. MS Ivanov and AK Rebrov, Eds. Novosibirsk*, pages 1079–1084, 2007. 5.2.6, 5.2.6

[Bab86] H. Babovsky. On a Simulation Scheme for the Boltzmann Equation. *Math. Methods Appl. Sci.*, 8(2):223–233, 1986. 3.1

[BF91] F. Brezzi and M. Fortin. *Mixed and Hybrid Finite Elements Methods*. Springer series in computational mathematics. Springer-Verlag, 1991. 4.4

[Bir95] G. A. Bird. *Molecular Gas Dynamics and the Direct Simulation of Gas Flows*, volume 42 of *Oxford Engineering Science Series*. The Clarendon Press, Oxford University Press, New York, 1995. Corrected reprint of the 1994 original, With 1 IBM-PC floppy disk (3.5 inch; DD), Oxford Science Publications. 3.1

[Bob88] A. V. Bobylev. THE THEORY OF THE NONLINEAR SPATIALLY UNIFORM BOLTZMANN EQUATION FOR MAXWELL MOLECULES. In *Mathematical physics reviews*, Soviet Sci. Rev. Sect. C Math. Phys. Rev., pages 111–233. Harwood Academic Publishers, 1988. 3.2.3, 5.1.1

[BR97a] F. Bassi and S. Rebay. A High-Order Accurate Discontinuous Finite Element Method for the Numerical Solution of the Compressible Navier-Stokes Equations. *Journal of Computational Physics*, 131(2):267 – 279, 1997. 4.1

[BR97b] F. Bassi and S. Rebay. High-Order Accurate Discontinuous Finite Element Solution of the 2D Euler Equations. *Journal of Computational Physics*, 138(2):251 – 285, 1997. 4.1

[Bra92] D. Braess. *Finite Elemente: Theorie, schnelle Löser und Anwendungen in der Elastizitätstheorie*. Springer-Lehrbuch. Springer, 1992. 4.1

[Bre] A. Bressan. Notes on the Boltzmann Equation. Lecture notes for a summer course given at S.I.S.S.A., Trieste. 2.2.1

[Bue96] C. Buet. A Discrete-Velocity Scheme for the Boltzmann Operator of Rarefied Gas Dynamics. *Transport Theory and Statistical Physics*, 25(1):33–60, 1996. 3.2.2, 3.2.2, 3.2.4

[Cer90] C. Cercignani. *Mathematical Methods in Kinetic Theory*. Plenum Press, New York, second edition, 1990. 1.1.1, 1.1.1, 1.1.1, 2, 2.1, 2.2, 2.2.1, 2.2.2, 2.3.2, 2.3.2, 2.4

[CFTV10] J. Carrillo, M. Fornasier, G. Toscani, and F. Vecil. Particle, kinetic, and hydrodynamic models of swarming. In G. Naldi, L. Pareschi, and G. Toscani, editors, *Mathematical Modeling of Collective Behavior in Socio-Economic and Life Sciences*, Modeling and Simulation in Science, Engineering and Technology, pages 297–336. Birkhuser Boston, 2010. 1.1.2

[CHS90] B. Cockburn, S. Hou, and C.-W. Shu. THE RUNGE KUTTA LOCAL PROJECTION DISCONTINUOUS GALERKIN FINITE ELEMENT METHOD FOR CONSERVATION LAWS IV: THE MULTIDIMENSIONAL CASE. *Mathematics of Computation*, 54(190):545–581, 1990. 4.1

[CIP94] C. Cercignani, R. Illner, and M. Pulvirenti. *The Mathematical Theory of Dilute Gases*, volume 106 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1994. 1.1.1, 1.1.1, 1.1.1, 1.1.1, 1.2.1, 2, 2.2.1, 2.2.1, 2.3.2, 2.4.1

[CKS00] B. Cockburn, G. Karniadakis, and C.-W. Shu. The development of discontinuous Galerkin methods. In B. Cockburn, G. Karniadakis, and C.-W. Shu, editors, *Discontinuous Galerkin Methods*, volume 11 of *Lecture Notes in Computational Science and Engineering*, pages 3–50. Springer Berlin Heidelberg, 2000. 4.1

[CLS89] B. Cockburn, S.-Y. Lin, and C.-W. Shu. TVB Runge-Kutta Local Projection Discontinuous Galerkin Finite Element Method for Cnservation Laws III: One-Dimensional Systems. *Journal of Computational Physics*, 84(1):90–113, 1989. 4.1

[CS89] B. Cockburn and C.-W. Shu. TVB Runge-Kutta Local Projection Discontinuous Galerkin Finite Element Method for Conservation Laws. II. General Framework. *Mathematics of computation*, 52(186):411–435, 1989. 4.1

[CS01] B. Cockburn and C.-W. Shu. Runge-Kutta Discontinuous Galerkin Methods for Convection-Dominated Problems. *Journal of scientific computing*, 16(3):173–261, 2001. 4.1

[DDCS12] B. A. De Dios, J. A. Carillo, and C.-W. Shu. Discontinuous Galerkin methods for the Multi dimensional Vlasov-Poisson problem. *Mathematical Models and Methods in Applied Sciences*, 22(12):1250042, 2012. 1.3, 4.3.6

[DM98] L. Dagum and R. Menon. OpenMP: an industry standard api for shared-memory programming. *Computational Science & Engineering, IEEE*, 5(1):46–55, 1998. 5

[EE99] A. Y. Ender and I. A. Ender. Polynomial expansions for the isotropic Boltzmann equation and invariance of the collision integral with respect to the choice of basis functions. *Phys. Fluids*, 11(9):2720–2730, 1999. 4.3.6

[EG04] A. Ern and J.-L. Guermond. *Theory and Practice of Finite Elements*. Applied mathematical sciences. Springer, New York, 2004. 4.1

[Ern84] M. H. Ernst. Exact Solutions of the Nonlinear Boltzmann Equation. *Journal of Statistical Physics*, 34(5-6):1001–1017, 1984. 5.1.1

[Fal00] R. Falk. ANALYSIS OF FINITE ELEMENT METHODS FOR LINEAR HYPERBOLIC PROBLEMS. In B. Cockburn, G. Karniadakis, and C.-W. Shu, editors, *Discontinuous Galerkin Methods*, volume 11 of *Lecture Notes in Computational Science and Engineering*, pages 103–112. Springer Berlin Heidelberg, 2000. 4.1

[FGH14] E. Fonn, P. Grohs, and R. Hiptmair. Polar Spectral Scheme for the Spatially Homogeneous Boltzmann Equation. Technical Report 2014-13, Seminar for Applied Mathematics, ETH Zürich, Switzerland, 2014. 4.3.6

[FM10] F. Filbet and C. Mouhot. Analysis of spectral methods for the homogeneous Boltzmann equation. *Transactions of the American Mathematical Society*, page To appear, 2010. 41 pages. 3.2

[GPT97] E. Gabetta, L. Pareschi, and G. Toscani. RELAXATION SCHEMES FOR NONLINEAR KINETIC EQUATIONS. *SIAM Journal on Numerical Analysis*, 34(6):2168–2194, 1997. 3.1.2

[Gra49] H. Grad. On the Kinetic Theory of Rarefied Gases. *Communications on Pure and Applied Mathematics*, 2(4):331–407, 1949. 1.3

[Gra58] H. Grad. Principles of the Kinetic Theory of Gases. In S. Flügge, editor, *Thermodynamik der Gase / Thermodynamics of Gases*, volume 3 / 12 of *Handbuch der Physik / Encyclopedia of Physics*, pages 205–294. Springer Berlin Heidelberg, 1958. 1.3

[HGMM12] R. E. Heath, I. M. Gamba, P. J. Morrison, and C. Michler. A discontinuous Galerkin method for the Vlasov-Poisson system. *J. Comput. Phys.*, 231(4):1140–1174, 2012. 1.3, 4.3.6

[HW79] G. Hardy and E. Wright. *AN INTRODUCTION TO THE THEORY OF NUMBERS*. Oxford science publications. Clarendon Press, 1979. 3.2.2

[HW08] J. Hesthaven and T. Warburton. *Nodal Discontinuous Galerkin Methods: Algorithms, Analysis, and Applications*. Texts in Applied Mathematics. Springer, 2008. 4.1

[Joh12] C. Johnson. *Numerical Solution of Partial Differential Equations by the Finite Element Method*. Dover Books on Mathematics Series. Dover Publications, Incorporated, 2012. 4.1

[JP86] C. Johnson and J. Pitkäranta. An Analysis of the Discontinuous Galerkin Method for a Scalar Hyperbolic Equation. *Math. Comp.*, 46(173):1–26, 1986. 4.1

[Jün09] A. Jüngel. *Transport Equations for Semiconductors*. Lecture Notes in Physics. Springer Berlin Heidelberg, 2009. 1.1.2

[JWC65] J. W. T. James W. Cooley. An Algorithm for the Machine Calculation of Complex Fourier Series. *Mathematics of Computation*, 19(90):297–301, 1965. 3.2.1

[Kru67] R. S. Krupp. *A Nonequilibrium Solution of the Fourier Transformed Boltzmann Equation*. PhD thesis, MIT, 1967. 5.1.1

[KS13] G. Kitzler and J. Schöberl. Efficient Spectral Methods for the spatially homogeneous Boltzmann equation. Technical Report 13, Institute for Analysis and Scientific Computing, Vienna UT, 2013. 4.3

[KS15] G. Kitzler and J. Schöberl. A high-order space momentum discontinuous Galerkin method for the Boltzmann equation. *Computers and Mathematics with Applications*, 70(7):1539 – 1554, 2015. High-Order Finite Element and Isogeometric Methods. 4.3

[LR74] P. Lesaint and P. Raviart. *On a Finite Element Method for Solving the Neutron Transport Equation*. Univ. Paris VI, Labo. Analyse Numérique, 1974. 4.1

[Nan80] K. Nanbu. Direct Simulation Scheme Derived from the Boltzmann Equation. I. Monocomponent Gases. *Journal of the Physical Society of Japan*, 49(5):2042–2049, 1980. 3.1, 3.1.1

[PH02] V. A. Panferov and A. G. Heintz. A new consistent-discrete velocity model for the Boltzmann equation. *Math. Methods Appl. Sci.*, 25(7):571–593, 2002. 3.2.5

[Pla10] R. Plato. *Numerische Mathematik kompakt: Grundlagenwissen für Studium und Praxis*. Vieweg Studium. Vieweg+Teubner Verlag, 2010. 4.1.3

[PP96] L. Pareschi and B. Perthame. A Fourier Spectral Method for Homogeneous Boltzmann Equations. *Transport Theory and Statistical Physics*, 25(3-5):369–382, 1996. 3.2, 3.2.1, 3.2.3

[PR] L. Pareschi and G. Russo. On the stability of spectral methods for the homogeneous Boltzmann equation. In *Proceedings of the Fifth International Workshop on Mathematical Aspects of Fluid and Plasma Dynamics (Maui, HI*, pages 369–383. 3.2

[PR00] L. Pareschi and G. Russo. NUMERICAL SOLUTION OF THE BOLTZMANN EQUATION I:SPECTRALLY ACCURATE APPROXIMATION OF THE COLLISION OPERATOR. *SIAM J. Numer. Anal.*, 37(4):1217–1245, 2000. 3.2, 3.2.1, 3.2.1, 3.2.1, 3.2.1, 3.2.1

[PT05] L. Pareschi and S. Trazzi. NUMERICAL SOLUTION OF THE BOLTZMANN EQUATION BY TIME RELAXED MONTE CARLO (TRMC) METHODS. *Int. J. Numer. Meth. Fluids*, pages 947–983, 2005. 3.1.2, 3.1.2

[RH73] W. H. Reed and T. R. Hill. TRIANGULAR MESH METHODS FOR THE NEUTRON TRANSPORT EQUATION. *Proceedings of the American Nuclear Society*, 1973. 4.1

[RSZ01] C. Ringhofer, C. Schmeiser, and A. Zwirchmayr. MOMENT METHODS FOR THE SEMICONDUCTOR BOLTZMANN EQUATION ON BOUNDED POSITION DOMAINS. *SIAM Journal on Numerical Analysis*, 39(3):1078–1095, 2001. 1.3

[RT77] P.-A. Raviart and J. M. Thomas. A MIXED FINITE ELEMENT METHOD FOR 2nd ORDER ELLIPTIC PROBLEMS. In *Mathematical aspects of finite element methods (Proc. Conf., Consiglio Naz. delle Ricerche (C.N.R.), Rome, 1975)*, pages 292–315. Lecture Notes in Math., Vol. 606. Springer, Berlin, 1977. 4.4

[RTS13] A. Rana, M. Torrilhon, and H. Struchtrup. A robust numerical method for the R13 equations of rarefied gas dynamics: Application to lid driven cavity. *Journal of Computational Physics*, 236:169 – 186, 2013. 1.3

[RW05] S. Rjasanow and W. Wagner. *Stochastic Numerics for the Boltzmann Equation*, volume 37 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 2005. 1.1.1, 2, 2.2, 3.1

[Sch] J. Schöberl. NGSOLVE Finite Element Library. 5

[Sch14] J. Schöberl. C++11 Implementation of Finite Elements in NGSolve. Technical Report 30, Institute for Analysis and Scientific Computing, Vienna UT, 2014. 5

[ST03] H. Struchtrup and M. Torrilhon. Regularization of Grad's 13 moment equations: Derivation and linear analysis. *Physics of Fluids*, 15(9):2668–2680, 2003. 1.3

[STW11] J. Shen, T. Tang, and L.-L. Wang. *Spectral Methods: Algorithms, Analysis and Applications*, volume 41. Berlin: Springer-Verlag, 2011. 4.1.3, 4.3.3

[Wil51] E. Wild. ON BOLTZMANN'S EQUATION IN THE KINETIC THEORY OF GASES. *Mathematical Proceedings of the Cambridge Philosophical Society*, 47:602–609, 7 1951. 3.1.2

[ZZ92a] O. C. Zienkiewicz and J. Z. Zhu. The superconvergent patch recovery and a posteriori error estimates. Part 1: The recovery technique. *International Journal for Numerical Methods in Engineering*, 33(7):1331–1364, 1992. 5.2.1, 5.2.3

[ZZ92b] O. C. Zienkiewicz and J. Z. Zhu. The superconvergent patch recovery and a posteriori error estimates. Part 2: Error estimates and adaptivity. *International Journal for Numerical Methods in Engineering*, 33(7):1365–1382, 1992. 5.2.1, 5.2.3