

Tensor-product discretization for the
spatially inhomogeneous and transient
Boltzmann equation in 2D

P. Grohs and R. Hiptmair and S. Pintarelli

Research Report No. 2015-38

November 2015

Latest revision: April 2016

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

Tensor-product discretization for the spatially inhomogeneous and transient Boltzmann equation in $2D$

P. Grohs R. Hiptmair S. Pintarelli

Abstract

In this paper we extend the previous work [E. FONN, P. GROHS, AND R. HIPTMAIR, *Polar spectral scheme for the spatially homogeneous Boltzmann equation*, Tech. Rep. 2014-13, Seminar for Applied Mathematics, ETH Zürich, 2014.] for the homogeneous nonlinear Boltzmann equation to the spatially inhomogeneous case. We consider a tensor-product discretization of the distribution function combining Laguerre polynomials times a Maxwellian in velocity with continuous, first order finite elements in the spatial domain. The advection problem in phase space is discretized through a Galerkin least squares technique and yields an implicit formulation in time. The discrete collision operators can be evaluated with an asymptotic effort of $\mathcal{O}(K^5)$, where K is the number of velocity degrees of freedom in a single direction. Numerical results in 2D are presented for different Mach and Knudsen numbers.

1 Introduction

The Boltzmann equation offers a mesoscopic description of rarefied gases and is a typical representative of a class of integro partial differential equations that model interacting particle systems. The binary particle interactions in d -dimensional space are modeled by a collision operator which involves a $2d - 1$ fold integral. Due to its non-linearity and the high dimension, the evaluation of the collision operator is computationally challenging. Stochastic simulation methods are widely used. A well-known example is the direct simulation Monte Carlo (DSMC) method developed by Bird and Nanbu in [1] and [2]. Among deterministic approaches Fourier methods are most popular. In [3] Pareschi et al. introduced a Fourier based method, related approaches have been introduced in [4–7]. Fourier methods are fairly efficient and accurate for short-time simulations, but they suffer from aliasing errors caused by the periodic truncation of the velocity domain.

To overcome this problem a spectral discretization in velocity based on Laguerre polynomials has been developed in [8] for the spatially homogeneous Boltzmann equation extending the work done in [9]. No truncation of the velocity domain is necessary. This approach has the advantage that the collision operator can be represented as a tensor, which enjoys considerable sparsity and whose entries can be precomputed with highly accurate quadrature.

In this work, we extend this idea to the spatially inhomogeneous Boltzmann equation, combining a truncation-free spectral Galerkin approximation in velocity with a least squares stabilized finite element discretization on the spatial domain. The tensor based local evaluation of the discrete collision operator involves an asymptotic computational effort of $\mathcal{O}(K^5)$, where K is the polynomial degree in one velocity direction, see Sec. 3. We also explore ways to ensure discrete conservation of mass, momentum, and energy, see Sec. 3.2. This can be achieved by modifying a few trial functions in the spirit of a Petrov-Galerkin discretization. An alternative is the direct enforcement of the constraints through Lagrangian multipliers.

In Sec. 5 we elaborate how to incorporate various physically relevant spatial boundary conditions into our new scheme.

For time-stepping we rely on a splitting scheme, which separately treats collisions and advection. For the former we opt for explicit time-stepping, whereas the latter is tackled by a time-implicit least squares formulation. This has the advantage, that for high Knudsen numbers we are not restricted by a CFL condition. However, one must note that for small Knudsen numbers, i.e. small mean free path length, the problem is stiff and the time-step must be chosen sufficiently small. Extensive numerical tests in various settings typical of flow problems for rarefied gases are reported in Sec. 6.

Closely related and conducted parallel to our investigations is the work by Kitzler and Schöberl [10, 11]. These authors also use a spectral polynomial discretization in velocity, but they rely on a Petrov-Galerkin discretization. The velocity distribution function (VDF) is represented by polynomials times a shifted Maxwellian, while the test functions are polynomials. The complexity for the evaluation of the collision operator is reduced from $\mathcal{O}(K^6)$ to $\mathcal{O}(K^5)$ by exploiting its translation invariance properties. They locally rescale the basis functions in velocity to fit macroscopic velocity and temperature. In physical space Kitzler and Schöberl use a discontinuous Galerkin scheme. On the one hand this offers great flexibility concerning the local choice of velocity spaces. On the other hand the DG method involves evaluating interface fluxes and thus requires projection of the velocity distribution function between adjacent elements. Then stability issues impose constraints on the temperature differences between neighboring elements.

1.1 The Boltzmann equation

The time-dependent distribution function $f = f(\mathbf{x}, \mathbf{v}, t)$ is sought on the $2 + 2$ -dimensional phase space $\Omega = D \times \mathbb{R}^2$, where D denotes a spatial domain with piecewise smooth boundary.

We consider the inhomogeneous and time dependent Boltzmann equation

$$\partial_t f + \mathbf{v} \cdot \nabla_{\mathbf{x}} f = \frac{1}{kn} Q(f, f)(\mathbf{v}), \quad (\mathbf{x}, \mathbf{v}) \in \Omega = D \times \mathbb{R}^2, \quad (1)$$

with initial distribution

$$f(\mathbf{x}, \mathbf{v}, t = 0) = f_0(\mathbf{x}, \mathbf{v}). \quad (2)$$

The Knudsen number kn represents the mean free path in its nondimensional form. Boundary conditions are prescribed on the inflow boundary $\Gamma^- := \{(\mathbf{x}, \mathbf{v}) : \mathbf{x} \in \partial D \wedge \mathbf{v} \cdot \mathbf{n} \leq 0\}$, where \mathbf{n} denotes the outward unit normal vector. Common types of boundary conditions are inflow, specular reflective and diffusive reflective boundary conditions [12, Sec. 1.5].

Inflow boundary conditions

$$f(t, \mathbf{x}, \mathbf{v}) = f_{in}(t, \mathbf{x}, \mathbf{v}), \quad (\mathbf{x}, \mathbf{v}) \in \Gamma^- \quad (3)$$

Specular reflective boundary conditions

$$f(t, \mathbf{x}, \mathbf{v}) = f(t, \mathbf{x}, \mathbf{v} - 2\mathbf{v} \cdot \mathbf{nn}), \quad (\mathbf{x}, \mathbf{v}) \in \Gamma^- \quad (4)$$

(The particles behave like billiard balls at the wall.)

Diffusive reflective boundary conditions The particles are absorbed at the wall and re-emitted with Maxwellian distribution $M_w(\mathbf{x}, \mathbf{v})$.

$$f(t, \mathbf{x}, \mathbf{v}) = M_w(\mathbf{x}, \mathbf{v}) \rho_+(f), \quad (\mathbf{x}, \mathbf{v}) \in \Gamma^- \quad (5)$$

where

$$M_w(\mathbf{x}, \mathbf{v}) := \left(\frac{1}{2\pi} \right)^{\frac{1}{2}} T_w^{\frac{3}{2}} e^{-\frac{\|\mathbf{v}\|^2}{2T_w}}, \quad (6)$$

is a Maxwellian distribution at the boundary, which may depend on \mathbf{x} implicitly through the wall temperature $T_w(\mathbf{x})$, and

$$\rho_+(f) := \int_{\Gamma^+} \mathbf{n} \cdot \mathbf{w} f(t, \mathbf{x}, \mathbf{w}) d\mathbf{w}.$$

M_w is normalized such that $\int_{\Gamma^+} \mathbf{n} \cdot \mathbf{v} M_w(\mathbf{x}, \mathbf{v}) d\mathbf{v} = 1$. Macroscopic quantities of the gas can be computed in terms of moments of the distribution function f .

$$\begin{aligned} \text{Mass} \quad \rho(t, \mathbf{x}) &= \int_{\mathbb{R}^2} f(t, \mathbf{x}, \mathbf{v}) d\mathbf{v} \\ \text{Momentum} \quad \mathbf{u}(t, \mathbf{x}) &= \frac{1}{\rho} \int_{\mathbb{R}^2} \mathbf{v} f(t, \mathbf{x}, \mathbf{v}) d\mathbf{v} \\ \text{Energy} \quad E(t, \mathbf{x}) &= \frac{1}{\rho} \int_{\mathbb{R}^2} \|\mathbf{v}\|^2 f(t, \mathbf{x}, \mathbf{v}) d\mathbf{v} \\ \text{Temperature} \quad T(t, \mathbf{x}) &= \frac{1}{2} (E(t, \mathbf{x}) - \|\mathbf{u}(t, \mathbf{x})\|^2) \end{aligned}$$

The Boltzmann collision operator Q in $2D$ is represented by a 3 fold integral:

$$Q(f, h)(\mathbf{v}) = \int_{\mathbb{R}^d} \int_{\mathbb{S}^{d-1}} B(\|\mathbf{v} - \mathbf{v}_*\|, \cos \theta) (h'_* f' - h_* f) d\sigma d\mathbf{v}_* \quad (7)$$

It is common to split Q into gain Q^+ and loss Q^- part

$$Q^+(f, h)(\mathbf{v}) = \int_{\mathbb{R}^d} \int_{\mathbb{S}^{d-1}} B(\|\mathbf{v} - \mathbf{v}_*\|, \cos \theta) h'_* f' d\sigma d\mathbf{v}_* \quad (8)$$

$$Q^-(f, h)(\mathbf{v}) = \int_{\mathbb{R}^d} \int_{\mathbb{S}^{d-1}} B(\|\mathbf{v} - \mathbf{v}_*\|, \cos \theta) h_* f d\sigma d\mathbf{v}_*, \quad (9)$$

where $f = f(\mathbf{v})$, $f' = f(\mathbf{v}')$, $h_* = h(\mathbf{v}_*)$, $h'_* = h(\mathbf{v}'_*)$. For elastic scattering, the post-collisional velocities \mathbf{v}' , \mathbf{v}'_* are given by, see Fig. 1:

$$\begin{aligned} \mathbf{v}' &= \frac{\mathbf{v} + \mathbf{v}_*}{2} + \sigma \frac{\|\mathbf{v} - \mathbf{v}_*\|}{2} \\ \mathbf{v}'_* &= \frac{\mathbf{v} + \mathbf{v}_*}{2} - \sigma \frac{\|\mathbf{v} - \mathbf{v}_*\|}{2} \end{aligned} \quad \sigma \in \mathbb{S}^1. \quad (10)$$

We assume that the interaction potential governing collisions is described by the collision kernel B of the form [12]:

$$B(\|\mathbf{v} - \mathbf{v}_*\|, \cos \theta) = C(\cos \theta) \|\mathbf{v} - \mathbf{v}_*\|^\lambda, \quad (11)$$

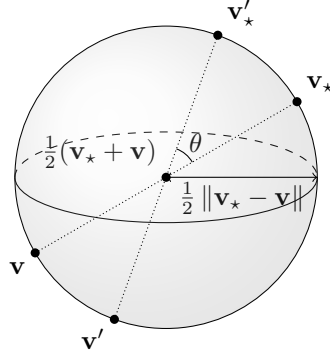


Figure 1

and that $C(\cos \theta)$ satisfies Grad's cutoff assumption [13]:

$$\int_0^{2\pi} C(\cos \theta) d\theta < \infty$$

In the following, we will restrict ourselves to the variable hard spheres model, i.e. we set $C \equiv \frac{1}{2\pi}$ and consider $\lambda \geq 0$. The case $\lambda = 0$ is known as Maxwellian molecules. In order to reduce the computational complexity we will make use of the rotational and translational invariance of the collision operator Q .

Definition 1.1 (Translation and rotation operator). *The pullbacks induced by the translation $\tau^*(\mathbf{c})$ and rotation operator $\rho^*(\omega)$ act on a function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ as follows:*

$$\begin{aligned} \tau^*(\mathbf{c})f(\mathbf{v}) &:= f(\mathbf{v} + \mathbf{c}), & \text{for } \mathbf{c} \in \mathbb{R}^2 \quad (\text{in Cartesian coordinates}) \\ \rho^*(\omega)f(\varphi, r) &:= f(\varphi + \omega, r), & \text{for } \omega \in [0, 2\pi[\quad (\text{in polar coordinates}) \end{aligned}$$

It is easy to see that the collision operator enjoys the following covariance properties:

$$Q(\rho^*(\omega)f, \rho^*(\omega)g)(\varphi, r) = \rho^*(\omega)Q(f, g)(\varphi, r) \quad (12)$$

$$Q(\tau^*(\mathbf{c})f, \tau^*(\mathbf{c})g)(\varphi, r) = \tau^*(\mathbf{c})Q(f, g)(\varphi, r), \quad (13)$$

for any $\omega \in [0, 2\pi[$ and $\mathbf{c} \in \mathbb{R}^2$.

2 Spectral Velocity Space

For discretization in the velocity coordinate, we use the *Polar-Laguerre* basis developed in [10, Sec. 2.1]. It can be shown, that the basis is equivalent to weighted polynomials in \mathbb{R}^2 of total degree $\leq K$, with weight $e^{-r^2/2}$. Throughout we designate by (φ, r) polar coordinates in \mathbb{R}^2 .

Definition 2.1 (Polar-Laguerre basis functions $\Psi_{k,j}^a(\varphi, r)$).

$$\Psi_{k,j}^a(\varphi, r) := \begin{cases} a(2j\varphi) r^{2j} L_{\frac{k}{2}-j}^{(2j)}(r^2) e^{-r^2/2} & k \in 2\mathbb{N} \\ a((2j+1)\varphi) r^{2j+1} L_{\frac{k-1}{2}-j}^{(2j+1)}(r^2) e^{-r^2/2} & k \in 2\mathbb{N} + 1 \end{cases} \quad (14)$$

where $a = \cos, \sin$ and $L_n^{(\alpha)}$ are the associated Laguerre polynomials.

The basis functions $\Psi_{k,j}$ are orthogonal in the inner product $\langle f, g \rangle := \int_{\mathbb{R}^2} f(\mathbf{v})g(\mathbf{v}) \, d\mathbf{v}$ [14, Chap. 22]. We define the spectral basis \mathfrak{B}_V^N of maximal polynomial degree K and total number of elements N :

$$\mathfrak{B}_V^N := \{\mathbb{L}_k^{\cos} : k = 0, \dots, K\} \cup \{\mathbb{L}_k^{\sin} : k = 0, \dots, K\}, \quad (15)$$

where

$$\begin{aligned} \mathbb{L}_k^{\cos} &:= \{\Psi_{k,j}^{\cos} : j = 0 \dots \lfloor \frac{k}{2} \rfloor\} \\ \mathbb{L}_k^{\sin} &:= \{\Psi_{k,j}^{\sin} : j = 1 - (k \bmod 2) \dots \lfloor \frac{k}{2} \rfloor\}. \end{aligned} \quad (16)$$

For later usage we define the function space $V_V^N := \text{span}\{\mathfrak{B}_V^N\}$.

Notation: Unless specified, N will always denote the number of basis functions used to discretize the velocity domain and has therefore been included in the superscript of the symbols \mathfrak{B}_V^N and V_V^N .

Remark 2.2. In [8], the test and trial functions in radial direction have the following form:

$$\Psi_k(r) = e^{-r^2/2} \begin{cases} \sqrt{2} L_{\frac{k}{2}}^{(0)}(r^2) & k \text{ even} \\ \sqrt{\frac{1}{k+1}} r L_{\frac{k-1}{2}}^{(1)}(r^2) & k \text{ odd} \end{cases}$$

The Ψ_k , for $k = 0, \dots, K$ are then combined with the Fourier modes $e^{il\varphi}$ in angle, for $l = 0, \dots, L$, such that $k \equiv l \pmod{2}$. Consider for example $k = 1, l = 1$:

$$e^{il\varphi} \Psi_1(r) = \sqrt{\frac{1}{2}} e^{il\varphi} e^{-r^2/2},$$

which is singular at $r = 0, \varphi \in [0, 2\pi[$. The same problem appears for all $k \leq l$ and causes rapidly oscillating line integrals during the assembly of the collision tensor entries.

Lemma 2.3. *In Cartesian coordinates, the Polar-Laguerre basis functions $\Psi_{k,j}^{\cos, \sin}$ are polynomials of total degree k weighted by $e^{-r^2/2}$.*

Proof. [10, Lemma 5] Use that

$$\cos n\varphi = \sum_{j=0}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{2j} \sin(\varphi)^{2j} \cos(\varphi)^{n-2j}, \quad \sin n\varphi = \sum_{j=0}^{\lfloor \frac{n-1}{2} \rfloor} \binom{n}{2j+1} \sin(\varphi)^{2j+1} \cos(\varphi)^{n-2j-1} \quad (17)$$

For k even:

$$\begin{aligned} \Psi_{k,j}^{\cos} e^{r^2/2} &= \sum_{i=0}^j \binom{2j}{2i} \sin(\varphi)^{2j} \cos(\varphi)^{2j-2i} r^{2j} L_{\frac{k}{2}-j}^{(2j)}(r^2) \\ &= \sum_{i=0}^j \binom{2j}{2i} (r \sin(\varphi))^{2j} (r \cos(\varphi))^{2j-2i} L_{\frac{k}{2}-j}^{(2j)}(r^2) \\ &= \sum_{i=0}^j \binom{2j}{2i} y^{2j} x^{2j-2i} L_{\frac{k}{2}-j}^{(2j)}(r^2) \end{aligned} \quad (18)$$

$L_{\frac{k}{2}-j}^{(2j)}(r^2)$ is a polynomial of total degree $2(\frac{k}{2} - j)$ and thus multiplication with $y^{2j} x^{2j-i}$ yields a polynomial of total degree k . \square

Whenever convenient, we will drop the double index (k, j) of $\Psi_{k,j}$ and denote elements of \mathfrak{B}_V^N by $b_i, i = 0, \dots, N - 1$. Thus we may formally write the expansion f^P with Polar-Laguerre coefficients $c_j^P, j = 0, \dots, N - 1$, of a function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$:

$$f^P(\varphi, r) = \sum_{j=0}^{N-1} c_j^P b_j(\varphi, r). \quad (19)$$

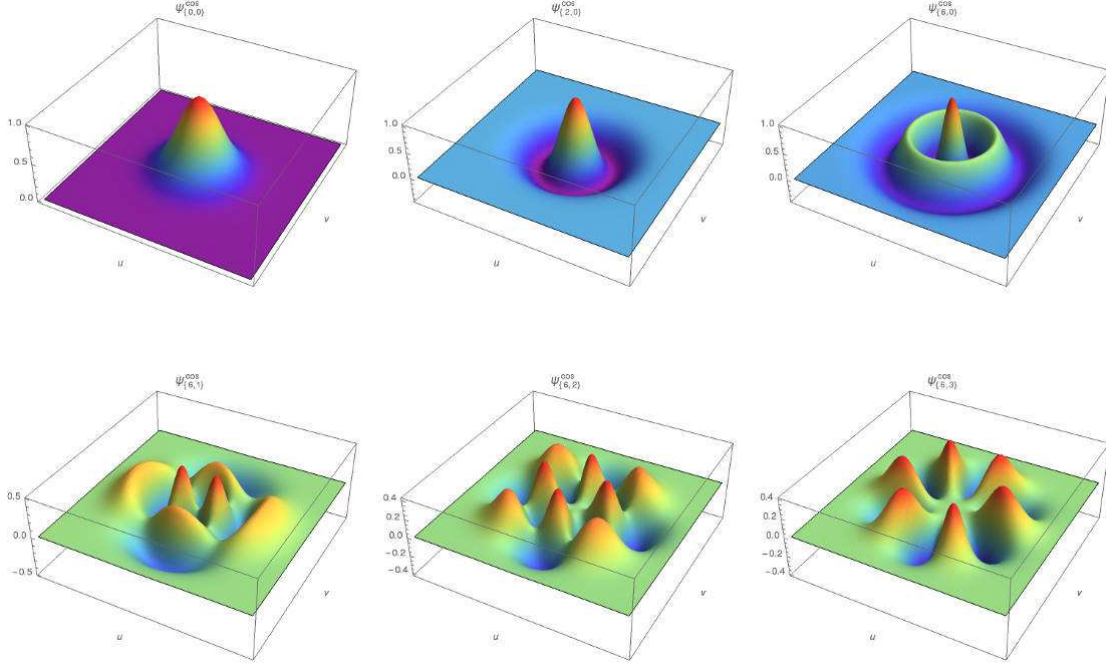


Figure 2: Polar-Laguerre basis functions $\Psi_{k,j}^{\text{cos}}(\mathbf{v})$, $\mathbf{v} \in [-5, 5]^2$.
First row: $j = 0, k = 0, 2, 6$, **Second row:** $k = 6, j = 1, 2, 3$.

Definition 2.4 (Hermite basis). *The expansion of a function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ in Hermite polynomials of total degree $\leq K$ reads:*

$$f^H(x, y) = \sum_{k=0}^K \sum_{s=0}^k c_{s, k-s} h_s(x) e^{-\frac{x^2}{2}} h_{k-s}(y) e^{-\frac{y^2}{2}}, \quad (20)$$

where $h_i(x)$ are suitably normalized Hermite polynomials [14], such that $\int_{\mathbb{R}} h_i(x) h_j(x) e^{-x^2} dx = \delta_{i,j}$.

Definition 2.5 (Nodal basis). *The expansion of a function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ in Lagrange polynomials of degree K reads:*

$$f^N(x, y) = \sum_{i=0}^K \sum_{j=0}^K c_{i,j}^N l_i(x) e^{-\frac{x^2}{2}} l_j(y) e^{-\frac{y^2}{2}}, \quad (21)$$

where ℓ_i denote the Lagrange polynomials at the Gauss-Hermite quadrature nodes x_i with weights w_i [14].

$$\ell_i(x) = \frac{1}{\sqrt{w_i}} \prod_{\substack{0 \leq m \leq K \\ m \neq i}} \frac{x - x_m}{x_i - x_m}.$$

We normalize the $\ell_i(x)$ such that $\langle \ell_i(x), \ell_j(x)e^{-x^2} \rangle = \delta_{i,j}$.

Notation In the following, we will tag coefficient vectors \mathbf{c} with a superscript P, H, N to indicate that they belong to the Polar-Laguerre, Hermite or the nodal basis.

3 Treatment of the Collision Operator

In this section we will discuss the discretization of the collision operator.

3.1 Discretization in velocity coordinate

The following derivation is identical to the one presented in [8], except that we use a real valued basis in φ . Consider the homogeneous Boltzmann equation

$$\partial_t f = Q(f, f). \quad (22)$$

Temporarily let \hat{V}_V^N stand for a generic function space. Specific choices will be given in Sec. 3.2. Multiplication of (22) with a test function $g \in \hat{V}_V^N$ and integration over \mathbb{R}^d gives

$$\partial_t \int_{\mathbb{R}^d} f(t, \mathbf{v}) g(\mathbf{v}) d\mathbf{v} = \int_{\mathbb{R}^d} Q(f, f) g(\mathbf{v}) d\mathbf{v}. \quad (23)$$

Making the ansatz $f \in V_V^N$ in (23) and choosing $g = \hat{b}_i$ and $\{\hat{b}_i\}_1^N$ as a basis of \hat{V}_V^N gives rise to a 3-dimensional tensor Q_N . One may think of it as an array of $N \times N$ matrices $\mathbf{S}_i, i = 0, \dots, N-1$, where slice \mathbf{S}_i is obtained by testing with $\hat{b}_i \in \hat{V}_V^N$:

$$(\mathbf{S}_i)_{i_1, i_2} := \left\langle Q(b_{i_1}, b_{i_2}), \hat{b}_i \right\rangle_{L^2(\mathbb{R}^2)}, \quad b_{i_1}, b_{i_2} \in \mathfrak{B}_V^N \quad (24)$$

We split $Q(f, f) = Q^+(f, f) - Q^-(f, f)$, as in (8) and (9), and accordingly $\mathbf{S} = \mathbf{S}^+ - \mathbf{S}^-$.

$$\begin{aligned} (\mathbf{S}_i^-)_{i_1, i_2} &= \left\langle Q^-(b_{i_1}, b_{i_2}), \hat{b}_i \right\rangle = \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} B(\|\mathbf{v} - \mathbf{v}_*\|, \cos \theta) b_{i_1}(\mathbf{v}) b_{i_2}(\mathbf{v}_*) \hat{b}_i(\mathbf{v}) d\sigma d\mathbf{v}_* d\mathbf{v} \\ &= \int_{\mathbb{R}^2} b_{i_1}(\mathbf{v}) \hat{b}_i(\mathbf{v}) \int_{\mathbb{R}^2} b_{i_2}(\mathbf{v}_*) \mathcal{I}^-(\mathbf{v}, \mathbf{v}_*) d\mathbf{v}_* d\mathbf{v} \end{aligned} \quad (25)$$

where the *inner integral* \mathcal{I}^- is given by

$$\mathcal{I}^- = \int_{\mathbb{S}^1} B(\|\mathbf{v} - \mathbf{v}_*\|, \cos \theta) d\sigma = \|\mathbf{v} - \mathbf{v}_*\|^\lambda \int_{\mathbb{S}^1} C(\cos \theta) d\sigma, \quad (26)$$

and as stated in the beginning $C \equiv \frac{1}{2\pi}$.

$$\begin{aligned} (\mathbf{S}_i^+)_{(i_1, i_2)} &= \left\langle Q^+(b_{i_1}, b_{i_2}), \hat{b}_i \right\rangle_{L^2(\mathbb{R}^2)} = \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} B(\|\mathbf{v} - \mathbf{v}_*\|, \cos \theta) b_{i_1}(\mathbf{v}') b_{i_2}(\mathbf{v}'_*) \hat{b}_i(\mathbf{v}) \, d\sigma \, d\mathbf{v}_* \, d\mathbf{v} \\ &= C \int_{\mathbb{R}^2} b_{i_1}(\mathbf{v}) \int_{\mathbb{R}^2} b_{i_2}(\mathbf{v}_*) \mathcal{I}_i^{(+)}(\mathbf{v}, \mathbf{v}_*) \, d\mathbf{v}_* \, d\mathbf{v} \end{aligned} \quad (27)$$

with

$$\mathcal{I}^{(+)}(\mathbf{v}', \mathbf{v}'_*; \hat{b}_i) = \int_{\mathbb{S}^1} B(\|\mathbf{v}' - \mathbf{v}'_*\|, \cos \theta) \hat{b}_i(\mathbf{v}') \, d\sigma, \quad (28)$$

see (10) and Fig. 1 for the definition of θ and \mathbf{v}' , \mathbf{v}'_* . Note that, in the second line of (27), we have made the change of variables $\mathbf{v}, \mathbf{v}_* \leftrightarrow \mathbf{v}', \mathbf{v}'_*$. Next, we substitute $\mathbf{w}' = R_\alpha \mathbf{v}'$ for $\alpha = -\arg(\mathbf{v} + \mathbf{v}_*)$. R_α denotes the rotation by α around the origin in counter clockwise direction. Taking the test function \hat{b}_i are from Def. 2.1, we assume that they are of the form $a(l\varphi) \phi_r(r)$, where a is either sin or cos.

$$\begin{aligned} \mathcal{I}^{(+)}(\mathbf{v}', \mathbf{v}'_*; \hat{b}_i) &= \|\mathbf{v}' - \mathbf{v}'_*\|^\lambda C \int_{\mathbb{S}^1} \hat{b}_i(\arg(\mathbf{w}') + \alpha, \|\mathbf{w}'\|) \, d\sigma \\ &= \|\mathbf{v}' - \mathbf{v}'_*\|^\lambda C \int_{\mathbb{S}^1} a(l \arg(\mathbf{w}') + \alpha) \phi_r(\|\mathbf{w}'\|) \, d\sigma \end{aligned} \quad (29)$$

We simplify (29) for $a = \sin$

$$\begin{aligned} \mathcal{I}^{(+)}(\mathbf{v}', \mathbf{v}'_*; \hat{b}_i) &= \|\mathbf{v}' - \mathbf{v}'_*\|^\lambda C \int_{\mathbb{S}^1} \left[\sin(l \arg(\mathbf{w}')) \cos(l\alpha) + \cos(l \arg(\mathbf{w}')) \sin(l\alpha) \right] \phi_r(\|\mathbf{w}'\|) \, d\sigma \\ &= \sin(l\alpha) \|\mathbf{v}' - \mathbf{v}'_*\|^\lambda C \int_{\mathbb{S}^1} \cos(l \arg(\mathbf{w}')) \phi_r(\|\mathbf{w}'\|) \, d\sigma, \end{aligned} \quad (30)$$

and for $a = \cos$

$$\begin{aligned} \mathcal{I}^{(+)}(\mathbf{v}', \mathbf{v}'_*; \hat{b}_i) &= \|\mathbf{v}' - \mathbf{v}'_*\|^\lambda C \int_{\mathbb{S}^1} \left[\cos(l \arg(\mathbf{w}')) \cos(l\alpha) - \sin(l \arg(\mathbf{w}')) \sin(l\alpha) \right] \phi_r(\|\mathbf{w}'\|) \, d\sigma \\ &= \cos(l\alpha) \|\mathbf{v}' - \mathbf{v}'_*\|^\lambda C \int_{\mathbb{S}^1} \cos(l \arg(\mathbf{w}')) \phi_r(\|\mathbf{w}'\|) \, d\sigma. \end{aligned} \quad (31)$$

Thus we have found that, up to a factor, the integral $\mathcal{I}^+(\mathbf{v}', \mathbf{v}'_*; \hat{b}_i)$, which is cheap to compute, depends only on $d := \|\mathbf{v}' - \mathbf{v}'_*\|$ and on $c := \|\mathbf{v}' + \mathbf{v}'_*\|$.

3.2 Conservative discretization

An important property of (22) is that mass, momentum and energy are conserved. In particular it holds that

$$\partial_t \begin{pmatrix} \rho(f) \\ \rho \mathbf{u}(f) \\ \rho E(f) \end{pmatrix} = \int_{\mathbb{R}^2} Q(f, f) \begin{pmatrix} 1 \\ \mathbf{v} \\ \|\mathbf{v}\|^2 \end{pmatrix} \, d\mathbf{v} \equiv 0, \quad (32)$$

by fundamental properties of the Boltzmann collision operator [15, sec. 5]. In the following we present two options for conservative time-stepping schemes for the homogeneous Boltzmann equation (22).

Option I: Petrov-Galerkin discretization Condition (32) can be naturally enforced for the ordinary differential equation (23) by choosing the test space \hat{V}_V^N such that it contains $1, \mathbf{v}, \|\mathbf{v}\|^2$. Inspection of the basis functions from \mathfrak{B}_V^N reveals that it is sufficient to multiply a few of them by a factor $e^{r^2/2}$ to conserve mass, momentum and energy:

$$\begin{aligned} \Psi_{0,0}^{\cos} \exp(r^2/2) &= 1, & \Psi_{1,0}^{\sin} \exp(r^2/2) &= \sin(\varphi) r \\ \Psi_{1,0}^{\cos} \exp(r^2/2) &= \cos(\varphi) r & \Psi_{2,0}^{\cos} \exp(r^2/2) &= (1 - r^2) \end{aligned} \quad (33)$$

Therefore we use a test space \hat{V}_V^N which is identical to V_V^N , except that $\Psi_{0,0}^{\cos}, \Psi_{1,0}^{\sin}, \Psi_{1,0}^{\cos}, \Psi_{2,0}^{\cos}$ have been multiplied by the weight $\exp(r^2/2)$. The discretized collision operator Q_N has the following expansion into basis functions:

$$Q_N(f^P, g^P)(\mathbf{v}) = \sum_{i=1}^N \left(\mathbf{M}^{-1} [\mathbf{c}^T \mathbf{S}_j \mathbf{d}]_{j=1}^N \right)_i b_i(\mathbf{v}), \quad (34)$$

where $(\mathbf{M})_{j,j'} = \langle b_j, b_{j'} \rangle$, $f^P, g^P \in V_V^N$ with coefficient vectors \mathbf{c}, \mathbf{d} with respect to the basis. The mass matrix \mathbf{M} is *diagonal*, except for dense blocks in the rows corresponding to $\{\Psi_{0,0}^{\cos}, \Psi_{1,0}^{\sin}, \Psi_{1,0}^{\cos}, \Psi_{2,0}^{\cos}\} \times e^{r^2/2}$ of size at most $1 \times K$. For now, the cost for applying Q_N is $\mathcal{O}(K^6)$. In the next section we show that, due to the polar representation, the complexity can actually be reduced by a factor K .

Option II: Galerkin discretization with Lagrange multipliers Alternatively one can also use a Galerkin discretization and solve a constrained minimization problem with respect to the L_2 -norm such that mass, momentum and energy are conserved. This has been proposed in [6] for the Fourier-spectral method. In the context of a time-stepping method, let \mathbf{c}^k be the coefficient vector in the Polar-Laguerre basis at time t_k .

1. Compute coefficients at the next time-step by a single step of an explicit time-stepping scheme, here explicit Euler:

$$\tilde{\mathbf{c}}^{k+1} = \mathbf{c}^k + \Delta t_k Q^N(\mathbf{c}^k, \mathbf{c}^k)$$

2. Solve the saddle point problem:

$$\mathbf{c}^{k+1} = \arg \min_{\mathbf{c}_*^{k+1} \in \mathbb{R}^N} \left\| \mathbf{c}_*^{k+1} - \tilde{\mathbf{c}}^{k+1} \right\|^2 + \underbrace{\lambda^T \mathbf{H}^T (\mathbf{c}_*^{k+1} - \tilde{\mathbf{c}}^{k+1})}_{\text{conservation of mass, momentum and energy}}, \quad (35)$$

where $\mathbf{H}^T \in \mathbb{R}^{2+2 \times N}$, $\mathbf{H}^T \mathbf{c} = (\rho, \rho \mathbf{u}, \rho E)^T$, with Lagrange multiplier $\lambda \in \mathbb{R}^{2+2}$. The entries of \mathbf{H}^T are given by:

$$\left. \begin{aligned} [\mathbf{H}^T]_{1,i} &= \int_{\mathbb{R}^2} b_i(\mathbf{v}) \, d\mathbf{v} \\ [\mathbf{H}^T]_{2,i} &= \int_{\mathbb{R}^2} \mathbf{v}_x b_i(\mathbf{v}) \, d\mathbf{v} \\ [\mathbf{H}^T]_{3,i} &= \int_{\mathbb{R}^2} \mathbf{v}_y b_i(\mathbf{v}) \, d\mathbf{v} \\ [\mathbf{H}^T]_{4,i} &= \int_{\mathbb{R}^2} \|\mathbf{v}\|^2 b_i(\mathbf{v}) \, d\mathbf{v}. \end{aligned} \right\} \text{for } b_i \in \mathfrak{B}_V^N, i = 1, \dots, N \quad (36)$$

The solution to (35) is

$$\mathbf{c}^{k+1} = \tilde{\mathbf{c}}^{k+1} - \frac{1}{2}\mathbf{H}\lambda, \quad (37)$$

with

$$\lambda = 2(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T(\tilde{\mathbf{c}}^{k+1} - \mathbf{c}^k). \quad (38)$$

Also note that $\mathbf{H}^T\mathbf{H}$ is positive definite.

3.3 Computational aspects

We repeat the definition 2.1 of the Polar-Laguerre basis functions $\Psi_{k,j}$, and, for the sake of simplicity in the current discussion, replace the real valued Fourier modes by their complex counterparts:

$$\Psi_{k,j}(\varphi, r) := \begin{cases} e^{i2j\varphi} r^{2j} L_{\frac{k}{2}-j}^{(2j)}(r^2) e^{-r^2/2} & k \in 2\mathbb{N} \\ e^{i(2j+1)\varphi} r^{2j+1} L_{\frac{k-1}{2}-j}^{(2j+1)}(r^2) e^{-r^2/2} & k \in 2\mathbb{N} + 1 \end{cases} \quad (39)$$

First, we observe that the $\Psi_{k,j}$'s are of the form $f_\varphi(l\varphi) f_r(r)$ with angular frequency $l \in \mathbb{Z}$.

Corollary 3.1. *Let f and g be represented in polar coordinates as*

$$f(r, \varphi) = f_r(r) e^{ik\varphi}, \quad g(r, \varphi) = g_r(r) e^{il\varphi}$$

for some functions f_r, g_r and $l, k \in \mathbb{Z}$. Then,

$$Q(f, g)(r, \varphi) = C(r) e^{-i(k+l)\varphi} \quad (40)$$

Proof. [8] We get $\rho^*(\omega)f = e^{ik\omega}f$, and correspondingly for g . Using (12) and the bilinearity of Q we obtain

$$\rho_\omega Q(f, g)(r, \varphi) = e^{i(k+l)\omega} Q(f, g)(r, \varphi). \quad (41)$$

Choose $\omega = -\varphi$ and rearrange to find

$$Q(f, g)(r, \varphi) = e^{-i(k+l)\varphi} \rho_\varphi Q(f, g)(r, \varphi).$$

The result follows since $\rho_\varphi Q(f, g)(r, \varphi) = Q(f, g)(r, 0)$ is independent of φ . \square

As a direct consequence of Cor. 3.1, in the complex Fourier basis, the collision tensor contains nonzero entries for $l + k = j$ only, where l, k and j are the angular frequencies of the trial function and the test function respectively. In the real valued Fourier basis, we have nonzero entries for $k+l = j$ or $|k-l| = j$ only. The derivation can be found in Appendix 7.1.

Corollary 3.2. *The consequence of 3.1 is that each \mathbf{S}_i from (34) only has $\mathcal{O}(K^3)$ nonzero entries, and therefore the tensor representation of Q_N has $\mathcal{O}(K^5)$ nonzero entries.*

Quadrature is carried out in polar coordinates. We use Gauss quadrature nodes and weights in the radial direction r on the interval $[0, \infty]$ with weight $r e^{-r^2/2}$, which are computed via the Golub-Welsch algorithm [16]. Recursion formulas for the coefficients contained in the Jacobi matrix can be found in [17]. Due to numerical instabilities, both the recursion formulas and the eigenvalue problem have to be computed with extended precision. We compute the quadrature nodes and weights with 128 digit accuracy, which is sufficient for degrees up to order ≈ 100 , and store them in tables.

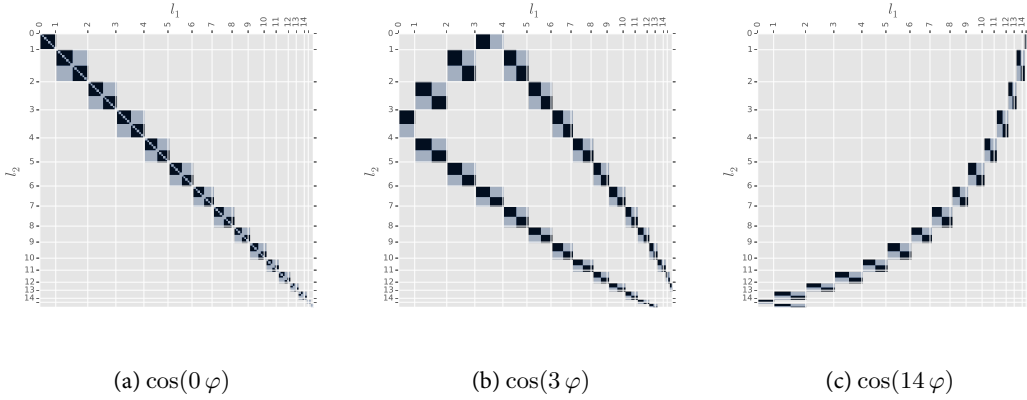


Figure 3: Nonzero entries for a few slices of the collision tensor for $K = 16$. The plots are labeled by the angular part of the test function, since the location of the nonzero entries depend on it solely. The basis functions are sorted by $(l, \cos / \sin, k)$, where l is the angular frequency and k denotes the polynomial degree in radial direction.

3.4 Exploiting the translational invariance of Q

We have used the rotational invariance of the collision operator for efficient computation and storage of its discrete analogue. According to (13), Q is also invariant to translation. A Maxwellian at temperature $T = 1$ with momentum $\mathbf{u} = 0$ is represented in the polar basis by a single non-zero coefficient. In order to represent the same Maxwellian with momentum $\mathbf{u} \neq 0$ with same accuracy, the required polynomial degree K grows with $\|\mathbf{u}\|$, cf. Sec. 3.4.1. If one wants to apply the collision operator to a given function, it would be beneficial to perform first a change of variables such that it has zero momentum, apply the collision operator and then shift it back to the original position. This has the advantage that a given function with zero momentum will have faster decaying coefficients compared to its nonzero momentum counterpart and thus one might truncate at a lower K without loss of accuracy. The straightforward way to translate a given function in its polar representation to zero momentum is to compute the expansion of $f(\mathbf{v} + \mathbf{u})$ in the Polar-Laguerre basis, where \mathbf{u} denotes the momentum. This entails the evaluation of f , which costs $\mathcal{O}(K^2)$, at $\mathcal{O}(K^2)$ quadrature points, resulting in a total cost of $\mathcal{O}(K^4)$. In the following, we will show that this can be done with complexity $\mathcal{O}(K^3)$ if we temporarily switch to the Hermite basis. The Hermite expansion with coefficients $c_{s,k-s}$ of a function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ reads

$$f(x, y) = \sum_{k=0}^{K-1} \sum_{s=0}^k c_{s,k-s} h_s(x) h_{k-s}(y) e^{-\frac{x^2+y^2}{2}}, \quad (42)$$

where $h_s(x)$, $h_{k-s}(y)$ are Hermite polynomials orthogonal with respect to the weights e^{-x^2} and e^{-y^2} . As a consequence of Lemma (2.3), any function in the Polar-Laguerre basis of degree K has an exact representation through Hermite polynomials of total degree K . Let us formally define the coefficient transformations matrices $\mathbf{T}_{P \rightarrow H}$, $\mathbf{T}_{H \rightarrow P}$ used to transform Polar-Laguerre to Hermite coefficients and vice versa:

$$\begin{aligned} \mathbf{c}^H &= \mathbf{T}_{P \rightarrow H} \mathbf{c}^P \\ \mathbf{c}^P &= \mathbf{T}_{H \rightarrow P} \mathbf{c}^H, \end{aligned}$$

where $\mathbf{T}_{P \rightarrow H}, \mathbf{T}_{H \rightarrow P} \in \mathbb{R}^{N \times N}$. Because of their block-diagonal structure with dense blocks of size $k+1, k=0, \dots, K-1$, the cost to transform the coefficients from the Polar-Laguerre to the Hermite basis is $\mathcal{O}(K^3)$. The derivation of the Polar-Laguerre to Hermite transformation matrices can be found in [11, Sec. 3.2].

Let c_k denote the coefficients of a 1-dimensional Hermite expansion g with maximal polynomial degree K and momentum \bar{x} . We are looking for the Hermite expansion of $\bar{g}(x) = g(x + \bar{x})$.

$$\bar{g}(x) = g(x + \bar{x}) = \sum_{k=0}^{K-1} c_k h_k(x + \bar{x}) e^{-\frac{(x+\bar{x})^2}{2}} \approx \sum_{k=0}^{K-1} \bar{c}_k h_k(x) e^{-\frac{x^2}{2}} \quad (43)$$

Note that $\bar{g}(x)$ has zero momentum. The coefficients \bar{c}_i are computed by forming L_2 -inner products.

$$\begin{aligned} \bar{c}_i &= \frac{1}{s_i} \int_{\mathbb{R}} \sum_{k=0}^{K-1} c_k h_k(x + \bar{x}) e^{-\frac{(x+\bar{x})^2}{2}} h_i(x) e^{-\frac{x^2}{2}} dx \\ &= \sum_{k=0}^{K-1} c_k \frac{1}{s_i} \int_{\mathbb{R}} h_k(x + \bar{x}) h_i(x) e^{-\frac{(x+\bar{x})^2}{2}} e^{-\frac{x^2}{2}} dx =: \sum_{k=0}^{K-1} (\mathbf{S}^{\bar{x}})_{i,k} c_k, \end{aligned} \quad (44)$$

where $s_i = \int_{\mathbb{R}} h_i(x) h_i(x) e^{-x^2} dx$. The above can be written as a matrix-vector-product $\bar{\mathbf{c}} = \mathbf{S}^{\bar{x}} \mathbf{c}$, where $\mathbf{S}^{\bar{x}} \in \mathbb{R}^{K,K}$. To further simplify the expression for the matrix entries $(\mathbf{S}^{\bar{x}})_{i,j}$, we substitute $x = x - \frac{\bar{x}}{2}$

$$(\mathbf{S}^{\bar{x}})_{i,j} = \frac{1}{s_i} \int_{\mathbb{R}} h_j(x + \frac{\bar{x}}{2}) h_i(x - \frac{\bar{x}}{2}) e^{-x^2} e^{-\frac{\bar{x}^2}{4}} dx \quad (45)$$

and use the identity

$$h_n(x + \bar{x}) = \sum_{k=0}^n \binom{n}{k} (2\bar{x})^{n-k} h_k(x), \quad (46)$$

to expand $h_k(x + \frac{\bar{x}}{2}), h_i(x - \frac{\bar{x}}{2})$ and find

$$\begin{aligned} (\mathbf{S}^{\bar{x}})_{i,j} &= \frac{1}{s_i} \sum_{s=0}^i \sum_{t=0}^j \binom{i}{s} \binom{j}{t} (-\bar{x})^{i-s} (\bar{x})^{j-t} e^{-\frac{\bar{x}^2}{4}} \delta_{t,s} \sqrt{\pi} 2^t t! \\ &= \frac{\sqrt{\pi}}{s_i} e^{-\frac{\bar{x}^2}{4}} \sum_{t=0}^{\min(i,j)} \binom{i}{t} \binom{j}{t} (-\bar{x})^{i-t} (\bar{x})^{j-t} 2^t t!, \end{aligned} \quad (47)$$

where we have used the orthogonality of the Hermite polynomials.

To carry out the shifting in $2D$, we rearrange the Hermite coefficient vectors $\mathbf{c}, \bar{\mathbf{c}}$ into lower triangular matrices $\mathbf{C}, \bar{\mathbf{C}} \in \mathbb{R}^{K,K}$:

$$f^H(x, y) = \sum_{i=1}^K \sum_{j=1}^K [\mathbf{C}]_{i,j} h_i(x) e^{-\frac{x^2}{2}} h_j(y) e^{-\frac{y^2}{2}}, \quad (48)$$

The matrix $\mathbf{S}^{\bar{x}}$ applied along the columns of \mathbf{C} performs the shift in x -direction, subsequent row-wise application of $\mathbf{S}^{\bar{y}}$ shifts in y -direction:

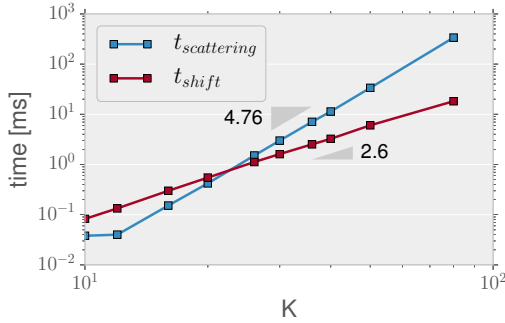
$$\begin{aligned}\bar{\mathbf{C}}^T &= \mathbf{S}^{\bar{y}}(\mathbf{S}^{\bar{x}}\mathbf{C})^T \\ \Leftrightarrow \bar{\mathbf{C}} &= \mathbf{S}^{\bar{x}}\mathbf{C}\mathbf{S}^{\bar{y},T}.\end{aligned}\tag{49}$$

We use orthonormal Hermite polynomials in the implementation to avoid numerical overflow. The procedure described above is summarized in Algorithm 1, whose total cost without evaluating the collision operator is $\mathcal{O}(K^3)$.

Algorithm 1

Collision operator in re-centered basis via Hermite representation. (Superscripts P, H denote coefficients in Polar-Laguerre / Hermite basis.)

- 1: **procedure** APPLY Q IN RE-CENTERED BASIS(\mathbf{c}^P)
 - 2: $\mathbf{c}^H \leftarrow \mathbf{T}_{P \rightarrow H}\mathbf{c}^P$ ▷ Transform to Hermite basis
 - 3: $\bar{\mathbf{c}}^H \leftarrow \mathbf{S}^{\bar{x}}\mathbf{c}^H$ ▷ Transform to zero momentum
 - 4: $\bar{\mathbf{c}}^P \leftarrow \mathbf{T}_{H \rightarrow P}\bar{\mathbf{c}}^H$ ▷ Go back to Polar-Laguerre basis
 - 5: $\bar{\mathbf{c}}^P \leftarrow$ update with Q in truncated basis
 - 6: $\bar{\mathbf{c}}^H \leftarrow \mathbf{T}_{P \rightarrow H}\bar{\mathbf{c}}^P$ ▷ Transform to Hermite basis
 - 7: $\mathbf{c}^H \leftarrow \mathbf{S}^{-\bar{x}}\bar{\mathbf{c}}^H$ ▷ Shift back
 - 8: $\mathbf{c}^P \leftarrow \mathbf{T}_{H \rightarrow P}\mathbf{c}^H$ ▷ Transform to Polar-Laguerre basis
 - 9: **end procedure**
-



(a)

K	$t_{\text{shift}}[ms]$	$t_{\text{collision op.}}[ms]$
10	0.08	0.04
12	0.13	0.04
16	0.3	0.15
20	0.55	0.42
26	1.12	1.51
30	1.61	2.99
36	2.52	7.08
40	3.25	11.4
50	6.04	33.6
80	18.2	337

(b)

Figure 4: CPU-time: Intel Core i7 4790K (4GHz, single threaded), Linux 4.2.3, GCC 5.2.0, relevant compiler flags: `-O3 -mssse2 -mavx2`. t_{shift} is the time for the execution of Algorithm 1 except the application of the collision operator.

Fig. 4 displays timings for the shifting procedure and the application of the collision operator for varying polynomial degree K . For $K < 40$ the shifting does not pay off, because we have observed that it is slower than the application of the collision operator.

3.4.1 Example: Decay of coefficients

The following example is to demonstrate that the Polar-Laguerre coefficients decay fastest if the approximand is centered such that it has zero momentum.

$$f(\mathbf{v}) = \exp(-\mathbf{v}^T \mathbf{M} \mathbf{v}) + \exp\left(-\frac{\|\mathbf{v}-\mathbf{v}_c\|^2}{2}\right), \quad (50)$$

where

$$\mathbf{M} = \frac{1}{8} \begin{bmatrix} 7 & \sqrt{3} \\ \sqrt{3} & 5 \end{bmatrix}, \quad \mathbf{v}_c = \left[\frac{1}{5}, 0\right]. \quad (51)$$

The decay of the absolute values of the Polar-Laguerre coefficients $|\mathbf{c}|$ with respect to angular index $l := 2j + k \pmod{2}$ and radial index k is shown in Fig. 5.

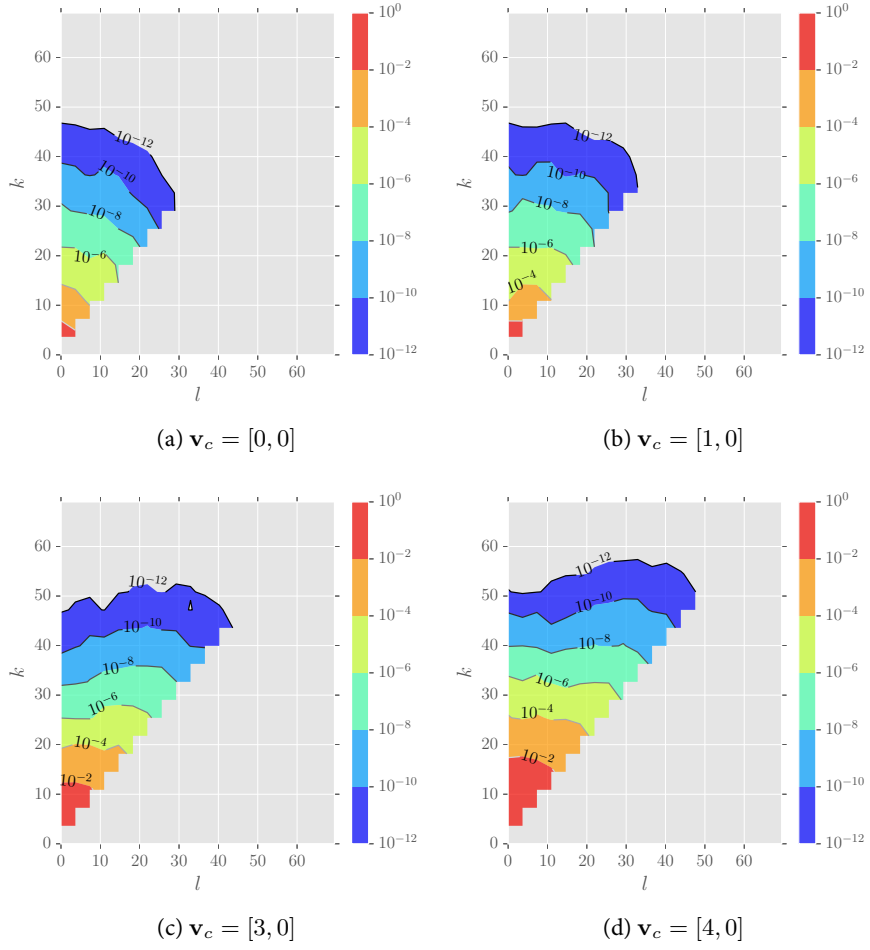


Figure 5: Decay of Polar-Laguerre coefficients for $f(\mathbf{v} - \mathbf{v}_c)$, defined in (50), with respect to angular index l and radial index k .

4 Discretization in Physical Space

In this section we present the spatial discretization in $D \subset \mathbb{R}^2$. It is well known that the advection part in (1) requires stabilization. We use a least squares formulation, which has the advantage that, after partial integration, the term $\langle \mathbf{v} \cdot \mathbf{n} \Phi, f \rangle_\Gamma$ appears in the variational formulation, which comes handy to include inflow-type boundary conditions. The advection part of (1) reads

$$\partial_t f + \mathbf{v} \cdot \nabla_{\mathbf{x}} f = 0. \quad (52)$$

We replace $\partial_t f$ in (52) by a backwards difference quotient and write down the least squares functional $J(f^{(n)}; f^{(n-1)})$ for the pure transport problem [18, Ch. 10.3.1]:

$$J(f^{(n)}; f^{(n-1)}) := \left\| \frac{1}{\Delta t} (f^{(n)} - f^{(n-1)}) + \mathbf{v} \cdot \nabla_{\mathbf{x}} f^{(n)} \right\|_{L^2(\Omega)}^2 \quad (53)$$

The bilinear form a and right hand side linear form b of the associated variational problem are given by

$$a(\Phi, f^{(n)}) = \frac{1}{\Delta t^2} \langle \Phi, f^{(n)} \rangle_\Omega + \frac{1}{\Delta t} \langle \mathbf{v} \cdot \mathbf{n} \Phi, f^{(n)} \rangle_\Gamma + \langle \mathbf{v} \cdot \nabla_{\mathbf{x}} \Phi, \mathbf{v} \cdot \nabla_{\mathbf{x}} f^{(n)} \rangle_\Omega, \quad (54)$$

where we have used partial integration in \mathbf{x} to obtain the boundary term, and

$$b(\Phi, f^{(n-1)}) := \frac{1}{\Delta t^2} \langle \Phi, f^{(n-1)} \rangle_\Omega + \frac{1}{\Delta t} \langle \mathbf{v} \cdot \nabla_{\mathbf{x}} \Phi, f^{(n-1)} \rangle_\Omega, \quad (55)$$

where \mathbf{n} is the unit outward normal vector on ∂D and $\Gamma := \partial D \times \mathbb{R}^2$. In the following we use $\langle \cdot, \cdot \rangle$ to denote the L_2 -inner product. V_D^L is the space of linear, piecewise continuous finite elements on quadrilateral triangulations of $D \subset \mathbb{R}^2$. The VDF on phase space $\Omega = D \times \mathbb{R}^2$ is approximated in the tensor product space $V^{L,N} = V_D^L \otimes V_V^N$. The test functions Φ are also taken from $V^{L,N}$. The superscript L will denote the number of degrees of freedom in physical space. The inclusion of boundary conditions is done in a weak sense, details will be discussed in the next section. When inflow boundary conditions are present, the corresponding parts of $\langle \mathbf{v} \cdot \mathbf{n} \Phi, f \rangle_\Gamma$ enter the right hand side.

For integration in time we separate (1) into advection and collision part and use a first order split time-stepping.

1. **Advection:** $\partial_t f + \mathbf{v} \cdot \nabla_{\mathbf{x}} f = 0$ (implicit Euler):

$$\begin{aligned} \frac{1}{\Delta t_k^2} \langle \Phi, f^{(n+1/2)} \rangle_\Omega + \frac{1}{\Delta t_k} \langle \mathbf{v} \cdot \mathbf{n} \Phi, f^{(n+1/2)} \rangle_\Gamma + \langle \mathbf{v} \cdot \nabla_{\mathbf{x}} \Phi, \mathbf{v} \cdot \nabla_{\mathbf{x}} f^{(n+1/2)} \rangle_\Omega \\ = \frac{1}{\Delta t_k^2} \langle \Phi, f^{(n)} \rangle_\Omega + \frac{1}{\Delta t_k} \langle \mathbf{v} \cdot \nabla_{\mathbf{x}} \Phi, f^{(n)} \rangle_\Omega \end{aligned} \quad (56)$$

2. **Collision operator** (explicit Euler):

$$f^{(n+1)} = f^{(n+1/2)} + \frac{\Delta t_k}{kn} Q(f^{(n+1/2)}, f^{(n+1/2)}) \quad (57)$$

5 Treatment of Boundary Conditions

We discuss inflow, specular reflective and diffusive reflective boundary conditions, which were defined in (3), (4) and (5). These are the simplest with physical significance. There exist other models, a detailed discussion can be found in [19, Ch. 1.11] and the references therein.

We split the second term of $a(\Phi, f)$, cf. (54), into in- and outflow part:

$$\frac{1}{\Delta t} \langle \mathbf{v} \cdot \mathbf{n} \Phi, f^{(n)} \rangle_{\Gamma} = \frac{1}{\Delta t} \left(\langle \mathbf{v} \cdot \mathbf{n} \Phi, f^{(n)} \rangle_{\Gamma^-} + \langle \mathbf{v} \cdot \mathbf{n} \Phi, f^{(n)} \rangle_{\Gamma^+} \right), \quad (58)$$

where $\Gamma = \partial D \times \mathbb{R}^2$. The function $f^{(n)}$ in $\langle \mathbf{v} \cdot \mathbf{n}, f^{(n)} \rangle_{\Gamma^-}$ can be replaced by the conditions for specular reflective, diffusive reflective or inflow boundary conditions. In the following we will discuss the conservation of moments for specular reflective and diffusive reflective boundary conditions.

Theorem 5.1. *Specular reflective boundary conditions (4) conserve mass and energy in the continuous formulation.*

Proof. Multiply (1) by $(1, \mathbf{v}, \|\mathbf{v}\|^2)^T$ and integrate over Ω to obtain:

$$\partial_t \int_D \int_{\mathbb{R}^2} f(t, \mathbf{x}, \mathbf{v}) \begin{pmatrix} 1 \\ \mathbf{v} \\ \|\mathbf{v}\|^2 \end{pmatrix} d\mathbf{v} d\mathbf{x} \equiv \partial_t \begin{pmatrix} \rho \\ \rho \mathbf{u} \\ \rho E \end{pmatrix} = - \int_D \int_{\mathbb{R}^2} \begin{pmatrix} 1 \\ \mathbf{v} \\ \|\mathbf{v}\|^2 \end{pmatrix} \mathbf{v} \cdot \nabla_{\mathbf{x}} f(t, \mathbf{x}, \mathbf{v}) d\mathbf{v} d\mathbf{x} \quad (59)$$

$$\Leftrightarrow \partial_t \begin{pmatrix} \rho \\ \rho \mathbf{u} \\ \rho E \end{pmatrix} = - \int_{\Gamma^+} \mathbf{n}(\mathbf{x}) \cdot \mathbf{v} \begin{pmatrix} 1 \\ \mathbf{v} \\ \|\mathbf{v}\|^2 \end{pmatrix} f(t, \mathbf{x}, \mathbf{v}) d\mathbf{v} d\mathbf{x} - \int_{\Gamma^-} \mathbf{n}(\mathbf{x}) \cdot \mathbf{v} \begin{pmatrix} 1 \\ \mathbf{v} \\ \|\mathbf{v}\|^2 \end{pmatrix} f(t, \mathbf{x}, \mathbf{v}) d\mathbf{v} d\mathbf{x}. \quad (60)$$

Next we insert the specular reflective boundary condition on the inflow boundary Γ^- which is $f(t, \mathbf{x}, \mathbf{v}) = f(t, \mathbf{x}, \mathbf{v} - 2\mathbf{n} \mathbf{n} \cdot \mathbf{v})$, $(\mathbf{x}, \mathbf{v}) \in \Gamma^-$ and obtain

$$\int_{\Gamma^-} \mathbf{n} \cdot \mathbf{v} \begin{pmatrix} 1 \\ \mathbf{v} \\ \|\mathbf{v}\|^2 \end{pmatrix} f(t, \mathbf{x}, \mathbf{v} - 2\mathbf{n} \mathbf{n} \cdot \mathbf{v}) d\mathbf{v} d\mathbf{x} d\mathbf{v} = - \int_{\Gamma^+} \mathbf{n} \cdot \mathbf{v} \begin{pmatrix} 1 \\ \mathbf{v} - 2\mathbf{n} \mathbf{n} \cdot \mathbf{v} \\ \|\mathbf{v}\|^2 \end{pmatrix} f(t, \mathbf{x}, \mathbf{v}) d\mathbf{x} d\mathbf{v}, \quad (61)$$

the right hand side follows after making the substitution $\mathbf{v} \rightarrow \mathbf{v} - 2\mathbf{n} \mathbf{n} \cdot \mathbf{v}$. Inserting back into (60) reveals that $\partial_t \rho \equiv 0$ and $\partial_t(\rho E) \equiv 0$, which finishes the proof. \square

Theorem 5.2. *Specular reflective boundary conditions conserve mass and energy in the discrete formulation.*

Proof. The discretized solution with coefficients $c_{k,j,i_x}^{(n)}$ at time t_n can be written as:

$$f^{(n)} = \sum_{i_x, j, k} c_{k,j,i_x}^{(n)} \Psi_{k,j}(\mathbf{v}) \phi_{i_x}(\mathbf{x}),$$

where ϕ_{i_x} are the basis functions spanning V_D^L . Insert a test function Φ which is constant in \mathbf{x} into the variational formulation $a(\Phi, f^{(n)}) = b(\Phi, f^{(n-1)})$:

$$\frac{1}{\Delta t} \left(\langle \Phi, f^{(n)} \rangle_{\Omega} - \langle \Phi, f^{(n-1)} \rangle_{\Omega} \right) + \langle \mathbf{v} \cdot \mathbf{n} \Phi, f^{(n)} \rangle_{\Gamma} + \Delta t \langle \mathbf{v} \cdot \nabla_{\mathbf{x}} \Phi, \mathbf{v} \cdot \nabla_{\mathbf{x}} f^{(n)} \rangle_{\Omega} = 0. \quad (62)$$

Note that

$$\left\langle \mathbf{v} \cdot \mathbf{n} \Phi, f^{(n)} \right\rangle_{\Omega} = \int_{\Gamma} \int_{\mathbb{R}^2} \mathbf{v} \cdot \mathbf{n} \Phi(\mathbf{v}) f^{(n)}(\mathbf{x}, \mathbf{v}) \, d\mathbf{x} \, d\mathbf{v}. \quad (63)$$

All terms in (62) involving $\nabla_{\mathbf{x}} \Phi$ evaluate to zero and we obtain

$$\left(\left\langle \Phi, f^{(n)} \right\rangle_{\Omega} - \left\langle \Phi, f^{(n-1)} \right\rangle_{\Omega} \right) = -\Delta t \left\langle \mathbf{v} \cdot \mathbf{n} \Phi, f^{(n)} \right\rangle_{\Gamma}, \quad (64)$$

since, on Γ , $f^{(n)}$ satisfies (4), we see that the right hand side of (64) vanishes whenever $\Phi \in V^{L,N}$ is chosen to be rotationally symmetric in the velocity coordinate, thus let $\Phi(\mathbf{x}, \mathbf{v}) = \Psi_{k',0}^{\cos}(\mathbf{v})$, k' even, recall Def. 2.1 for the definition of $\Psi_{k,j}(\mathbf{v})$:

$$\left\langle \Psi_{k',0}^{\cos}(\mathbf{v}), \sum_{k,i_x} \underbrace{\left(c_{k,0,i_x}^{(n)} - c_{k,0,i_x}^{(n-1)} \right)}_{:=\Delta c_{k,0,i_x}} \Psi_{k,j}(\mathbf{v}) \phi_{i_x}(\mathbf{x}) \right\rangle_{\Omega} = 0 \quad (65)$$

$$\Rightarrow \sum_{i_x} \Delta c_{k,0,i_x} \int_D \phi_{i_x} = 0, \quad \forall k \text{ even} \quad (66)$$

where we have used the L_2 -orthogonality of the spectral basis in the last line. The change in mass between times t_{n-1} , t_n is given by:

$$\int_D \rho^{(n)}(\mathbf{x}) - \rho^{(n-1)}(\mathbf{x}) \, d\mathbf{x} = \sum_{k,j} \sum_{i_x} \Delta c_{k,j,i_x} \underbrace{\int_{\mathbb{R}^2} \Psi_{j,k,i_x}(\mathbf{v}) \, d\mathbf{v}}_{=0 \text{ if } k \text{ odd or } j \neq 0} \int_D \phi_{i_x}(\mathbf{x}) \, d\mathbf{x} \quad (67)$$

$$= \pi \sum_{k \text{ even}} \sum_{i_x} \underbrace{\Delta c_{k,0,i_x} \int_D \phi_{i_x}}_{=0}. \quad (68)$$

In the last line we have used (66). Conservation of energy can be shown in the same way, except that one has an additional $\|\mathbf{v}\|^2$ in the integral over \mathbb{R}^2 in (67) which also evaluates to zero for k odd or $j \neq 0$. \square

Theorem 5.3. *In the presence of diffusive reflective boundary conditions mass is conserved in the continuous formulation.*

Proof. It must hold that

$$\begin{aligned} \int_{\mathbf{x} \in \partial D} \int_{\mathbf{v} \cdot \mathbf{n}(\mathbf{x}) > 0} \mathbf{n}(\mathbf{x}) \cdot \mathbf{v} f(t, \mathbf{x}, \mathbf{v}) \, d\mathbf{v} \, d\mathbf{x} &= - \int_{\mathbf{x} \in \partial D} \int_{\mathbf{v} \cdot \mathbf{n}(\mathbf{x}) < 0} \mathbf{n}(\mathbf{x}) \cdot \mathbf{v} f(t, \mathbf{x}, \mathbf{v}) \, d\mathbf{v} \, d\mathbf{x} \\ &= + \int_{\mathbf{x} \in \partial D} \rho_+(f) \underbrace{\int_{\mathbf{v} \cdot \mathbf{n}(\mathbf{x}) > 0} \mathbf{n}(\mathbf{x}) \cdot \mathbf{v} M_w(\|\mathbf{v}\|) \, d\mathbf{v}}_{\equiv 1} \, d\mathbf{x}, \quad (69) \end{aligned}$$

where we have made the changes of variables $\mathbf{v} \rightarrow \mathbf{v} - 2\mathbf{n} \mathbf{n} \cdot \mathbf{v}$ in the third line. By definition, the left hand side of (69) is $\int_{\mathbf{x} \in \partial D} \rho_+(f) \, d\mathbf{x}$, which finishes the proof. \square

Remark 5.4. The discrete formulation does not conserve mass for diffusive reflective boundary conditions, because in general, the velocity distribution function will have jumps across the line $\mathbf{v} \cdot \mathbf{n} \equiv 0$. Discontinuous functions cannot be represented exactly in the Polar-Laguerre basis and therefore mass is not conserved.

6 Numerical Experiments

We have implemented all the techniques discussed in C++. The finite element part is taken from the deal.II v8.3 [20] library. The collision operator is independent of \mathbf{x} and it is thus natural to parallelize via domain decomposition in the physical domain. The system matrix arising from the advection problem is assembled once and reused in every time-step. We use a block-diagonal, incomplete LU-factorization as preconditioner. Often it is observed that the ILU-preconditioned¹ GMRES solver converges in less than 5 iterations. We use the distributed vector, sparse matrix, iterative solvers and preconditioners offered by Trilinos v12.2.1 [21].

The numerical experiments in this section are carried out for Maxwellian molecules. The entries of the collision tensor were computed with 81, 131 quadrature points in radial direction and angular direction. For the inner integral (28) 131 quadrature points were used. Thus it can be assumed that the quadrature error is negligible.

6.1 Homogeneous case

In order to validate the implementation of the collision operator and to study the approximation properties of the Polar-Laguerre basis, we consider the homogeneous Boltzmann equation

$$\partial_t f = Q(f, f), \quad (22)$$

for which a non-stationary, analytical solution is available. This is the so-called BKW solution:

$$f(t, \mathbf{v}) = e^{-\frac{\|\mathbf{v}\|^2}{2s}} \frac{\|\mathbf{v}\|^2 - (2 + \|\mathbf{v}\|^2)s + 4s^2}{4\pi s^3}, \quad t > 0 \quad (70)$$

where $s = 1 - \exp(-\frac{1}{8}(t + 8 \log 2))$. It is valid for Maxwellian molecules. Taking the limit $t \rightarrow \infty$ of (70) shows that the equilibrium solution agrees with a single nonzero coefficient in the Polar-Laguerre basis:

$$\lim_{t \rightarrow \infty} f(t, \mathbf{v}) = \frac{1}{2\pi} e^{-\|\mathbf{v}\|^2/2}$$

The BKW solution (70) has temperature $T=1$, and thus we call it to be a temperature-normalized solution in the Polar-Laguerre basis.

Theorem 6.1. *Let $f(t, \mathbf{v})$ be a solution to (22) with a collision kernel of the form (11). Let $\alpha, \gamma > 0$ be given, and define $\eta = \alpha/\gamma^{\lambda+2}$. Then*

$$h(t, \mathbf{v}) = \alpha f(\eta t, \gamma \mathbf{v})$$

is also a solution to (22).

The proof can be found in [8]. Making use of Theorem 6.1 and rescaling $f(t, \mathbf{v})$ accordingly, we can construct analytical solutions with different temperatures. Numerical results for the BKW solution are reported in Fig. 6. In order to demonstrate the approximation properties of the Polar-Laguerre basis, the simulations were carried out for temperature-normalized initial distributions with $T=0.5$, 1 centered at $\mathbf{v}_c=[0, 0]$ and $\mathbf{v}_c=[1, 1]$. Also, we compare the two different methods to conserve mass, momentum and

¹Block-diagonal ILU preconditioner with zero fill-in from Trilinos IFPACK.

energy described in Sec. 3.1. We used RK4 with step size $\Delta t=10^{-3}$ for $T=1$ and $\Delta t=2 \times 10^{-3}$ for the $T=0.5$ -normalized initial distribution. The equilibrium state was reached after $15k$ time-steps. Relative L_2 -errors are reported in Fig. 6. As expected we observe the fastest decay in the L_2 -errors wrt. time for the normalized temperature, i.e. $T=1$, BKW initial distribution with zero momentum, cf. Fig. 6b. For $t > 10$ and for sufficiently high polynomial degree K , the errors are of the size of the machine precision. The numerical results reveal that the Galerkin discretization of the collision operator in conjunction with the Lagrange multipliers yields considerably smaller errors than the Petrov-Galerkin approach.

Remark 6.2. The authors in [8] consider the homogeneous Boltzmann equation. Initial conditions are always rescaled such that they are temperature-normalized, which allows to obtain best possible approximation properties. When working with continuous finite elements, a per element rescaled basis in the velocity coordinate would create dense sub-blocks of size $N \times N$ in the system matrix associated with the advection problem, and furthermore make it necessary to reassemble the system matrix in every time-step. The costs in regards to memory requirements and computational effort would be prohibitively high.

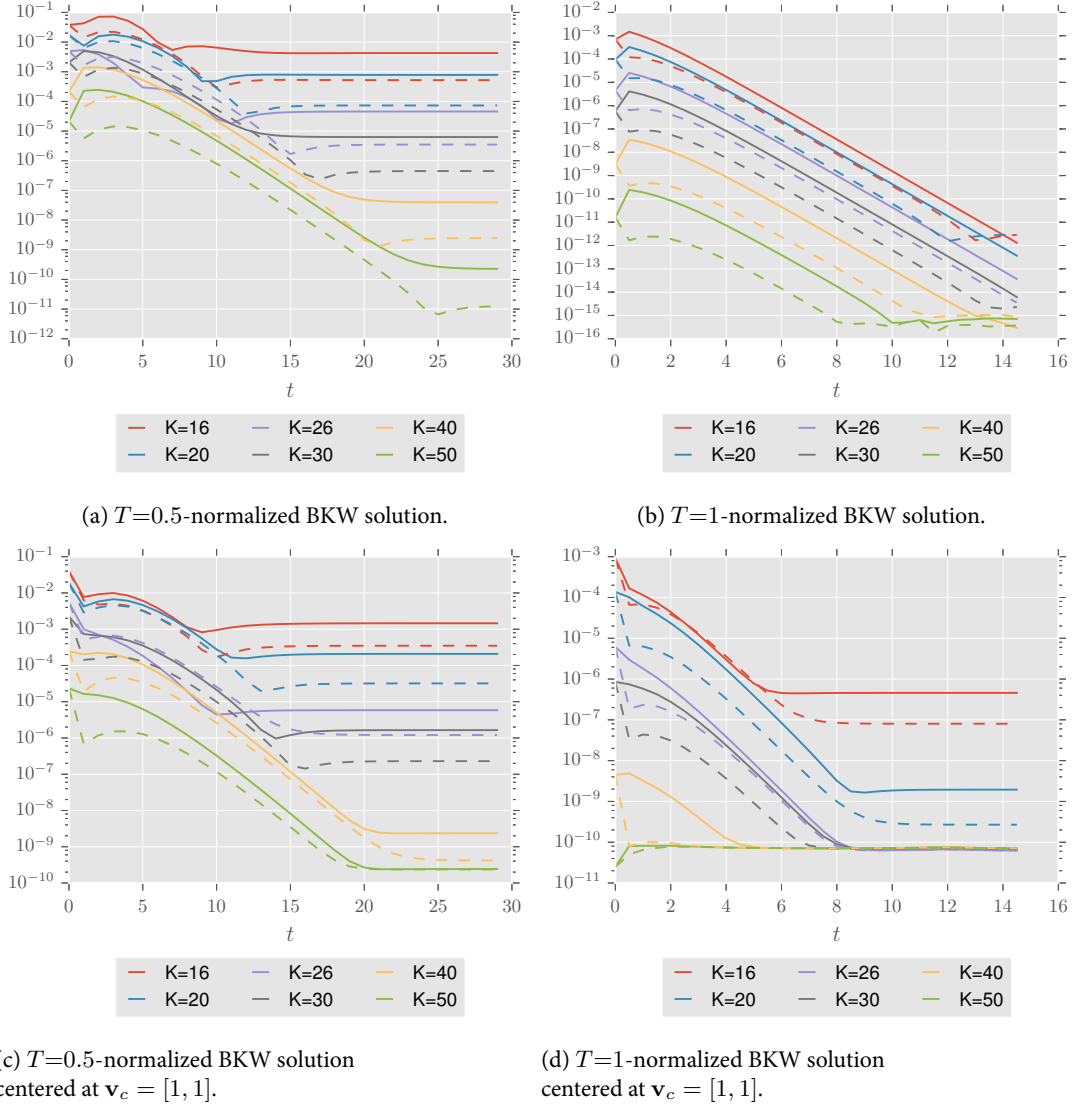


Figure 6: Relative L_2 -errors for the BKW solution versus time t . **Solid lines:** Petrov-Galerkin scheme, **dashed lines:** Galerkin scheme with Lagrange multipliers. The errors were measured against an expansion of the exact solution in the spectral basis with degree $K = 60$, the coefficients were modified by the method described in Sec. 3.1 in order to yield the same mass, momentum and energy as the exact solution does.

6.2 Mach 3 flow in a wind tunnel

We show numerical results for the well known Mach 3 wind tunnel experiment, which was first introduced in [22]. The computational domain describes a wind tunnel with a forward facing step at position $x = 0.6$ with height 0.2. The gas is initially at equilibrium with temperature $T_0=1$, $\mathbf{v}=[3, 0]$, $\rho = 1.4$. At $x=0$ inflow boundary conditions with $T=1$, $\mathbf{v}=[3, 0]$, $\rho=1.4$ are imposed and outflow (zero inflow) boundary conditions at $x=3$, the other walls are specularly reflective. The Knudsen number was $kn = 2.5 \times 10^{-3}$ for a Maxwellian gas. In Fig. 8 the pressure is shown at different times $t \in [0, 1]$. We have used a time-step of length $\Delta t=2.5 \times 10^{-5}$. The solution shown in Fig. 8 qualitatively agrees with simulation results obtained from the compressible Euler equations, which can be found, for example in [23]. We have observed that on coarse meshes, the distribution function can become negative at the re-entrant corner. For small Knudsen numbers, for example $kn=2.5 \times 10^{-3}$, this may cause the solution to diverge when the collision operator is applied. A possible remedy is to project onto positive distribution values in \mathbf{v} , see discussion below. The projection step was not required for the results reported here.

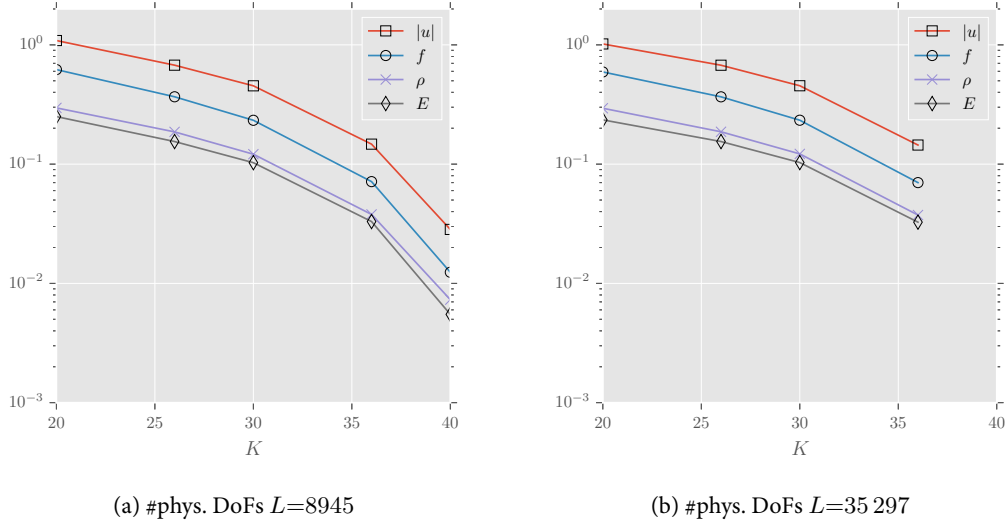


Figure 7: Rel. L_2 -errors vs. polynomial degree K for two physical grids with different numbers of vertices. The solution on the finest grid with highest polynomial degree $K=40$ was used as reference. Errors are shown for the velocity distribution function f and the macroscopic observables: mass ρ , momentum \mathbf{u} and energy E . The errors are dominated by the polynomial degree K .

Ensuring positivity We have observed that in the vicinity of singularities, for example near re-entrant corners, the distribution function might locally become negative. In combination with a low Knudsen number this can cause numerical blow-up of the solution by the collision operator. A possible remedy is to evaluate the distribution function after each time-step at the quadrature nodes, set negative values to zero and project back onto the Polar-Laguerre basis. A naive implementation requires the evaluation of $f(\mathbf{v})$ at $\mathcal{O}(K^2)$ quadrature nodes, where the evaluation requires $\mathcal{O}(K^2)$ operations per node and thus has a total cost of $\mathcal{O}(K^4)$. The machinery developed in [10] provides an elegant solution by transforming first to the Hermite and then to the nodal basis. As already noted in Sec. 3.4, the transformation between

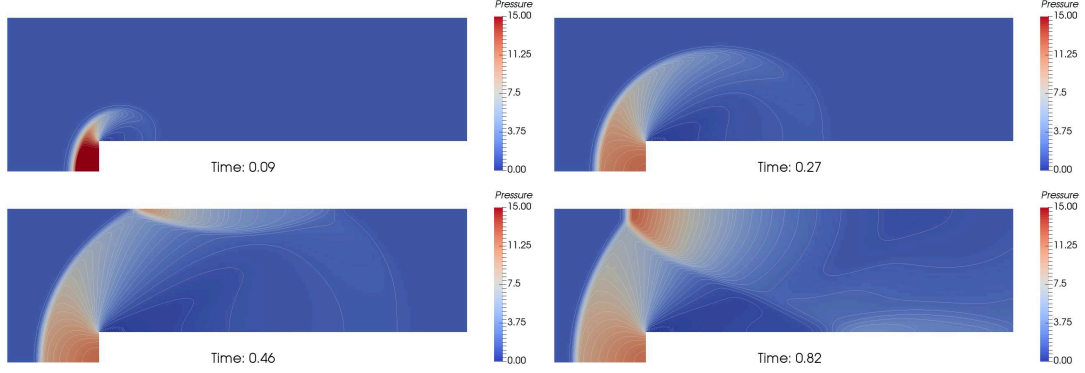


Figure 8: Mach 3 wind tunnel: polynomial degree $K = 40$, $35k$ vertices, Maxwellian molecules, $28.9M$ total DoFs. Coloring: pressure, contour lines: density. Computations were carried out on the Euler cluster of ETH Zurich (Xeon E5-2697 v2) using 360 cores.

the Polar-Laguerre and the Hermite basis can be done with effort $\mathcal{O}(K^3)$. The transformation between the Hermite and nodal basis again costs $\mathcal{O}(K^3)$, this time because it can be performed separately along each coordinate axis and therefore the transformation matrices are of size $K \times K$ only. The entries of the Hermite to nodal transformation matrix $\mathbf{T}_{\text{H} \rightarrow \text{N}} \in \mathbb{R}^{K,K}$ are given by:

$$(\mathbf{T}_{\text{H} \rightarrow \text{N}})_{i,j} = \int_{\mathbb{R}} h_j(x) e^{-\frac{x^2}{2}} \ell_i(x) e^{-\frac{x^2}{2}} dx = \sum_{k=0}^K h_j(x_k) \ell_i(x_k) w_k = \sum_{k=0}^K h_j(x_k) \frac{\delta_{i,k}}{\sqrt{w_k}} w_k = h_j(x_i) \sqrt{w_i}, \quad (71)$$

where $x_i, w_i, i = 0, \dots, K$ are the Gauss-Hermite quadrature nodes and weights. We have that $(\mathbf{T}_{\text{N} \rightarrow \text{H}})^{-1} := \mathbf{T}_{\text{H} \rightarrow \text{N}}^T$, since $\mathbf{T}_{\text{H} \rightarrow \text{N}}$ is an orthonormal matrix:

$$\begin{aligned} (\mathbf{T}_{\text{H} \rightarrow \text{N}})^T \mathbf{T}_{\text{H} \rightarrow \text{N}} &= \sum_{k=0}^K (h_i(x_k) \sqrt{w_k}) (h_j(x_k) \sqrt{w_k}) = \sum_{k=0}^K h_i(x_k) h_j(x_k) w_k \\ &= \int_{\mathbb{R}^2} h_i(x) h_j(x) e^{-x^2} dx = \delta_{i,j} \quad (72) \end{aligned}$$

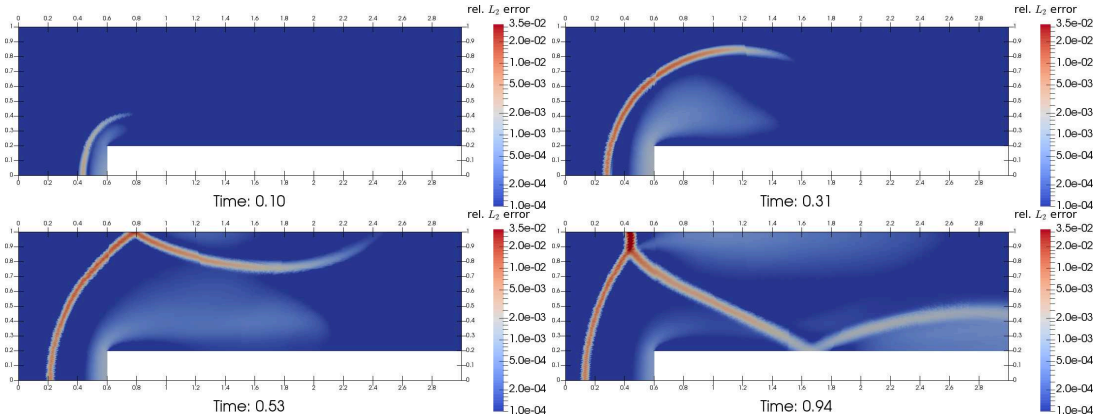


Figure 9: Mach 3 wind tunnel. Relative L_2 -errors for $K=36$, $L=35k$ on the physical grid (reference solution with $K = 40$).

Algorithm 2 Project to positive velocity distribution values

- 1: **procedure** APPLY THE COLLISION OP. IN RE-CENTERED BASIS(\mathbf{c}^P)
 - 2: $\mathbf{c}^H \leftarrow \mathbf{T}_{P \rightarrow H} \mathbf{c}^P$ ▷ Transform to Hermite basis
 - 3: $\mathbf{c}^N \leftarrow \mathbf{T}_{H \rightarrow N}$ ▷ Transform to Nodal basis
 - 4: **for all** $(\mathbf{c}^N)_i < 0$ **do**
 - 5: $(\mathbf{c}^N)_i \leftarrow 0$ ▷ Set negative coefficients to zero
 - 6: **end for**
 - 7: $\mathbf{c}^H \leftarrow \mathbf{T}_{N \rightarrow H} \mathbf{c}^N$ ▷ Transform to Hermite basis
 - 8: $\mathbf{c}^P \leftarrow \mathbf{T}_{H \rightarrow P} \mathbf{c}^H$ ▷ Transform to Polar-Laguerre basis
 - 9: **end procedure**
-

6.3 Nozzle flow

We consider the flow of a rarefied gas with $kn = 0.1$ in a nozzle, see Fig. 10. Inflow boundary conditions are placed at the left boundary with $T = 1$, $\mathbf{v}_0 = [2.5, 0]$, $\rho_0 = 1.4$, and outflow b.c. at $x = 4$, the other walls are specularly reflecting. The initial distribution was

$$f(t = 0, \mathbf{x}, \mathbf{v}) = \frac{\rho_0}{2\pi T} \exp\left(-\frac{\|\mathbf{v} - \mathbf{v}_0\|^2}{2}\right).$$

Convergence plots for the L_2 -errors are reported in Fig. 11, the reference solution was computed on a mesh with 18 500 vertices and for polynomial degree $K=40$. For the lowest resolution in space, i.e. $L=1200$ and for $K > 26$, we find that the error is dominated by the mesh size, whereas for $L=4700$ the errors mainly depend on K . Compared to the Mach 3 wind tunnel experiment, we obtain smaller errors and faster convergence wrt. K , which is attributed to the absence of shocks.

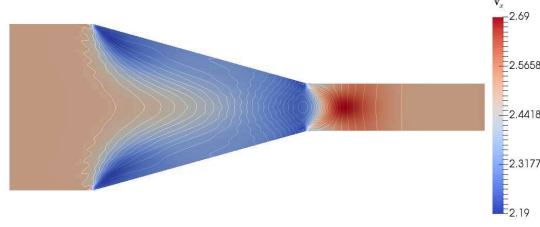


Figure 10: Nozzle flow: $L=18\,529$, $K=36$, $N=666$, velocity in x -direction. Pressure as contour lines.

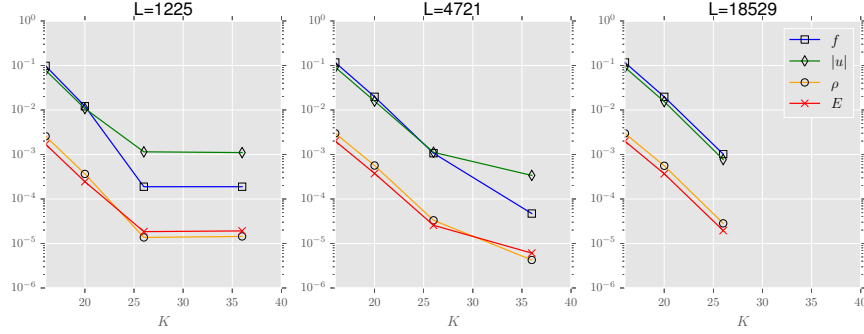


Figure 11: Nozzle flow: relative L_2 -errors at time $t = 3.75$. Reference solution with $K = 40$, $L = 18\,529$, $\Delta t = 2.5 \times 10^{-4}$. The collision operator was discretized with the Petrov-Galerkin scheme.

Figure 12

6.4 Shock tube

A gas is placed in a tube of unit length. Initially the gas is at equilibrium in the left and right half with densities ρ_l, ρ_r and temperatures T_l, T_r :

$$f_l(t=0, \mathbf{x}, \mathbf{v}) = \frac{\rho_l}{2\pi T_l} \exp\left(-\frac{\|\mathbf{v}\|^2}{2T_l}\right), \quad x < 0.5 \quad (73)$$

$$f_r(t=0, \mathbf{x}, \mathbf{v}) = \frac{\rho_r}{2\pi T_r} \exp\left(-\frac{\|\mathbf{v}\|^2}{2T_r}\right), \quad x \geq 0.5, \quad (74)$$

where $\rho_l=1, \rho_r=1$ and $T_l=1.25, T_r=1$. Specular reflective boundary conditions are imposed on the top and bottom wall, at $x=0, x=1$ we use inflow boundary conditions with densities ρ_l, ρ_r and temperatures T_l, T_r . The calculations were carried out on a structured grid with element size $h_x=1.48 \times 10^{-3}$ in x -direction for different $kn=0.01, 0.1, 1$, and with polynomial degrees $K=16, 20, 26, 30, 36, 40$. The calculation with $K=40$ is used as reference to compute L_2 -errors in the VDF $f(\mathbf{v}, \mathbf{x})$ and for the macroscopic quantities $\rho, |\mathbf{u}|$ and E . L_2 -errors are shown in Fig. 13, the errors for $kn=0.01$ are an order of magnitude smaller compared to the calculations with $kn=0.1$. This is because for $kn=0.01$, the smoothing by the collision operator is stronger and therefore better approximation in the velocity do-

main is obtained. In Fig. 14, the density and momentum \mathbf{u}_x are compared for $K=30, 40$ at different times along the line $x(s)=s$, $s \in [0, 1]$.

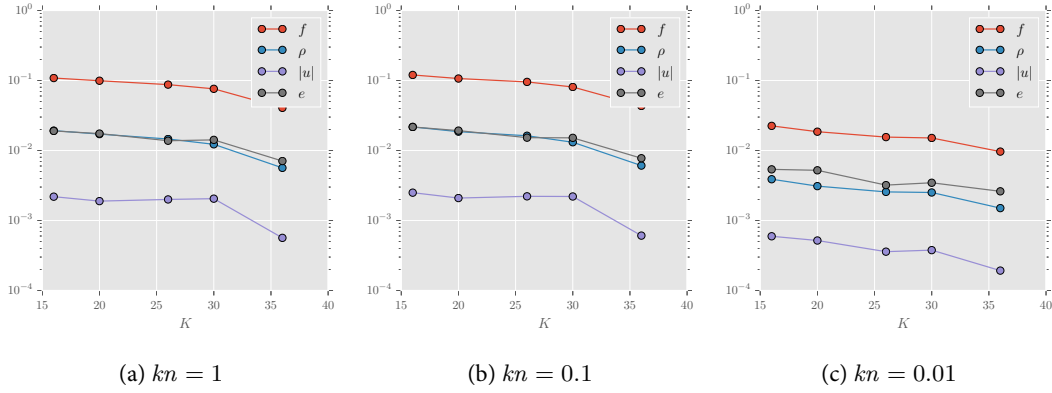


Figure 13: Relative L_2 -errors for the shock tube with varying polynomial degree K . Reference computation with $K=40$ at time $t=0.1$.

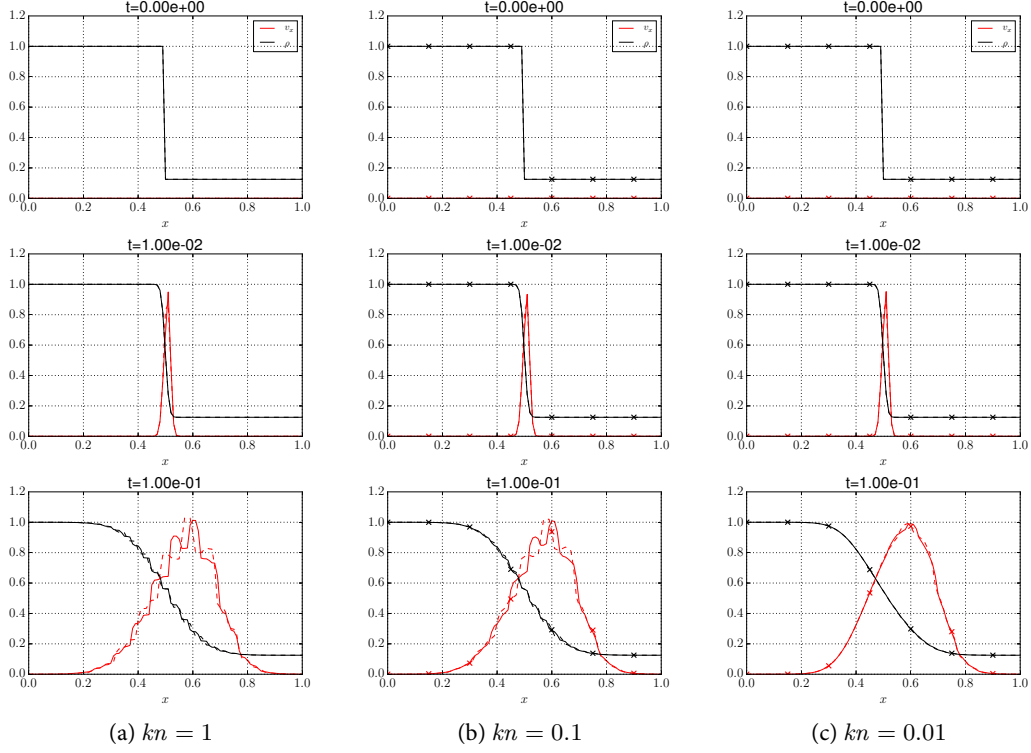


Figure 14: Shock tube: Macroscopic density and momentum in x -direction plotted along the line $x = [0, 1]$. **Solid line:** Polynomial degree $K = 40$, **Dashed line:** $K = 30$

6.5 Sudden change in wall temperature

We consider a gas initially at rest with temperature $T=1$ confined between to parallel plates at $y=0, 1$ and with periodic b.c. in x -direction. For $t>0$ diffusive reflective boundary conditions with temperature T_w are imposed on the walls. Computations were performed for two different temperatures $T_w=1.3, 1.7$. As it has been discussed in Sec. 5, mass is not conserved exactly for diffusive boundary conditions. Table (1) shows that if the polynomial degree K is chosen sufficiently large, mass is preserved up to $\approx 0.01\%$. The time evolution for temperature and mass in y -direction is shown in Fig. (15) and (16). The results agree qualitatively, but the fluctuations are too large to compute convergence rates. The inaccuracy originates from the temperature shock present at time $t=0$ at the walls. In order to satisfy the boundary condition, the velocity distribution function is required to be discontinuous perpendicular to the normal vector, what is not approximated well by the Polar-Laguerre basis.

T_w	K20	K26	K30	K36	K40	K50
1.3	0.15	$-6.59 \cdot 10^{-2}$	$-6.08 \cdot 10^{-2}$	$6.18 \cdot 10^{-2}$	$5.75 \cdot 10^{-2}$	$-4.50 \cdot 10^{-2}$
1.7	0.15	-0.11	-0.1	0.11	$9.85 \cdot 10^{-2}$	$-7.58 \cdot 10^{-2}$

Table 1: Deviation in mass [in percent] for different polynomial degrees K at time $t = 0.25$, $\Delta t = 10^{-4}$, mesh width: $h = 512^{-1}$.

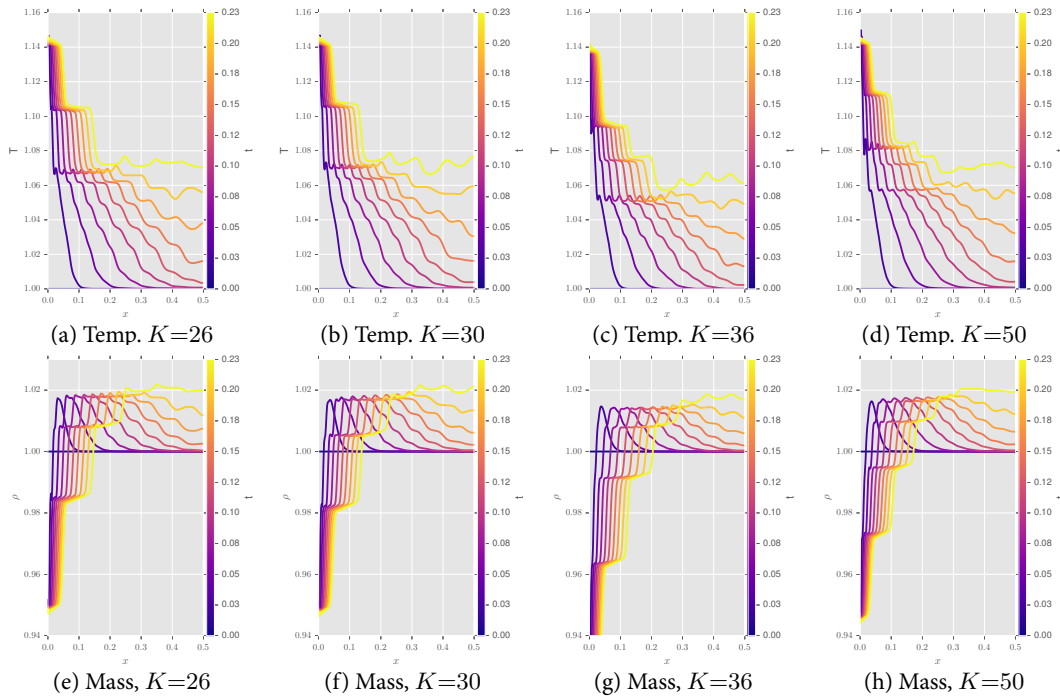


Figure 15: Sudden change in wall temperature $T_w=1.3$: Evolution of the **temperature** $T(x, t)$ and **mass** ρ for $x \in [0, 0.5]$, $t \in [0, 0.23]$. The time evolution is encoded in the color map.

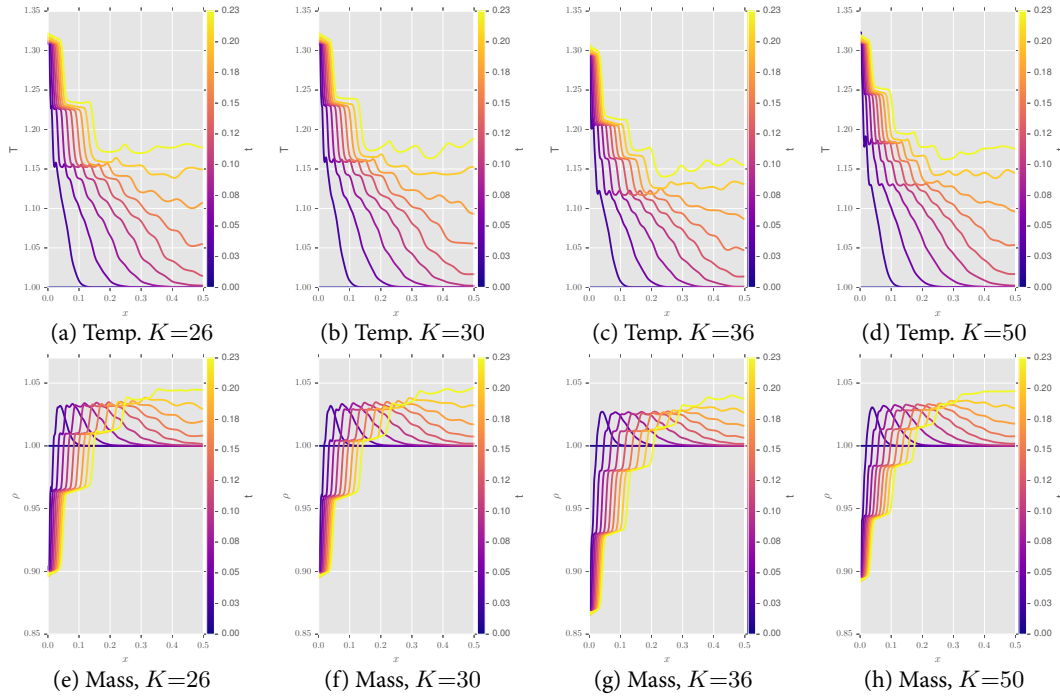


Figure 16: Sudden change in wall temperature $T_w=1.7$: Evolution of $T(x,t)$ and mass ρ for $x \in [0, 0.5], t \in [0, 0.23]$. The time evolution is encoded in the color map.

6.6 Flow generated by a temperature gradient

We consider the same geometry as in the previous example. Diffusive reflective boundary conditions are imposed with $T_l=1$, $T_u=1.44$ at the lower and upper wall. We choose the initial distribution

$$f(t=0, y, \mathbf{v}) = \frac{1}{2\pi T(y)} e^{-\frac{\|\mathbf{v}\|^2}{2T(y)}},$$

$$T(y) = 1 + 1.44y.$$

The simulations were carried out for Knudsen numbers $kn=0.025, 0.1, 1$, until a stationary state was reached with time-step $\Delta t = 10^{-3}$. We observe good agreement in the temperature profiles for $K=20, \dots, 40$.

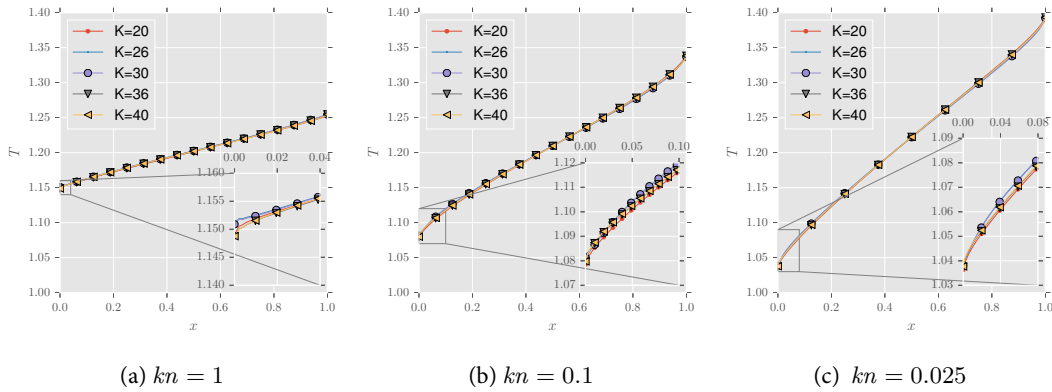


Figure 17: Temperature profiles for the stationary states at time $t = 6, 25, 75$ for $kn = 1, 0.1, 0.025$.

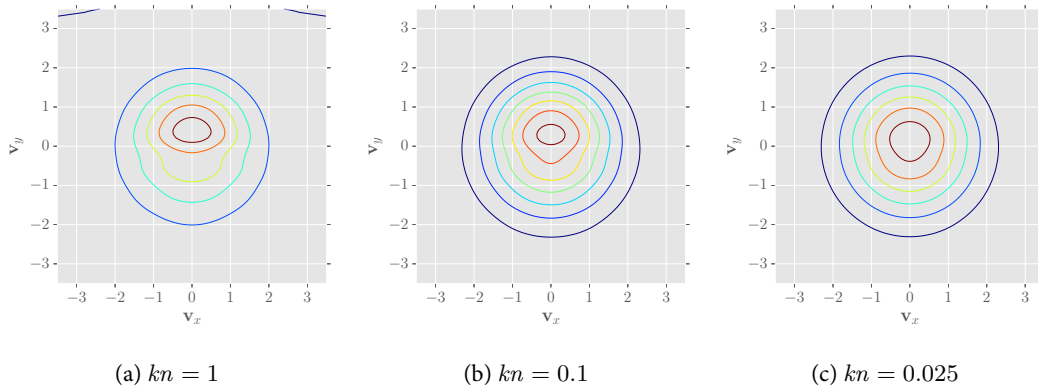


Figure 18: Mass distribution function at the upper wall: $f(t=t_{\text{end}}, y=1, \mathbf{v})$

7 Conclusion

We have presented a combined spectral polynomial and finite element method for the spatially inhomogeneous Boltzmann equation. It can be extended to conserve the lowest moments and include all relevant boundary conditions. We have elaborated it for elastic collisions in the variable hard spheres model. The simulations were carried out for Maxwellian molecules, but in general, any separable collision kernel of the form $C(\cos \theta) \|v - v_*\|$ can be tackled by our scheme. Conservation of mass, momentum and energy can be achieved either by the Petrov-Galerkin approach or the Lagrange multiplier method. The latter way seems to offer better accuracy. Further investigations of this difference in performance will be conducted.

For numerical testing we have implemented an extensive simulation framework in C++ which can deal with different types of boundary conditions on realistic geometries in $2D$. The code has been parallelized using MPI, and provided that the spatial mesh is sufficiently fine, scales well up to a few hundred processors. Details of this implementation will be published separately. We have reported numerical results for low and high-speed flows from the hydrodynamic to the rarefied regime. The polar spectral basis offers fast convergence for smooth solutions. For initial distributions with discontinuities we observe a degradation in convergence with respect to the velocity degrees of freedom. The same holds true for discontinuities in the velocity distribution function imposed by hot or cold walls.

References

- [1] *Molecular gas dynamics and the direct simulation of gas flows*. Ed. by G. A. Bird. Oxford engineering science series 42. Oxford: Clarendon, 1994. 458 pp.
- [2] Kenichi Nanbu. “Direct Simulation Scheme Derived from the Boltzmann Equation. I. Monocomponent Gases”. In: *Journal of the Physical Society of Japan* 49.5 (Nov. 15, 1980), pp. 2042–2049. DOI: 10.1143/JPSJ.49.2042.
- [3] Lorenzo Pareschi and Benoit Perthame. “A Fourier spectral method for homogeneous boltzmann equations”. In: *Transport Theory and Statistical Physics* 25.3 (Apr. 1, 1996), pp. 369–382. DOI: 10.1080/00411459608220707.
- [4] A. Bobylev and S. Rjasanow. “Difference scheme for the Boltzmann equation based on the Fast Fourier Transform”. In: *European Journal of Mechanics, B/Fluids* 16.2 (1997), pp. 293–306.
- [5] A. V. Bobylev and S. Rjasanow. “Fast deterministic method of solving the Boltzmann equation for hard spheres”. In: *European Journal of Mechanics - B/Fluids* 18.5 (Sept. 1999), pp. 869–887. DOI: 10.1016/S0997-7546(99)00121-1.
- [6] Irene M. Gamba and Sri Harsha Tharkabhushanam. “Spectral-Lagrangian methods for collisional models of non-equilibrium statistical states”. In: *Journal of Computational Physics* 228.6 (Apr. 1, 2009), pp. 2012–2036. DOI: 10.1016/j.jcp.2008.09.033.
- [7] Lei Wu et al. “Deterministic numerical solutions of the Boltzmann equation using the fast spectral method”. In: *Journal of Computational Physics* 250 (Oct. 1, 2013), pp. 27–52. DOI: 10.1016/j.jcp.2013.05.003.
- [8] E. Fonn, P. Grohs, and R. Hiptmair. *Polar Spectral Scheme for the Spatially Homogeneous Boltzmann Equation*. 2014-13. Switzerland: Seminar for Applied Mathematics, ETH Zürich, 2014.

- [9] A. Ya 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”. In: *Physics of Fluids (1994-present)* 11.9 (Sept. 1, 1999), pp. 2720–2730. DOI: 10.1063/1.870131.
- [10] G. Kitzler and J. Schöberl. *Efficient Spectral Methods for the spatially homogeneous Boltzmann equation*. 13/2013. Austria: Institute for Analysis and Scientific Computing, TU Wien, 2013.
- [11] G. Kitzler and J. Schöberl. “A high order space–momentum discontinuous Galerkin method for the Boltzmann equation”. In: *Computers & Mathematics with Applications*. High-Order Finite Element and Isogeometric Methods 70.7 (Oct. 2015), pp. 1539–1554. DOI: 10.1016/j.camwa.2015.06.011.
- [12] C Villani. “A review of mathematical topics in collisional kinetic theory”. In: *Handbook of Mathematical Fluid Dynamics, Vol. 1*. Ed. by S Friedlander and D Serre. Elsevier, 2002, pp. 71–305. DOI: 10.1016/S1874-5792(02)80004-0.
- [13] Harold Grad. “Principles of the Kinetic Theory of Gases”. In: *Thermodynamik der Gase / Thermodynamics of Gases*. Ed. by S. Flügge. Handbuch der Physik / Encyclopedia of Physics 3 / 12. Springer Berlin Heidelberg, 1958, pp. 205–294.
- [14] Milton Abramowitz and Irene Stegun. *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*. Dover, June 1972. 358 pp.
- [15] C. Cercignani. “Chapter 1 - The Boltzmann Equation and Fluid Dynamics”. In: *Handbook of Mathematical Fluid Dynamics*. Ed. by S. Friedlander D. Serre. Vol. 1. North-Holland, 2002, pp. 1–69. DOI: 10.1016/S1874-5792(02)80003-9.
- [16] Gene H. Golub and John H. Welsch. “Calculation of Gauss Quadrature Rules”. In: *Mathematics of Computation* 23.106 (1969), 221–s10.
- [17] B Shizgal. “A Gaussian quadrature procedure for use in the solution of the Boltzmann equation and related problems”. In: *Journal of Computational Physics* 41.2 (June 1981), pp. 309–328. DOI: 10.1016/0021-9991(81)90099-1.
- [18] Pavel B. Bochev and Max D. Gunzburger. *Least-squares finite element methods*. Vol. 166. Applied mathematical sciences. New York: Springer, 2009. 660 pp. DOI: 10.1007/b13382.
- [19] Carlo Cercignani. *Rarefied gas dynamics : from basic concepts to actual calculations*. Cambridge texts in applied mathematics. Cambridge: Cambridge University Press, 2000. 320 pp.
- [20] W. Bangerth et al. *The deal.II Library, Version 8.3*.
- [21] Michael A. Heroux et al. “An overview of the Trilinos project”. In: *ACM Trans. Math. Softw.* 31.3 (2005), pp. 397–423. DOI: 10.1145/1089014.1089021.
- [22] Ashley F Emery. “An evaluation of several differencing methods for inviscid fluid flow problems”. In: *Journal of Computational Physics* 2.3 (Feb. 1, 1968), pp. 306–331. DOI: 10.1016/0021-9991(68)90060-0.
- [23] A. Huerta, E. Casoni, and J. Peraire. “A simple shock-capturing technique for high-order discontinuous Galerkin methods”. In: *International Journal for Numerical Methods in Fluids* 69.10 (Aug. 10, 2012), pp. 1614–1632. DOI: 10.1002/flid.2654.

Appendix

7.1 Location of the nonzero entries in the collision tensor

We compute the locations of the nonzero entries for the discretized collision operator

$$\langle Q(\phi_{\tau_1, l_1}, \phi_{\tau_2, l_2}), \phi_{\tau, l} \rangle_{L^2(\mathbb{R}^2)} \quad (75)$$

with trial functions $\phi_{\tau_1, l_1}, \phi_{\tau_2, l_2} \in V_V^N$ and a test function $\phi_{\tau, l} \in \hat{V}_V^N$. The rotation invariance (12) and the bilinearity of the collision operator will be used. Since in general, we do not obtain sparsity with respect to the radial part of the ansatz function we drop the index k of the $\Psi_{k, j}^{\sin, \cos}$ and denote them instead by $\phi_{\tau, l}$:

$$\phi_{\tau_i, l_i}(\varphi, r) := \Psi_{\cdot, j}^{\sin, \cos}(\varphi, r),$$

where l_i is the angular frequency. E.g. $\phi_{\tau_i, l_i} = \tau_i(l_i \varphi) f_r(r)$ for $\tau_i = \sin, \cos, i = 1, 2$ and analogously for the test function $\phi_{\tau, l}$. In the following we will use the rotation operator ρ_ω defined as $\rho_\omega h(\varphi, r) = h(\varphi + \omega, r)$. The rotation invariance applied to the four possible combinations of inputs in τ_1, τ_2 gives the following equations:

$$\begin{aligned} Q(\rho_\omega \phi_{c, l_1}, \rho_\omega \phi_{c, l_2}) &= \rho_\omega Q(\phi_{c, l_1}, \phi_{c, l_2}) \\ Q(\rho_\omega \phi_{c, l_1}, \rho_\omega \phi_{s, l_2}) &= \rho_\omega Q(\phi_{c, l_1}, \phi_{s, l_2}) \\ Q(\rho_\omega \phi_{s, l_1}, \rho_\omega \phi_{c, l_2}) &= \rho_\omega Q(\phi_{s, l_1}, \phi_{c, l_2}) \\ Q(\rho_\omega \phi_{s, l_1}, \rho_\omega \phi_{s, l_2}) &= \rho_\omega Q(\phi_{s, l_1}, \phi_{s, l_2}) \end{aligned} \quad (76)$$

Using the trigonometric identities

$$\begin{aligned} \rho_\omega \sin(l\varphi) &= \cos(l\varphi) \sin(l\omega) + \sin(l\varphi) \cos(l\omega) \\ \rho_\omega \cos(l\varphi) &= \cos(l\varphi) \cos(l\omega) - \sin(l\varphi) \sin(l\omega) \end{aligned} \quad (77)$$

and the bilinearity of Q , (76) is transformed into a 4×4 system of linear equations in the unknowns $Q(\phi_{c, l_1}, \phi_{c, l_2}), Q(\phi_{c, l_1}, \phi_{s, l_2}), Q(\phi_{s, l_1}, \phi_{c, l_2}), Q(\phi_{s, l_1}, \phi_{s, l_2})$:

$$\mathbf{A}(\omega) \mathbf{q}(\varphi) = \mathbf{q}_0(\omega), \quad (78)$$

where

$$\begin{aligned} \mathbf{A}(\omega) &:= \begin{pmatrix} \cos(l_2 \omega) \cos(l_1 \omega) & -\cos(l_1 \omega) \sin(l_2 \omega) & -\cos(l_2 \omega) \sin(l_1 \omega) & \sin(l_2 \omega) \sin(l_1 \omega) \\ \cos(l_1 \omega) \sin(l_2 \omega) & \cos(l_2 \omega) \cos(l_1 \omega) & -\sin(l_2 \omega) \sin(l_1 \omega) & -\cos(l_2 \omega) \sin(l_1 \omega) \\ \cos(l_2 \omega) \sin(l_1 \omega) & -\sin(l_2 \omega) \sin(l_1 \omega) & \cos(l_2 \omega) \cos(l_1 \omega) & -\cos(l_1 \omega) \sin(l_2 \omega) \\ \sin(l_2 \omega) \sin(l_1 \omega) & \cos(l_2 \omega) \sin(l_1 \omega) & \cos(l_1 \omega) \sin(l_2 \omega) & \cos(l_2 \omega) \cos(l_1 \omega) \end{pmatrix} \\ \mathbf{q}(\varphi) &:= \begin{pmatrix} Q(\phi_{c, l_1}, \phi_{c, l_2})(\varphi, r) \\ Q(\phi_{c, l_1}, \phi_{s, l_2})(\varphi, r) \\ Q(\phi_{s, l_1}, \phi_{c, l_2})(\varphi, r) \\ Q(\phi_{s, l_1}, \phi_{s, l_2})(\varphi, r) \end{pmatrix} \end{aligned} \quad (79)$$

and

$$\mathbf{q}_0(\omega) := \begin{pmatrix} Q(\phi_{c,l_1}, \phi_{c,l_2})(\varphi + \omega, r) \\ Q(\phi_{c,l_1}, \phi_{s,l_2})(\varphi + \omega, r) \\ Q(\phi_{s,l_1}, \phi_{c,l_2})(\varphi + \omega, r) \\ Q(\phi_{s,l_1}, \phi_{s,l_2})(\varphi + \omega, r) \end{pmatrix}. \quad (80)$$

setting $\omega = -\varphi$ gives $\mathbf{q}_0(\varphi) = [Q(\phi_{c,l_1}, \phi_{c,l_2})(\varphi \equiv 0, r), 0, 0, 0]^T$. Explicit computation of the inverse of $A(-\varphi)$ gives:

$$\mathbf{A}(-\varphi)^{-1} = \begin{pmatrix} \cos(l_2\varphi) \cos(l_1\varphi) & -\cos(l_1\varphi) \sin(l_2\varphi) & -\cos(l_2\varphi) \sin(l_1\varphi) & \sin(l_2\varphi) \sin(l_1\varphi) \\ \cos(l_1\varphi) \sin(l_2\varphi) & \cos(l_2\varphi) \cos(l_1\varphi) & -\sin(l_2\varphi) \sin(l_1\varphi) & -\cos(l_2\varphi) \sin(l_1\varphi) \\ \cos(l_2\varphi) \sin(l_1\varphi) & -\sin(l_2\varphi) \sin(l_1\varphi) & \cos(l_2\varphi) \cos(l_1\varphi) & -\cos(l_1\varphi) \sin(l_2\varphi) \\ \sin(l_2\varphi) \sin(l_1\varphi) & \cos(l_2\varphi) \sin(l_1\varphi) & \cos(l_1\varphi) \sin(l_2\varphi) & \cos(l_2\varphi) \cos(l_1\varphi) \end{pmatrix}, \quad (81)$$

thus we have simplified (76) to

$$\begin{aligned} Q(\phi_{c,l_1}, \phi_{c,l_2}) &= \cos(l_2\varphi) \cos(l_1\varphi) Q(\phi_{c,l_1}, \phi_{c,l_2})(0, r) \\ Q(\phi_{c,l_1}, \phi_{s,l_2}) &= \cos(l_1\varphi) \sin(l_2\varphi) Q(\phi_{c,l_1}, \phi_{c,l_2})(0, r) \\ Q(\phi_{s,l_1}, \phi_{c,l_2}) &= \cos(l_2\varphi) \sin(l_1\varphi) Q(\phi_{c,l_1}, \phi_{c,l_2})(0, r) \\ Q(\phi_{s,l_1}, \phi_{s,l_2}) &= \sin(l_2\varphi) \sin(l_1\varphi) Q(\phi_{c,l_1}, \phi_{c,l_2})(0, r) \end{aligned} \quad (82)$$

We multiply (82) with the test function $\phi_{\tau,l}$ and integrate over φ .

$$\int_0^{2\pi} \begin{pmatrix} Q(\phi_{c,l_1}, \phi_{c,l_2}) \\ Q(\phi_{c,l_1}, \phi_{s,l_2}) \\ Q(\phi_{s,l_1}, \phi_{c,l_2}) \\ Q(\phi_{s,l_1}, \phi_{s,l_2}) \end{pmatrix} \phi_{\tau,l} d\varphi = Q(\phi_{c,l_1}, \phi_{c,l_2})(0, r) \int_0^{2\pi} \begin{pmatrix} \cos(l_2\varphi) \cos(l_1\varphi) \\ \cos(l_1\varphi) \sin(l_2\varphi) \\ \cos(l_2\varphi) \sin(l_1\varphi) \\ \sin(l_2\varphi) \sin(l_1\varphi) \end{pmatrix} \phi_{\tau,l} d\varphi \quad (83)$$

From (83) we obtain the locations of the nonzero entries depending on τ_1, τ_2, τ and l_1, l_2, l :

- Test function $\phi_{\tau,l}$ with $\tau = \cos$.

$$\langle Q(\phi_{\tau_1,l_1}, \phi_{\tau_2,l_2})(\mathbf{v}), \phi_{\tau,l}(\mathbf{v}) \rangle_{L^2(\mathbb{R}^2)} = \begin{cases} \neq 0 & ((l_1 + l_2) = l \vee |l_1 - l_2| = l) \wedge \tau_1 \neq \tau_2 \\ 0 & \text{otherwise} \end{cases} \quad (84)$$

- Test function $\phi_{\tau,l}$ with $\tau = \sin$.

$$\langle Q(\phi_{\tau_1,l_1}, \phi_{\tau_2,l_2})(\mathbf{v}), \phi_{\tau,l}(\mathbf{v}) \rangle_{L^2(\mathbb{R}^2)} = \begin{cases} \neq 0 & ((l_1 + l_2) = l \vee |l_1 - l_2| = l) \wedge \tau_1 = \tau_2 \\ 0 & \text{otherwise} \end{cases} \quad (85)$$