ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Seminar for Applied Mathematics SAM

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

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

Funding SNF: 146356

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 employ a (Petrov)-Galerkin discretization in the velocity variable of the Boltzmann collision operator based on Laguerre polynomials times a Maxwellian. The advection problem in phase space is discretized by combining the spectral basis with continuous first order finite elements in space resulting in an implicit in time Galerkin least squares formulation. 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 interaction in d-dimensional space are modeled by the 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 [2] and [15]. Among deterministic approaches Fourier methods are most popular. In [16] Pareschi et al. introduced a Fourier based method, related approaches have been introduced in [3, 4, 11, 21]. 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 [9] for the spatially homogeneous Boltzmann equation extending the work done in [7]. 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 $O(K^5)$, where K is the polynomial degree in one velocity direction, see Section 3. We also explore ways to ensure discrete conservation of mass, momentum, and energy, see Section 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 5 we elaborate how to incorporate various physically relevant spatial boundary conditions into our new scheme.

For timestepping we rely on a splitting scheme, which separately treats collisions and advection. For the former we opt for explicit timestepping, 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 admit 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 6.

Closely related and conducted parallel to our investigations is the work by Kitzler and Schöberl [10, 14]. 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 it's 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 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

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

with initial distribution

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

Boundary conditions are prescribed on the inflow boundary $\Gamma^-(\mathbf{V}_w)$ with drift velocity $\mathbf{V}_w$:

$$\begin{aligned} \text{inflow boundary} \quad & \Gamma^-(\mathbf{V}_w) := \{(\mathbf{x}, \mathbf{v}) : \mathbf{x} \in \partial D \wedge (\mathbf{v} - \mathbf{V}_w) \cdot \mathbf{n} \leq 0\}, \\ \text{outflow boundary} \quad & \Gamma^+(\mathbf{V}_w) := \{(\mathbf{x}, \mathbf{v}) : \mathbf{x} \in \partial D \wedge (\mathbf{v} - \mathbf{V}_w) \cdot \mathbf{n} > 0\}, \end{aligned}$$

where $\mathbf{n}$ denotes outward unit normal vector. Whenever we omit the argument of $\Gamma^\pm$ we mean $\Gamma^\pm(\mathbf{V}_w{=}\mathbf{0})$. Common types of boundary conditions are inflow, specular reflective and diffusive reflective boundary conditions [18, Sec. 1.5].

Inflow boundary conditions

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

Specular reflective boundary conditions

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

The particles behave like billard balls at the wall.

Diffusive reflective boundary conditions The particles are absorbed at the wall and reemitted with Maxwellian distribution M_w.

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

where

$$M_w := \frac{1}{(2\pi)^{\frac{1}{2}} T_w^{\frac{3}{2}}} e^{-\frac{\|\mathbf{v}-\mathbf{V}_w\|^2}{2T_w}}, \tag{6}$$

is a Maxwellian distribution at the boundary, and

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

M_w is normalized such that $\int_{\Gamma^+(\mathbf{V}_w)} \mathbf{n}\cdot(\mathbf{v}-\mathbf{V}_w) M_w(\mathbf{v})\,\mathrm{d}\mathbf{v} = 1$.

Macroscopic quantities of the gas can be computed in terms of moments of the distribution function f as follows.

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

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

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

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

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

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

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

$$\begin{aligned} \mathbf{v}' &= \frac{\mathbf{v}+\mathbf{v}_\star}{2} + \sigma\frac{\|\mathbf{v}-\mathbf{v}_\star\|}{2} \\ \mathbf{v}'_\star &= \frac{\mathbf{v}+\mathbf{v}_\star}{2} - \sigma\frac{\|\mathbf{v}-\mathbf{v}_\star\|}{2} \end{aligned} \qquad \sigma \in \mathbb{S}^1. \tag{10}$$

We assume that the interaction potential governing collisions is described by the collision kernel B of the form [18]:

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

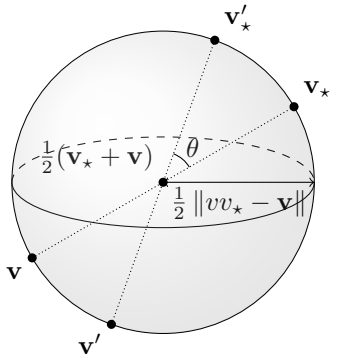


Figure 1

and that $C(\cos\theta)$ satisfies Grad's cutoff assumption [12]:

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

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 translation $\tau^*(\mathbf{c})$ and rotation operator $\rho^*(\omega)$ act on a function $f : \mathbb{R}^2 \to \mathbb{R}$ as follows:*

$$\tau^*(\mathbf{c})f(\mathbf{v}) := f(\mathbf{v}+\mathbf{c}), \qquad \textit{for } \mathbf{c} \in \mathbb{R}^2$$
$$\rho^*(\omega)f(\varphi, r) := f(\varphi+\omega, r), \qquad \textit{for } \omega \in [0, 2\pi[$$

Theorem 1.2. *It holds that*

$$Q(\rho^*(\omega)f, \rho^*(\omega)g)(\varphi, r) = \rho^*(\omega)Q(f,g)(\varphi, r) \tag{12}$$
$$Q(\tau^*(\mathbf{c})f, \tau^*(\mathbf{c})g)(\varphi, r) = \tau^*(\mathbf{c})Q(f,g)(\varphi, r) \tag{13}$$

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

2 Spectral Velocity Space

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 $e^{-r^2/2}$.

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

$$\Psi^a_{k,j}(\varphi, r) := \begin{cases} a(2j\varphi)\, r^{2j} L^{(2j)}_{\frac{k}{2}-j}(r^2) e^{-r^2/2} & k \in 2\mathbb{N} \\ a((2j+1)\varphi)\, r^{2j+1} L^{(2j+1)}_{\frac{k-1}{2}-j}(r^2) e^{-r^2/2} & k \in 2\mathbb{N}+1 \end{cases}$$

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

$\Psi_{k,j}$ are orthogonal in the inner product $\langle f, g\rangle := \int_{\mathbb{R}^2} f(\mathbf{v})g(\mathbf{v})\,\mathrm{d}\mathbf{v}$ [1, Chap. 22]. We define the spectral basis $V_{\mathcal{V}}^N$ of polynomial degree K and total number of elements N:

$$V_{\mathcal{V}}^N := \{\mathbb{L}_k^{\cos} : k = 0, \dots, K\} \cup \left\{\mathbb{L}_k^{\sin} : k = 0, \dots, K\right\}, \tag{14}$$

where

$$\begin{aligned} \mathbb{L}_k^{\cos} &:= \left\{\Psi_{k,j}^{\cos} : j = 0 \dots \lfloor \tfrac{k}{2} \rfloor\right\} \\ \mathbb{L}_k^{\sin} &:= \left\{\Psi_{k,j}^{\sin} : j = 1 - (k \bmod 2) \dots \lfloor \tfrac{k}{2} \rfloor\right\}. \end{aligned} \tag{15}$$

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 symbol $V_{\mathcal{V}}^N$.

Remark 2.2. *In [9] 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 $k \equiv l \mod 2$. Choose for example $k = 1, l = 1$:

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

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

Lemma 2.3. *[10, Lemma 5] The Polar-Laguerre basis functions $\Psi_{k,j}^{\cos,\sin}$ are polynomials of total degree k in Cartesian coordinates weighted by $e^{-r^2/2}$.*

Proof. [10, Lemma 5] Use that

$$\cos n\varphi = \sum_{j=0}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{2j} \sin(\varphi)^{2j} \cos(\varphi)^{n-2j}, \quad \sin n\varphi = \sum_{j=0}^{\lfloor \frac{n-1}{2} \rfloor} \binom{n}{2j+1} \sin(\varphi)^{2j+1} \cos(\varphi)^{n-2j-1} \tag{16}$$

For k even:

$$\begin{aligned} \Psi_{k,j}^{\cos} e^{r^2/2} &= \sum_{i=0}^{j} \binom{2j}{2i} \sin(\varphi)^{2j} \cos(\varphi)^{2j-2i} r^{2j} L_{\frac{k}{2}-j}^{(2j)}(r^2) \\ &= \sum_{i=0}^{j} \binom{2j}{2i} (r\sin(\varphi))^{2j} (r\cos(\varphi))^{2j-2i} L_{\frac{k}{2}-j}^{(2j)}(r^2) \\ &= \sum_{i=0}^{j} \binom{2j}{2i} y^{2j} x^{2j-2i} L_{\frac{k}{2}-j}^{(2j)}(r^2) \end{aligned} \tag{17}$$

$L_{\frac{k}{2}-j}(r^2)$ 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. □

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

$$f^{\mathrm{P}}(\varphi, r) = \sum_{j=0}^{N-1} c_j^{\mathrm{P}} b_j(\varphi, r). \tag{18}$$

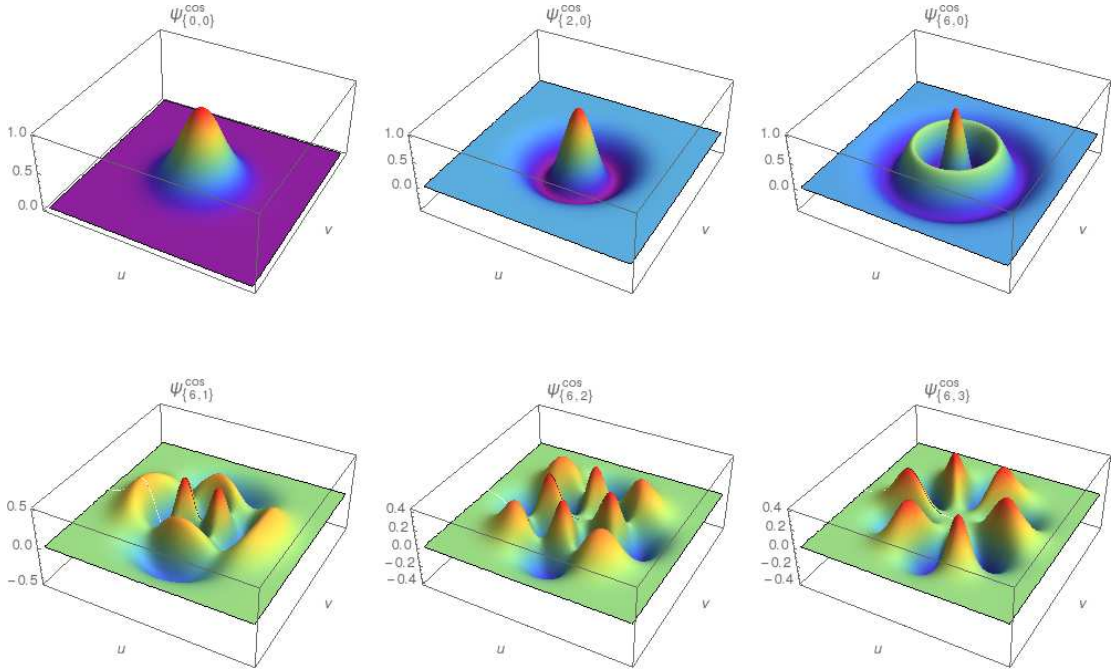


Figure 2: Polar-Laguerre basis functions $\Psi_{k,j}^{\cos}(\mathbf{v})$, $\mathbf{v} \in [-5, 5]^2$
First row: $j = 0, k = 0, 2, 6$ **Second row**: $k = 6, j = 1, 2, 3$

For later usage we define also the Hermite and nodal basis.

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

$$f^H(x, y) = \sum_{k=0}^{K} \sum_{s=0}^{k} c_{s,k-s} h_s(x) e^{-\frac{x^2}{2}} h_{k-s}(y) e^{-\frac{y^2}{2}}. \tag{19}$$

Where are $h_i(x)$ suitably normalized Hermite polynomials [1], such that $\int_{\mathbb{R}} h_i(x) h_j(x) e^{-x^2} \, \mathrm{d}x = \delta_{i,j}$.

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

degree K reads:

$$f^N(x,y) = \sum_{i=0}^{K}\sum_{j=0}^{K} c^{\mathrm{N}}_{i,j}\ell_i(x)e^{-\frac{x^2}{2}}\ell_j(y)e^{-\frac{y^2}{2}}, \tag{20}$$

where ℓ_i are the Lagrange polynomials at the Gauss-Hermite quadrature nodes x_i with weights w_i.

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

The normalization has been chosen such that $\langle \ell_i(x), \ell_j(x)e^{-x^2}\rangle = \delta_{i,j}$.

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

3 Treatment of the Collision Operator

3.1 Petrov-Galerkin discretization in velocity coordinate

The following derivation is identical to the one presented in [9], except that we use a real valued basis in φ. Consider the homogeneous Boltzmann equation

$$\partial_t f = Q(f,f). \tag{21}$$

Multiplication with a test function $\hat{b}_i$ from the yet unspecified function space $\hat{V}_{\mathcal{V}}^N$ and integration over $\mathbb{R}^d$ gives

$$\partial_t \int_{\mathbb{R}^d} f(t,\mathbf{v})g(\mathbf{v})\,\mathrm{d}\mathbf{v} = \int_{\mathbb{R}^d} Q(f,f)g(\mathbf{v})\,\mathrm{d}\mathbf{v}. \tag{22}$$

Making the ansatz $f \in V_{\mathcal{V}}^N$ in (22) gives rise to a 3-dimensional tensor Q_N. One may think of it as an array of matrices $\mathbf{S}_i, i = 0,\dots,N-1$, where slice $\mathbf{S}_i$ is obtained by testing with $\hat{b}_i \in \hat{V}_{\mathcal{V}}^N$:

$$(\mathbf{S}_i)_{i_1,i_2} := \left\langle Q(b_{i_1}, b_{i_2}), \hat{b}_i \right\rangle_{L^2(\mathbb{R}^2)}, \quad b_{i_1}, b_{i_2} \in V_{\mathcal{V}}^N \tag{23}$$

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

$$\begin{aligned}(\mathbf{S}_i^-)_{i_1,i_2} &= \left\langle Q^-(b_{i_1}, b_{i_2}), \hat{b}_i \right\rangle \\ &= \int_{\mathbb{R}^2}\int_{\mathbb{R}^2} B(\|\mathbf{v}-\mathbf{v}_\star\|, \cos\theta) b_{i_1}(\mathbf{v}) b_{i_2}(\mathbf{v}_\star)\hat{b}_i(\mathbf{v})\,\mathrm{d}\sigma\,\mathrm{d}\mathbf{v}_\star\,\mathrm{d}\mathbf{v} \\ &= \int_{\mathbb{R}^2} b_{i_1}(\mathbf{v})\hat{b}_i(\mathbf{v})\int_{\mathbb{R}^2} b_{i_2}(\mathbf{v}_\star)\mathcal{I}^-(\mathbf{v},\mathbf{v}_\star)\,\mathrm{d}\mathbf{v}_\star\,\mathrm{d}\mathbf{v}\end{aligned} \tag{24}$$

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

$$\mathcal{I}^- = \int_{\mathbb{S}^1} B(\|\mathbf{v}-\mathbf{v}_\star\|, \cos\theta)\,\mathrm{d}\sigma \tag{25}$$

$$= \|\mathbf{v}-\mathbf{v}_\star\|^\lambda \int_{\mathbb{S}^1} C(\cos\theta)\,\mathrm{d}\sigma, \tag{26}$$

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

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

with

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

Note that, in the second line of (27), we have made the change of variables $\mathbf{v}, \mathbf{v}_\star \leftrightarrow \mathbf{v}', \mathbf{v}'_\star$. Next, we substitute $\mathbf{w}' = \mathrm{R}_\alpha \mathbf{v}'$, $\alpha = -\arg(\mathbf{v}+\mathbf{v}_\star)$. R_α denotes the rotation by α around the origin in counter clockwise direction. Remember that the test function $\hat{b}_i \in \hat{V}_\mathcal{V}^N$ has the form $a(l\varphi)\,\phi_r(r)$, where a is either $\sin$ or $\cos$.

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

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

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

$$= \sin(l\alpha)\,\|\mathbf{v}'-\mathbf{v}'_\star\|^\lambda C \int_{\mathbb{S}^1} \cos(l\arg(\mathbf{w}'))\phi_r(\|\mathbf{w}'\|)\,\mathrm{d}\sigma, \tag{31}$$

and for $a = \cos$

$$\begin{aligned}\mathcal{I}^{(+)}(\mathbf{v}', \mathbf{v}'_\star; \hat{b}_i) = \quad & \|\mathbf{v}'-\mathbf{v}'_\star\|^\lambda C \int_{\mathbb{S}^1} \Big[\cos(l\arg(\mathbf{w}'))\cos(l\alpha) \\ & - \sin(l\arg(\mathbf{w}')\sin(l\alpha))\Big]\phi_r(\|\mathbf{w}'\|)\,\mathrm{d}\sigma\end{aligned} \tag{32}$$

$$= \cos(l\alpha)\,\|\mathbf{v}'-\mathbf{v}'_\star\|^\lambda C \int_{\mathbb{S}^1} \cos(l\arg(\mathbf{w}'))\phi_r(\|\mathbf{w}'\|)\,\mathrm{d}\sigma. \tag{33}$$

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

3.2 Conservative discretization

An important property of (21) is that mass, momentum and energy are conserved. In particular it holds that

$$\partial_t \begin{pmatrix} \rho(f) \\ \rho \mathbf{u}(f) \\ \rho E(f) \end{pmatrix} = \int_{\mathbb{R}^2} Q(f,f) \begin{pmatrix} 1 \\ \mathbf{v} \\ \|\mathbf{v}\|^2 \end{pmatrix} \mathrm{d}\mathbf{v} \equiv 0, \tag{34}$$

by fundamental properties of the Boltzmann collision operator [5, sec. 5]. This condition can be naturally enforced for (22) by choosing $\hat{V}_{\mathcal{V}}^N$ such that $1, \mathbf{v}, \|\mathbf{v}\|^2 \in \operatorname{span} \hat{V}_{\mathcal{V}}^N$. Inspection of the basis functions from $V_{\mathcal{V}}^N$ reveals that it is sufficient to multiply a few of them by a factor $e^{r^2/2}$ to conserve mass, momentum and energy:

$$\begin{aligned} \Psi_{0,0}^{\cos} \exp(r^2/2) &= 1 \\ \Psi_{1,0}^{\sin} \exp(r^2/2) &= \sin(\varphi)\, r \\ \Psi_{1,0}^{\cos} \exp(r^2/2) &= \cos(\varphi)\, r \\ \Psi_{2,0}^{\cos} \exp(r^2/2) &= (1 - r^2) \end{aligned} \tag{35}$$

Therefore we use a test space $\hat{V}_{\mathcal{V}}^N$ which is identical to $V_{\mathcal{V}}^N$, 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:

$$Q_N(f^P, g^P)(\mathbf{v}) = \sum_{i=1}^{N} \left(\mathbf{M}^{-1} [\mathbf{c}^T \mathbf{S}_j \mathbf{d}]_{j=1}^N \right)_i b_i(\mathbf{v}), \tag{36}$$

where $(\mathbf{M})_{j,j'} = \langle b_j, b_{j'} \rangle$, $f^P, g^P \in V_{\mathcal{V}}^N$ 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 the application of Q_N is $\mathcal{O}(K^6)$. In the next section we show that, due the polar representation, the complexity can actually be reduced by a factor K.

Galerkin discretization with Lagrange multipliers Alternatively one can also use a Galerkin discretization and solve a constrained minimization problem wrt. the L_2-norm such that mass, momentum and energy are conserved. This has been proposed in [11] for the Fourier-Spectral 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 without conservation:

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

2. Solve the constrained minimization problem:

$$\mathbf{c}^{k+1} = \min_{\mathbf{c}_\star^{k+1} \in \mathbb{R}^N} \left\| \mathbf{c}_\star^{k+1} - \tilde{\mathbf{c}}^{k+1} \right\|^2 + \underbrace{\lambda^T \mathbf{H}^T (\mathbf{c}_\star^{k+1} - \mathbf{c}^k)}_{\text{conservation of mass, momentum and energy}}, \tag{37}$$

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

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

The solution to (37) is

$$\mathbf{c}^{k+1} = \tilde{\mathbf{c}}^{k+1} - \frac{1}{2}\mathbf{H}\lambda, \tag{38}$$

with

$$\lambda = 2(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T(\tilde{\mathbf{c}}^{k+1} - \mathbf{c}^k). \tag{39}$$

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

3.3 Computational aspects

It can be shown that Q_N has nonzero entries for $k + l = j$ or $|k - l| = j$, where k, l denote the angular frequency of the trial functions and j is the angular frequency of the test function. The proof follows immediately with the help of trigonometric identities and the bilinearity of Q. The calculations can be found in the appendix 7.1. For the sake of simplicity, we quote the proof from [9], which uses exactly the same technique with a complex valued Fourier discretization for the angular part of the spectral basis.

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

$$f(r,\varphi) = f_r(r)e^{\mathrm{i}\,k\varphi}, \quad g(r,\varphi) = g_r(r)e^{\mathrm{i}\,l\varphi}$$

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

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

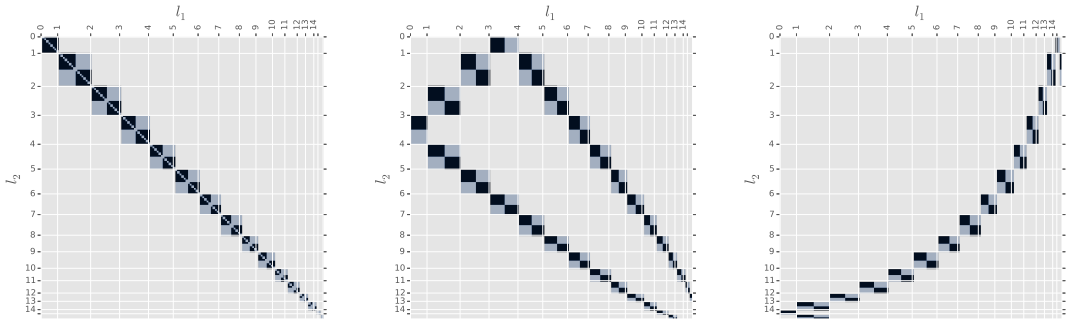
Proof. We get $\rho^*(\omega)f = e^{\mathrm{i}\,k\omega}f$, and correspondingly for g. Using (12) and the bilinearity of Q we obtain

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

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

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

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



(a) $\cos(0\varphi)$ (b) $\cos(3\varphi)$ (c) $\cos(14\varphi)$

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 lexically by $(l, \cos / \sin, k)$, where l is the angular frequency and k denotes the polynomial degree in radial direction.

Corollary 3.2. *The consequence of 3.1 is that each* $\mathbf{S}_i$ *from* (36) *only has* $\mathcal{O}(K^3)$ *nonzero entries.* Q_N *has a total of* $\mathcal{O}(K^5)$ *nonzero entries.*

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

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 theorem (1.2) Q is also invariant to translation. A Maxwellian at conforming temperature, i.e. $T = 1$, centered at the origin is represented in the polar basis by a single non-zero coefficient. In order to represented the same Maxwellian centered at $\mathbf{v}_0$ with same accuracy the required polynomial degree K grows with $\|\mathbf{v}_0\|$. On the other hand, if we want to apply the scattering operator to a given function, it would be beneficial to perform first a change of variables such that it has zero momentum, apply the scattering 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 non-zero momentum counterpart and thus one might truncate at a lower K without loss of accuracy. The straight forward 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 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 $f : \mathbb{R}^2 \to \mathbb{R}$ reads

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

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 $\mathbf{T}_{\mathrm{P}\to\mathrm{H}}, \mathbf{T}_{\mathrm{H}\to\mathrm{P}}$ used to transform Polar-Laguerre to Hermite coefficients and vice versa:

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

where $\mathbf{T}_{\mathrm{P}\to\mathrm{H}}, \mathbf{T}_{\mathrm{H}\to\mathrm{P}} \in \mathbb{R}^{N\times N}$, because of their block-diagonal structure with dense blocks of size $k+1, k = 0, K-1$, the cost to transform coefficients from the Polar-Laguerre to Hermite basis is $\mathcal{O}(K^3)$. The derivation of the Polar-Laguerre to Hermite transformation matrices can be found in [14, Sec. 3.2].

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

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

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

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

where $s_i = \int_{\mathbb{R}} h_i(x) h_i(x) e^{-x^2} \,\mathrm{d}x$. 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 $x = x - \frac{\bar{x}}{2}$

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

and use the identity

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

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

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

where we have used the orthogonality of the Hermite polynomials.

To carry out the shifting in $2D$, we reshape the coefficient vector into a lower triangular matrix $\mathbf{C} \in \mathbb{R}^{K,K}$ in the following way:

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

The shift matrices $\mathbf{S}^{\bar{x}}, \mathbf{S}^{\bar{y}}$ multiplied with from the left with a coefficient matrix act on its columns:

$$\begin{aligned}\bar{\mathbf{C}}^T &= \mathbf{S}^{\bar{y}}(\mathbf{S}^{\bar{x}}\mathbf{C})^T \\ \Leftrightarrow \bar{\mathbf{C}} &= \mathbf{S}^{\bar{x}}\,\mathbf{C}\,\mathbf{S}^{\bar{y},T}.\end{aligned} \tag{48}$$

To avoid numerical overflow orthonormal Hermite polynomials have been used in the implementation. The procedure described above is summarized in Algorithm 1, whose total cost without the scattering is $\mathcal{O}(K^3)$.

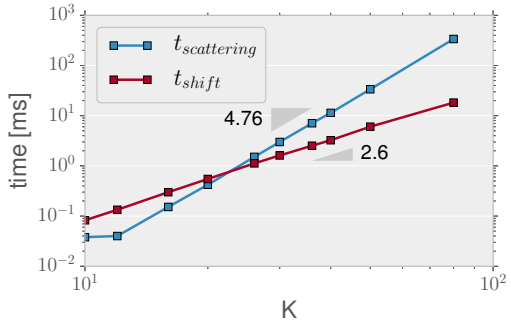
Algorithm 1 Scattering in re-centered basis via Hermite representation. (Superscripts P, H denote coefficients in Polar-Laguerre / Hermite basis.)

procedure APPLY SCATTERING IN RE-CENTERED BASIS($\mathbf{c}^{\mathrm{P}}$)
  $\mathbf{c}^{\mathrm{H}} \leftarrow \mathbf{T}_{\mathrm{P}\to\mathrm{H}}\mathbf{c}^{\mathrm{P}}$ ▷ Transform to Hermite basis
  $\bar{\mathbf{c}}^{\mathrm{H}} \leftarrow \mathbf{S}^{\bar{\mathbf{x}}}\mathbf{c}^{\mathrm{H}}$ ▷ Transform to zero momentum
  $\bar{\mathbf{c}}^{\mathrm{P}} \leftarrow \mathbf{T}_{\mathrm{H}\to\mathrm{P}}\bar{\mathbf{c}}^{\mathrm{H}}$ ▷ Go back to Polar-Laguerre basis
  $\bar{\mathbf{c}}^{\mathrm{P}} \leftarrow$ update with Q in truncated basis
  $\bar{\mathbf{c}}^{\mathrm{H}} \leftarrow \mathbf{T}_{\mathrm{P}\to\mathrm{H}}\bar{\mathbf{c}}^{\mathrm{P}}$ ▷ Transform to Hermite basis
  $\mathbf{c}^{\mathrm{H}} \leftarrow \mathbf{S}^{-\bar{\mathbf{x}}}\bar{\mathbf{c}}^{\mathrm{H}}$ ▷ Shift back
  $\mathbf{c}^{\mathrm{P}} \leftarrow \mathbf{T}_{\mathrm{H}\to\mathrm{P}}\mathbf{c}^{\mathrm{H}}$ ▷ Transform to Polar-Laguerre basis
end procedure

Fig. 4 displays timings for the shifting procedure and the scattering for varying polynomial degree K. For $K < 40$ the shifting does not payoff, since it is slower than the scattering operator.

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.

$$f(\mathbf{v}) = \exp(-\mathbf{v}^T\mathbf{M}\mathbf{v}) + \exp(-\tfrac{\|\mathbf{v}-\mathbf{v}_c\|^2}{2}), \tag{49}$$



(a)

K	$t_{\text{shift}}[ms]$	$t_{\text{scattering}}[ms]$
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

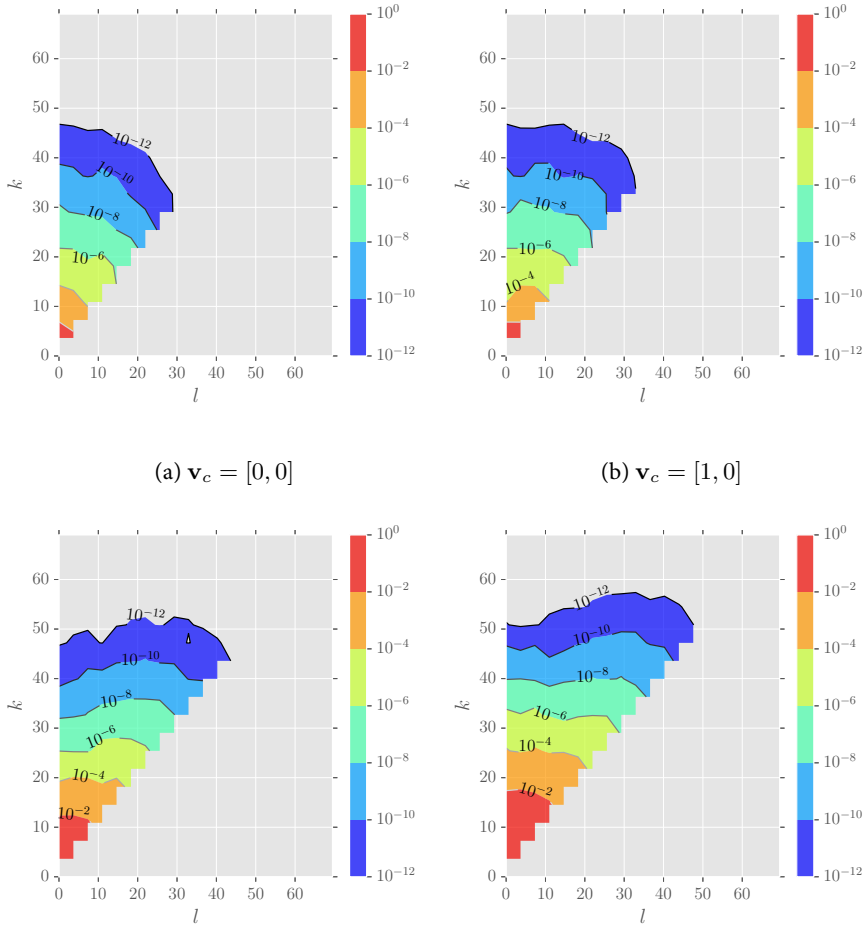
(b)

Figure 4: CPU-time: Intel Core i7 4790K (4GHz, single threaded), Linux 4.2.3, GCC 5.2.0, relevant compiler flags: `-O3 -msse2 -mavx`. t_{shift} contains all of Algorithm 1 except the scattering.

where

$$\mathbf{M} = \frac{1}{8}\begin{bmatrix} 7 & \sqrt{3} \\ \sqrt{3} & 5 \end{bmatrix}, \qquad \mathbf{v}_c = [0.2, 0]\,. \tag{50}$$

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



(a) $\mathbf{v}_c = [0, 0]$

(b) $\mathbf{v}_c = [1, 0]$

(c) $\mathbf{v}_c = [3, 0]$

(d) $\mathbf{v}_c = [4, 0]$

Figure 5: Decay of Polar-Laguerre coefficients for $f(\mathbf{v} - \mathbf{v}_c)$, defined in (49), with respect to angular index l and radial index k.

4 Discretization in Physical Space

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 $\left\langle \mathbf{v}\cdot\mathbf{n}\Phi, f\right\rangle_\Gamma$ appears in the variational formulation, which comes handy to include inflow-type boundary conditions.

The advection part of (1) reads

$$\partial_t f + \mathbf{v}\cdot\nabla_{\mathbf{x}} f = 0. \tag{51}$$

We replace $\partial_t f$ in (51) by a backwards difference quotient and write down the least squares functional $J(f^{(n)}; f^{(n-1)})$ for the pure transport problem:

$$J(f^{(n)}; f^{(n-1)}) := \left\| \frac{1}{\Delta t}\left(f^{(n)} - f^{(n-1)}\right) + \mathbf{v}\cdot\nabla_{\mathbf{x}} f^{(n)}\right\|^2_{L^2(\Omega)} \tag{52}$$

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

$$a(\Phi, f^{(n)}) = \frac{1}{\Delta t^2}\left\langle \Phi, f^{(n)}\right\rangle_\Omega + \frac{1}{\Delta t}\left\langle \mathbf{v}\cdot\mathbf{n}\,\Phi, f^{(n)}\right\rangle_\Gamma + \left\langle \mathbf{v}\cdot\nabla_{\mathbf{x}}\Phi, \mathbf{v}\cdot\nabla_{\mathbf{x}} f^{(n)}\right\rangle_\Omega, \tag{53}$$

and

$$b(\Phi, f^{(n-1)}) := \frac{1}{\Delta t^2}\left\langle \Phi, f^{(n-1)}\right\rangle_\Omega + \frac{1}{\Delta t}\left\langle \mathbf{v}\cdot\nabla_{\mathbf{x}}\Phi, f^{(n-1)}\right\rangle_\Omega, \tag{54}$$

where $\mathbf{n}$ is the unit outward normal vector on ∂D. In the following we use $\langle\cdot,\cdot\rangle$ to denote the L_2 inner product. $V_{\mathcal{D}}^L$ 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 by the tensor product space$V^{L,N} = V_{\mathcal{D}}^L \otimes V_{\mathcal{V}}^N$. The test functions Φ are also taken from $V^{L,N}$. The superscript L will denote the number of degrees of freedoms in physical space.

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

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

$$\begin{aligned}\frac{1}{\Delta t_k^2}\left\langle \Phi, f^{(n+1/2)}\right\rangle_\Omega + \frac{1}{\Delta t_k}\left\langle \mathbf{v}\cdot\mathbf{n}\,\Phi, f^{(n+1/2)}\right\rangle_\Gamma + \left\langle \mathbf{v}\cdot\nabla_{\mathbf{x}}\Phi, \mathbf{v}\cdot\nabla_{\mathbf{x}} f^{(n+1/2)}\right\rangle_\Omega \\ = \frac{1}{\Delta t_k^2}\left\langle \Phi, f^{(n)}\right\rangle_\Omega + \frac{1}{\Delta t_k}\left\langle \mathbf{v}\cdot\nabla_{\mathbf{x}}\Phi, f^{(n)}\right\rangle_\Omega\end{aligned} \tag{55}$$

2. **Scattering** (explicit Euler):

$$f^{(n+1)} = f^{(n+1/2)} + \frac{\Delta t_k}{kn} Q(f^{(n+1/2)}, f^{(n+1/2)}) \tag{56}$$

5 Boundary Conditions

We discuss inflow, specular reflective and diffusive reflective boundary conditions defined in (3),(4) and (5). These are of the simplest types available. There exist more physical models, a detailed disussion can be found in [6, Ch. 1.11] and the references therein.

5.1 Discretization

We integrate the second term of $a(\Phi, f)$, cf. (53), by parts and obtain

$$\frac{1}{\Delta t}\Big\langle \mathbf{v}\cdot\mathbf{n}\,\Phi, f^{(n)}\Big\rangle_{\Gamma} = \frac{1}{\Delta t}\Big\langle \mathbf{v}\cdot\mathbf{n}\,\Phi, f^{(n)}\Big\rangle_{\Gamma^+} + \frac{1}{\Delta t}\Big\langle \mathbf{v}\cdot\mathbf{n}\Phi, f^{(n)}\Big\rangle_{\Gamma^-}, \tag{57}$$

where $\Gamma = \partial D \times \mathbb{R}^2$.

$f^{(n)}$ in $\Big\langle \mathbf{v}\cdot\mathbf{n}, f^{(n)}\Big\rangle_{\Gamma^+}$ can be replaced by the respective boundary condition.

Conservation properties Multiply (1) by $\left(1, \mathbf{v}, \|\mathbf{v}\|^2\right)$ and integrate over Ω to obtain:

$$\partial_t \int_D \int_{\mathbb{R}^2} f(t,\mathbf{x},\mathbf{v}) \begin{pmatrix} 1 \\ \mathbf{v} \\ \|\mathbf{v}\|^2 \end{pmatrix} \mathrm{d}\mathbf{v}\,\mathrm{d}\mathbf{x} \equiv \partial_t \begin{pmatrix} \rho \\ \rho\mathbf{u} \\ \rho E \end{pmatrix} = -\int_D \int_{\mathbb{R}^2} \begin{pmatrix} 1 \\ \mathbf{v} \\ \|\mathbf{v}\|^2 \end{pmatrix} \mathbf{v}\cdot\nabla_{\mathbf{x}} f(t,\mathbf{x},\mathbf{v})\,\mathrm{d}\mathbf{v}\,\mathrm{d}\mathbf{x} \tag{58}$$

$$\Leftrightarrow \partial_t \begin{pmatrix} \rho \\ \rho\mathbf{u} \\ \rho E \end{pmatrix} = -\int_{\Gamma^+} \mathbf{n}(\mathbf{x})\cdot\mathbf{v} \begin{pmatrix} 1 \\ \mathbf{v} \\ \|\mathbf{v}\|^2 \end{pmatrix} f(t,\mathbf{x},\mathbf{v})\,\mathrm{d}\mathbf{v}\,\mathrm{d}\mathbf{x} - \int_{\Gamma^-} \mathbf{n}(\mathbf{x})\cdot\mathbf{v} \begin{pmatrix} 1 \\ \mathbf{v} \\ \|\mathbf{v}\|^2 \end{pmatrix} f(t,\mathbf{x},\mathbf{v})\,\mathrm{d}\mathbf{v}\,\mathrm{d}\mathbf{x} \tag{59}$$

If we insert the condition for specular reflection (4) into (59), one easily observes that the right-hand side of (59) evaluates to zero and therefore mass and energy are conserved.

Theorem 5.1. *For diffusive reflective boundary conditions mass is conserved.*

Proof. It must hold that

$$\begin{aligned} &\int_{\mathbf{x}\in\partial D} \int_{\mathbf{v}\cdot\mathbf{n}(\mathbf{x})>0} \mathbf{n}(\mathbf{x})\cdot\mathbf{v} f(t,\mathbf{x},\mathbf{v})\,\mathrm{d}\mathbf{v}\,\mathrm{d}\mathbf{x} \\ &\qquad = -\int_{\mathbf{x}\in\partial D} \int_{\mathbf{v}\cdot\mathbf{n}(\mathbf{x})<0} \mathbf{n}(\mathbf{x})\cdot\mathbf{v} f(t,\mathbf{x},\mathbf{v})\,\mathrm{d}\mathbf{v}\,\mathrm{d}\mathbf{x} \\ &\qquad = +\int_{\mathbf{x}\in\partial D} \rho_+(f) \underbrace{\int_{\mathbf{v}\cdot\mathbf{n}(\mathbf{x})>0} \mathbf{n}(\mathbf{x})\cdot\mathbf{v}\, M_w(\|\mathbf{v}\|)\,\mathrm{d}\mathbf{v}}_{\equiv 1}\,\mathrm{d}\mathbf{x}, \end{aligned} \tag{60}$$

where we have made the changes of variables $\mathbf{v} \to \mathbf{v} - 2\mathbf{n}\langle\mathbf{v},\mathbf{n}\rangle$ in the third line. By definition the LHS of (60) is $\int_{\mathbf{x}\in\partial D} \rho_+(f)\,\mathrm{d}\mathbf{x}$, which finishes the proof. □

Theorem 5.2. *Mass and energy are conserved in the discrete formulation with specular reflective boundary conditions.*

Proof. Insert a test function Φ which is constant in $\mathbf{x}$ into the variational formulation $\mathrm{a}(\Phi, f^{(n)}) = \mathrm{b}(\Phi, f^{(n-1)})$:

$$\begin{aligned} &\frac{1}{\Delta t}\left(\Big\langle \Phi, f^{(n)}\Big\rangle_{\Omega} - \Big\langle \Phi, f^{(n-1)}\Big\rangle_{\Omega}\right) \\ &\qquad = -\Big\langle \mathbf{v}\cdot\mathbf{n}\Phi, f^{(n)}\Big\rangle_{\Gamma} - \Delta t\Big\langle \mathbf{v}\cdot\nabla_{\mathbf{x}}\Phi, \mathbf{v}\cdot\nabla_{\mathbf{x}} f^{(n)}\Big\rangle_{\Omega} + \Big\langle \mathbf{v}\cdot\nabla_{\mathbf{x}}\Phi, f^{(n-1)}\Big\rangle_{\Omega} \end{aligned} \tag{61}$$

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

$$\left(\left\langle \Phi, f^{(n)} \right\rangle_{\Omega} - \left\langle \Phi, f^{(n-1)} \right\rangle_{\Omega}\right) = -\Delta t \left\langle \mathbf{v}\cdot\mathbf{n}\Phi, f^{(n)} \right\rangle_{\Gamma} \tag{62}$$

The right hand side of (62) vanishes whenever $\Phi \in V^{L,N}$ is chosen to be radially symmetric in the velocity coordinate. Let $\Phi = \Psi^{\cos}_{k,j=0}$:

$$\left\langle \Psi^{\cos}_{k,j=0}, \sum_{k,j,i_x} c^{(n)}_{k,j=0,i_x} \Psi_{k,j}\phi_{i_x}(\mathbf{x}) \right\rangle_{\Omega} = \left\langle \Psi^{\cos}_{k,j=0}, \sum_{k,j,i_x} c^{(n-1)}_{k,j=0,i_x} \Psi_{k,j}\phi_{i_x}(\mathbf{x}) \right\rangle_{\Omega} \tag{63}$$

$$\Rightarrow \sum_{i_x} c^{(n)}_{k,j=0,i_x} \int_D \phi_{i_x}(\mathbf{x})\,\mathrm{d}\mathbf{x} = \sum_{i_x} c^{(n-1)}_{k,j=0,i_x} \int_D \phi_{i_x}(\mathbf{x})\,\mathrm{d}\mathbf{x}, \qquad \forall k \tag{64}$$

Where we have used the L_2-orthogonality of the spectral basis in the last line. The proof follows because only basis functions with zero angular frequency make nonzero contributions to mass and energy. □

The discrete formulation does not conserve mass in the vicinity of diffusive reflective boundary conditions. This is because in general the velocity distribution function will have jumps across the line $\mathbf{v}\cdot\mathbf{n} \equiv 0$. Therefore mass is not conserved since the spectral basis cannot approximate discontinuous functions exactly.

5.2 Efficient evaluation of boundary conditions

Depending on the type of boundary condition and orientation of the normal vector, the term $\left\langle \mathbf{v}\cdot\mathbf{n}\Phi, f \right\rangle_{\Gamma}$ creates dense blocks of size $N \times N$ in the system matrix at locations corresponding to boundary faces. This is unfortunate, both with respect to memory consumption and the cost of the matrix $\times$ vector-product (MV-product). Since we use GMRES to solve the advection problem (55) we do not need to compute the matrix entries explicitly. Fortunately the MV-product of the boundary conditions can be computed in $\mathcal{O}(K^3)$, compared to $\mathcal{O}(K^4)$ for the naive approach, via a transformation to the nodal basis.

The evaluation in the nodal basis for a single test function Φ_{i_x,i_y} costs $\mathcal{O}(1)$ operations:

$$\left\langle \mathbf{v}\cdot\mathbf{n}\Phi_{i_x,i_y}, f^N \right\rangle_{\Gamma} = \sum_{q_1=0}^{K-1}\sum_{q_2=0}^{K-1} [x_{q_1}, x_{q_2}]\cdot\mathbf{n} \underbrace{\Phi_{i_x,i_y}(x_{q_1},x_{q_2})}_{=\delta_{i_x,i_{q_1}}/\sqrt{w_{q_1}},\delta_{i_y,i_{q_2}}/\sqrt{w_{q_2}}} f^N(x_{q_1}x_{q_2})w_{q_1}w_{q_2} \tag{65}$$

$$= (\mathbf{c}^N)_{q_1,q_2}\,[x_{q_1}, x_{q_2}]\cdot\mathbf{n}, \tag{66}$$

where x_i, w_i are the Hermite quadrature points and weights. The cost for evaluating (65) for all $i_x, i_y \in [0, K-1]^2$ is $\mathcal{O}(K^2)$.

The inclusion of the boundary condition is trivial when working in the nodal basis. The coefficients in the Polar basis can be rotated with linear complexity $\mathcal{O}(N) = \mathcal{O}(K^2)$ to the reference frame such that the normal vector points is aligned with the y-axis. As it was already noted in Section 3.4, the transformation from the Polar to the nodal basis can be done with cost $\mathcal{O}(K^3)$. Thus the total cost for the on-the-fly evaluation of the boundary condition is $\mathcal{O}(K^3)$.

Algorithm 2 Evaluate $\left\langle \mathbf{v}\cdot\mathbf{n}\phi_i, f\right\rangle_\Gamma, i=1,\dots,N.$

procedure EVALUATE BOUNDARY TERM($\mathbf{c}^{\mathrm{P}}$)
  $\mathbf{c}^{\mathrm{N}} \leftarrow \mathbf{T}_{\mathrm{P}\to\mathrm{N}}\,\mathbf{R}\,\mathbf{c}^{\mathrm{P}}$ ▷ rotate to reference frame, transform to Nodal basis: $\mathcal{O}(K^3)$
  $\tilde{\mathbf{c}}^{\mathrm{N}} \leftarrow$ apply_BC
  $\mathbf{q} \leftarrow \tilde{\mathbf{c}}\mathbf{v}_y\mathbf{w}$ ▷ Apply quad. $\mathcal{O}(K^2)$
  return $\mathbf{R}^{-1}\mathbf{T}_{\mathrm{N}\to\mathrm{P}}\mathbf{q}$ ▷ $\mathcal{O}(K^3)$
end procedure

5.3 Positivity

It has been observed that in the vicinity of singularities, for example near re-entrant corners, the distribution function might locally become negative. It has been observed, that in combination with a low Knudsen number this can lead the collision operator to numerically blow up the solution. 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 to the Polar-Laguerre basis. A naive implementation requires the evaluation of f 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 nodal basis.As already noted in section 3.4 the transformation Polar-Laguerre $\Leftrightarrow$ Hermite basis can be done with effort $\mathcal{O}(K^3)$. The transformation between the Hermite and nodal basis has again cost of $\mathcal{O}(K^3)$, this time because it can be performed separately along each coordinate axis and therefore the transformation matrices are of size $K\times K$ only.

Algorithm 3 Project to positive distribution values

procedure APPLY SCATTERING IN RE-CENTERED BASIS($\mathbf{c}^{\mathrm{P}}$)
  $\mathbf{c}^{\mathrm{H}} \leftarrow \mathbf{T}_{\mathrm{P}\to\mathrm{H}}\mathbf{c}^{\mathrm{P}}$ ▷ Transform to Hermite basis
  $\mathbf{c}^{\mathrm{N}} \leftarrow \mathbf{T}_{\mathrm{H}\to\mathrm{N}}$ ▷ Transform to Nodal basis
  for all $\left(\mathbf{c}^{\mathrm{N}}\right)_i < 0$ **do**
    $\left(\mathbf{c}^{\mathrm{N}}\right)_i \leftarrow 0$
  end for
  $\mathbf{c}^{\mathrm{H}} \leftarrow \mathbf{T}_{\mathrm{N}\to\mathrm{H}}\mathbf{c}^{\mathrm{N}}$ ▷ Transform to Hermite basis
  $\mathbf{c}^{\mathrm{P}} \leftarrow \mathbf{T}_{\mathrm{H}\to\mathrm{P}}\mathbf{c}^{\mathrm{H}}$ ▷ Transform to Polar-Laguerre basis
end procedure

The entries of the Hermite to Nodal transformation matrix $\mathbf{T}_{\mathrm{H}\to\mathrm{N}}\in\mathbb{R}^{K,K}$ are given by:

$$\left(\mathbf{T}_{\mathrm{H}\to\mathrm{N}}\right)_{i,j} = \int_{\mathbb{R}} h_j(x)e^{-\frac{x^2}{2}}\ell_i(x)e^{-\frac{x^2}{2}}\,\mathrm{d}x = \sum_{k=0}^{K} h_j(x_k)\ell_i(x_k)w_k = \sum_{k=0}^{K} h_j(x_k)\frac{\delta_{i,k}}{\sqrt{w_k}}w_k = h_j(x_i)\sqrt{w_i}, \tag{67}$$

where $x_i, w_i, i=0,\dots,K$ are the Gauss-Hermite quadrature nodes and weights.

Theorem 5.3. $\mathbf{T}_{\mathrm{H}\to\mathrm{N}}$ *is an orthonormal matrix.*

Proof.

$$(\mathbf{T}_{\mathrm{H}\to\mathrm{N}})^T\mathbf{T}_{\mathrm{H}\to\mathrm{N}} = \sum_{k=0}^{K}(h_i(x_k)\sqrt{w_k})\ (h_j(x_k)\sqrt{w_k}) = \sum_{k=0}^{K} h_i(x_k)h_j(x_k)w_k$$
$$= \int_{R^2} h_i(x)h_j(x)e^{-x^2}\,\mathrm{d}x = \delta_{i,j} \quad (68)$$

□

Thus we have that $(\mathbf{T}_{\mathrm{N}\to\mathrm{H}})^{-1} := \mathbf{T}^T_{\mathrm{H}\to\mathrm{N}}$.

6 Numerical Experiments

We have implemented all the techniques discussed in `C++`. For the finite element part we have used the `deal.II v8.3` [19] finite element library. Since the collision operator is independent of $\mathbf{x}$, it is trivial to parallelize via domain decomposition in the physical domain. The system matrix for the advection problem is assembled once and reused in every time-step, as preconditioner we use a block-diagonal incomplete LU-factorization. Often it is observed that the ILU-preconditioned GMRES solver converges in less than 5 iterations. For the solution of the advection part Trilinos `v12.2.1` [13] is used.

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 131 quadrature points were used. Thus it can be assumed that quadrature error is negligible.

6.1 Homogeneous case

The BKW solution [8] is the only non-stationary known analytical solution to the homogeneous Boltzmann equation

$$f(t,\mathbf{v}) = e^{-\frac{\|\mathbf{v}\|^2}{2s}}\frac{\|\mathbf{v}\|^2 - (2+\|v\|^2)s + 4s^2}{4\pi s^3}, \qquad t>0 \quad (69)$$

where $s = 1-\exp(-\frac{1}{8}(t+8\log 2))$. It is valid for Maxwellian molecules.

Taking the limit $t\to\infty$ of (69) shows that the equilibrium solution is represented by a single nonzero coefficient in the Polar-Laguerre basis:

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

$f(t,\mathbf{v})$ from (69) has temperature $T=1$, thus we call it to be a temperature-conforming solution for $T=1$.

Theorem 6.1. *Let $f(t,\mathbf{v})$ be a solution to (21) with a collision kernel of the form (11). Let $\alpha,\gamma>0$ be given, and define $\eta = \alpha/\gamma^{\lambda+2}$. Then*

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

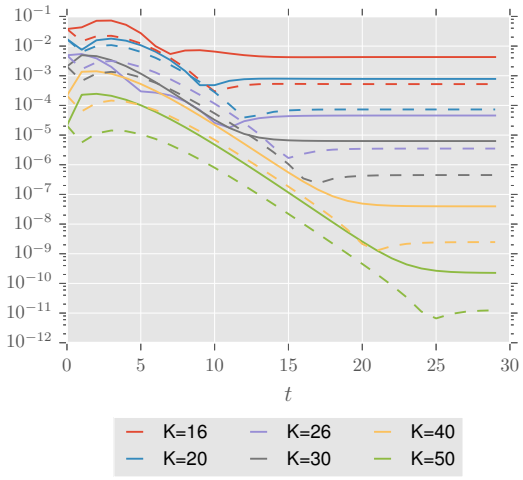
is also a solution to (21).

The proof can be found in [9, Theorem 3.2].

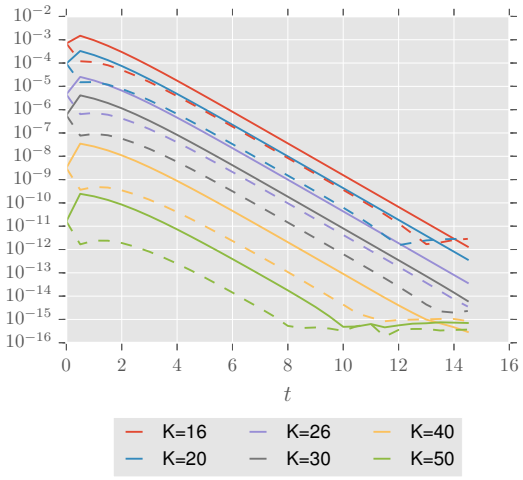
Making use of Theorem 6.1 and rescaling $f(t, \mathbf{v})$ accordingly we can construct a BKW solution conforming to a different temperature. Numerical results for the BKW solution are reported in Fig. 6. In order to demonstrate the approximation properties of the Polar-Laguerre basis the simulations were carried out for temperature-conforming 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 energy described in Section 3.1. We used RK4 with step size $\Delta t{=}0.001$ for the $T{=}1$ and $\Delta t{=}0.002$ for the $T{=}0.5$-conforming initial distribution. The equilibrium state was reached after $15k$ timesteps. Relative L_2-errors are reported in Fig. 6. As expected we observe the L_2-errors decay fastest wrt. to time for the $T{=}1$ conforming initial distribution centered at $\mathbf{v}_c{=}[0, 0]$, cf. Fig. 6b. For $t > 10$ and for suitably high polynomial degree K, the errors are of the size of the machine epsilon.

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.

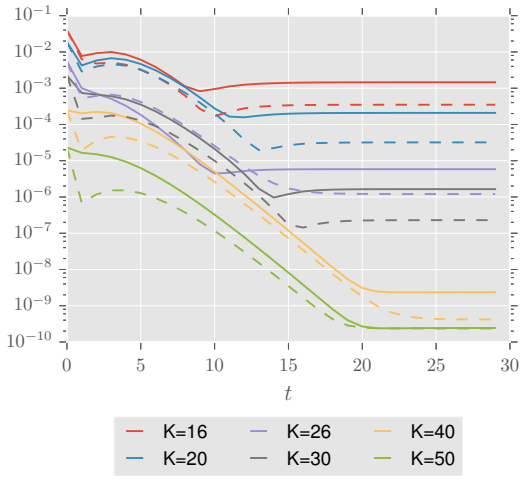
Remark 6.2. *In [9] initial conditions are always rescaled to be temperature-conforming which allows to obtain best possible approximation properties. It must be pointed out that in the inhomogeneous case this cannot be done when using continuous finite elements in the spatial domain this would lead to dense subblocks in the system matrix of size $N \times N$.*



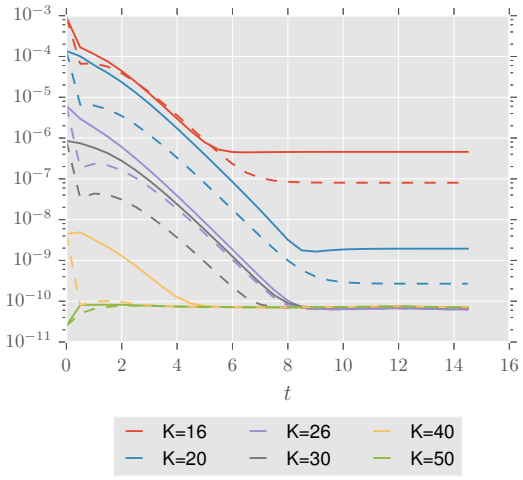
(a) T=0.5-conforming BKW solution.



(b) T=1-conforming BKW solution.



(c) T=0.5-conforming BKW solution centered at $\mathbf{v}_c = [1, 1]$.



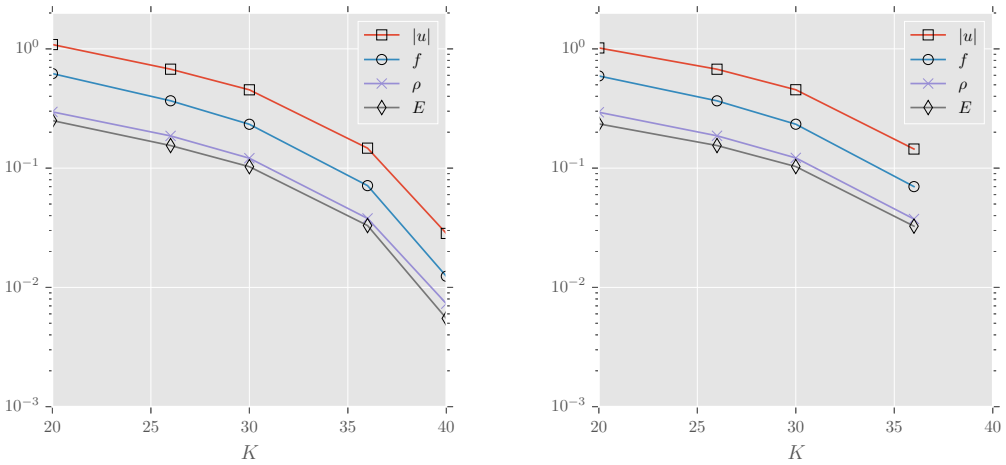
(d) T=1-conforming BKW solution centered at $\mathbf{v}_c = [1, 1]$.

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 famous Mach 3 wind tunnel experiment. The computational domain describes a wind tunnel with a forward facing step at position $x = 0.6$ weight height 0.2. The gas is initially at equilibrium with temperature T_0=1, $\mathbf{v}$=[3, 0], $\rho = 1.4$. At $x\equiv 0$ inflow boundary conditions with T=1, $\mathbf{v}$=[3, 0] are posed and outflow (zero inflow) boundary conditions at $x\equiv 3$, the remaining 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 $t \in [0, 1]$. AΔt=2.5×10^{-5}. The results qualitatively agree with real-world experiments carried out in a wind tunnel, cf. Woodward and Colella [20].

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=2.5×10^{-3} this may cause the solution to diverge when the scattering operator is applied. A possible remedy is to project to positive distribution function in $\mathbf{v}$, cf. discussion in Section 5.3. For the results shown here this projection step was not necessary.



(a) #phys. DoFs L=8945

(b) #phys. DoFs L=35 297

Figure 7: Rel. L_2-errors vs. polynomial degree K for two different physical meshes. The solution on the finer 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.

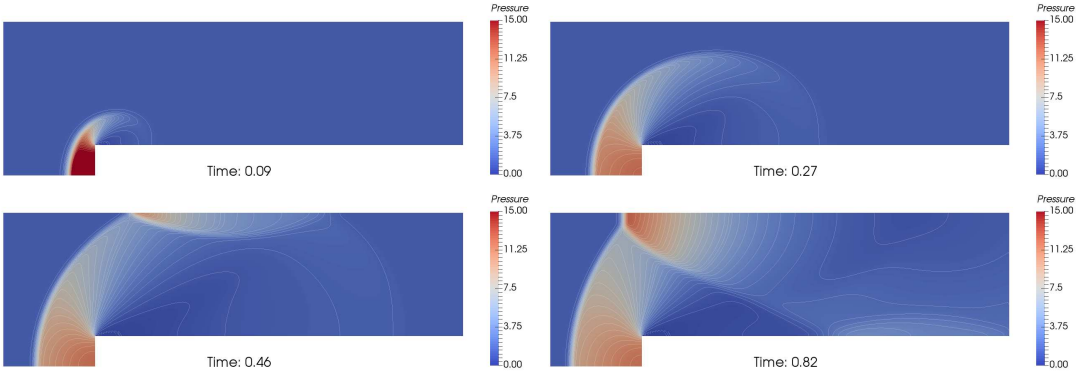


Figure 8: Mach 3 wind tunnel: Polynomial degree $K = 40$, $35k$ Vertices, Maxwellian molecules, $28.9M$ total DoFs. Coloring: pressure, contour lines: density. Computations were carried out on the Euler cluster of ETH Zurich (Xeon E5-2697 v2) with 360 cores.

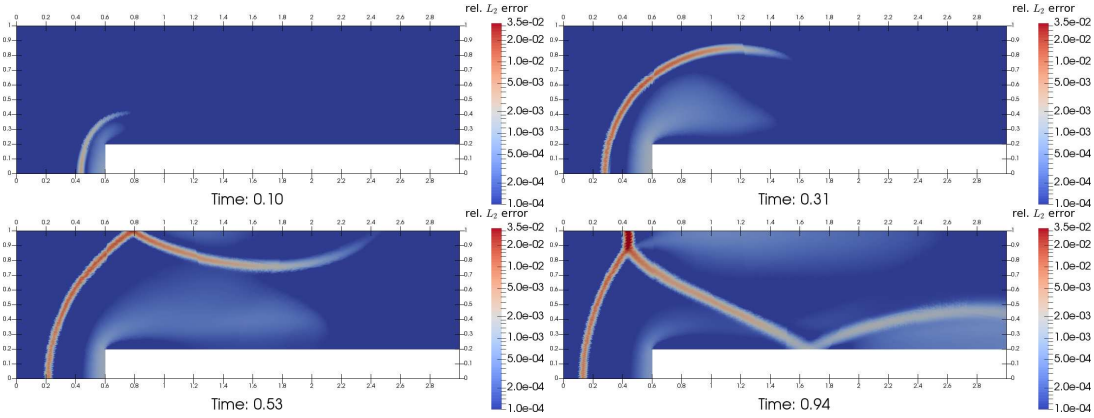
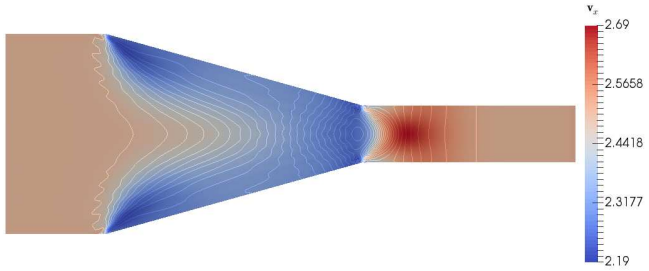


Figure 9: Mach 3 wind tunnel. Relative L_2-errors for K=36, $L = 35k$ on the physical grid measured against the reference solution ($K = 40$).

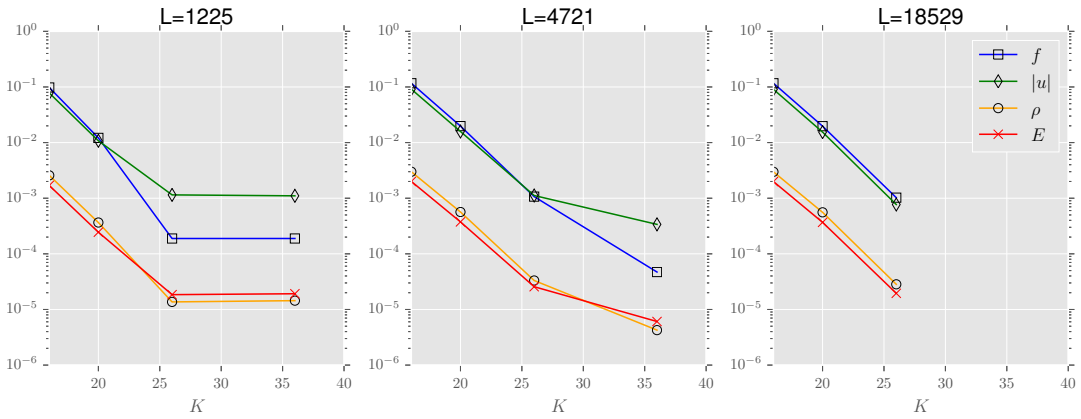
6.3 Nozzle flow

Inflow at the left boundary with $T = 1, \mathbf{v}_0 = [2.5, 0], \rho_0 = 1.4$. Zero inflow boundary condition at $x = 4$. The walls are specularly reflecting. The Knudsen number is $kn = 0.1$. Convergence plots for L_2-errors are reported in Fig. 10b, the reference solution was computed on a mesh with 18 529 vertices and for polynomial degree $K = 40$. We observe that for the lowest resolution in space, i.e. $L = 1225$, for $K > 26$ the error is dominated by the mesh size. Whereas for $L = 4721$ the errors mainly depend on K. We observe much smaller errors and faster convergence in the polynomial degree K compared to the Mach 3 wind tunnel experiment. This is attributed to the absence of shocks.

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



(a) Nozzle flow: L=18 529, K=36, N=666, velocity in x-direction. Pressure as contour lines.



(b) 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.

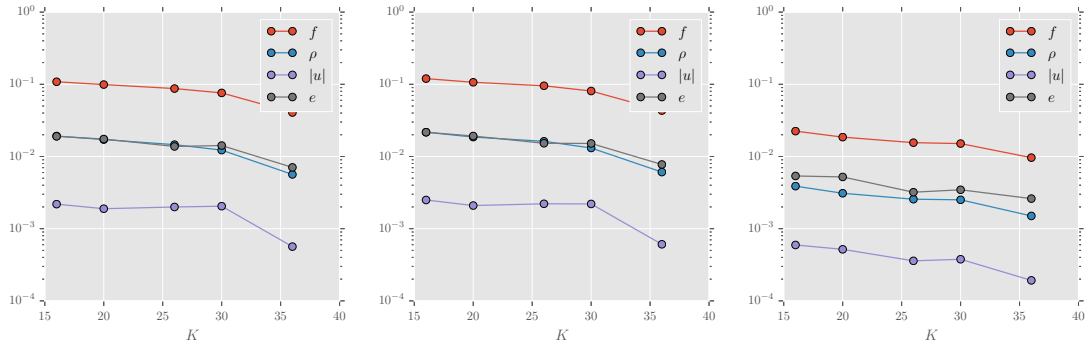
6.4 Shock tube

A gas is placed in a tube with length 1. Initially the gas is at equilibrium in the left and right half with densities ρ_l, ρ_r and temperatures T_l, T_r:

$$f_l(t{=}0, \mathbf{x}, \mathbf{v}) = \frac{\rho_l}{2\pi T_l} \exp\left(-\frac{\|\mathbf{v}\|^2}{2T_l}\right), \quad x < 0.5 \tag{70}$$

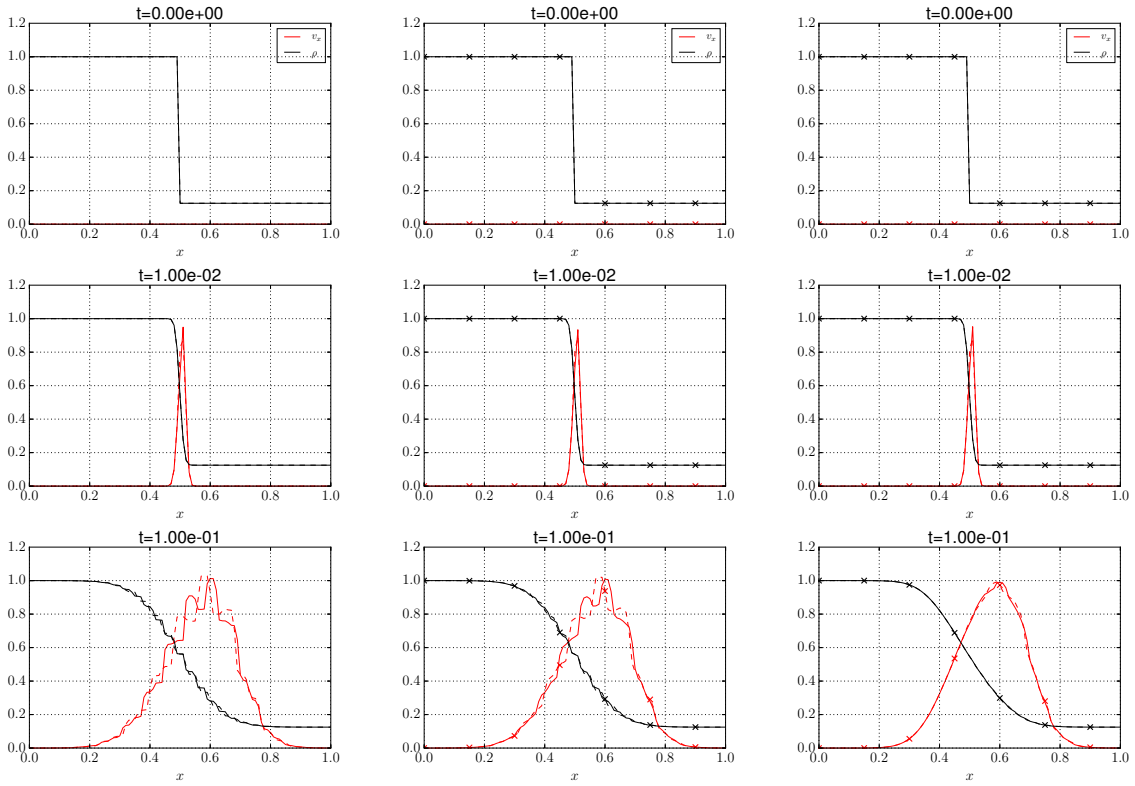
$$f_r(t{=}0, \mathbf{x}, \mathbf{v}) = \frac{\rho_r}{2\pi T_r} \exp\left(-\frac{\|\mathbf{v}\|^2}{2T_r}\right), \quad x \geq 0.5, \tag{71}$$

where $\rho_l{=}1, \rho_r{=}1$ and temperatures $T_l{=}1.25, T_r{=}1$. Specular reflective boundary conditions are used at the top and bottom wall and inflow boundary conditions with densities ρ_l, ρ_r and temperatures T_l, T_r at $x{=}0$, $x{=}1$. The calculations have been carried out on a structured grid with element size $h_x{=}1.48 \times 10^{-3}$ in x-direction and for $kn{=}0.01$, 0.1, 1 with polynomial degree $K{=}16$, 20, 26, 30, 36, 40. The calculation with $K{=}40$ has been used as reference to compute L_2-errors in the distribution function $f(\mathbf{v}, \mathbf{x})$ and for the macroscopic quantities $\rho, |\mathbf{u}|$ and E. L_2 errors are shown in Fig. 11, the errors for $kn{=}0.01$ are an order of magnitude smaller compared to the calculation with $kn{=}0.1$. This is because, for $kn{=}0.01$, the smoothing by the collision operator is stronger and therefore better approximation by the velocity basis is obtained. In Fig. 12 the density and momentum $\mathbf{u}_x$ are compared for $K{=}30$, 40 at different times along the path $x(t){=}t, t \in [0, 1]$.



(a) $kn = 1$ (b) $kn = 0.1$ (c) $kn = 0.01$

Figure 11: Shock tube: relative L_2 errors for varying polynomial degrees K against reference computation with $K{=}40$. Errors measured at $t{=}0.1$.

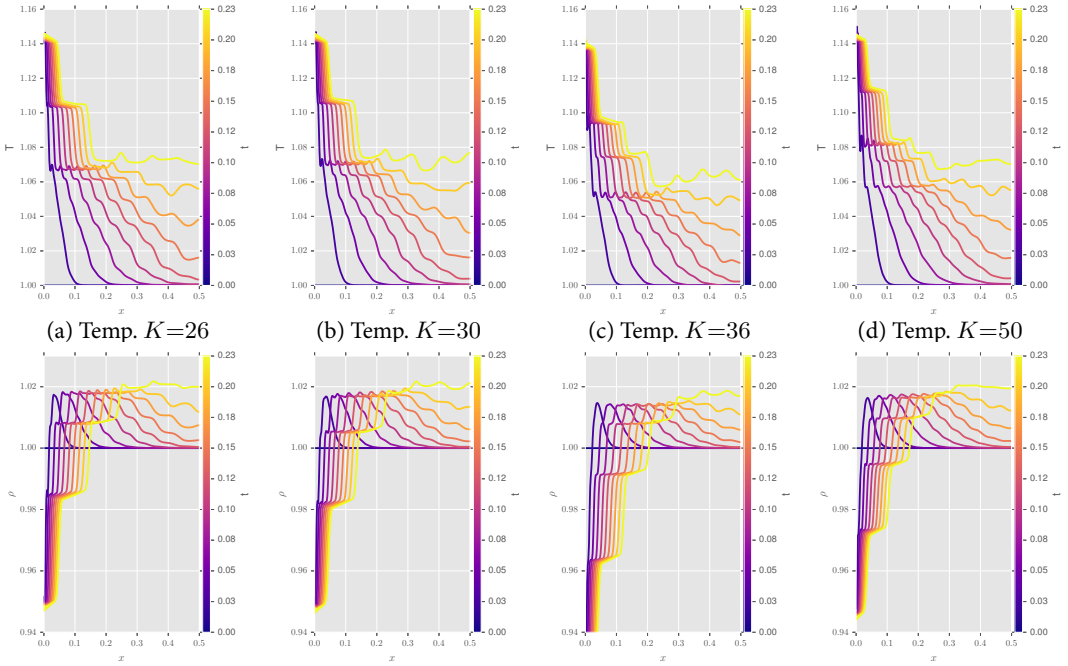


(a) $kn = 1$ (b) $kn = 0.1$ (c) $kn = 0.01$

Figure 12: 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 periodic 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 Section 5, mass is not conserved exactly for diffusive boundary conditions. However, 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. (13) and (14). The results agree qualitatively, but the fluctuations are too large to compute convergence rates. The inaccuracy originates from the temperature shock present at $t{=}0$ 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 cannot be approximated well in the Polar-Laguerre basis.



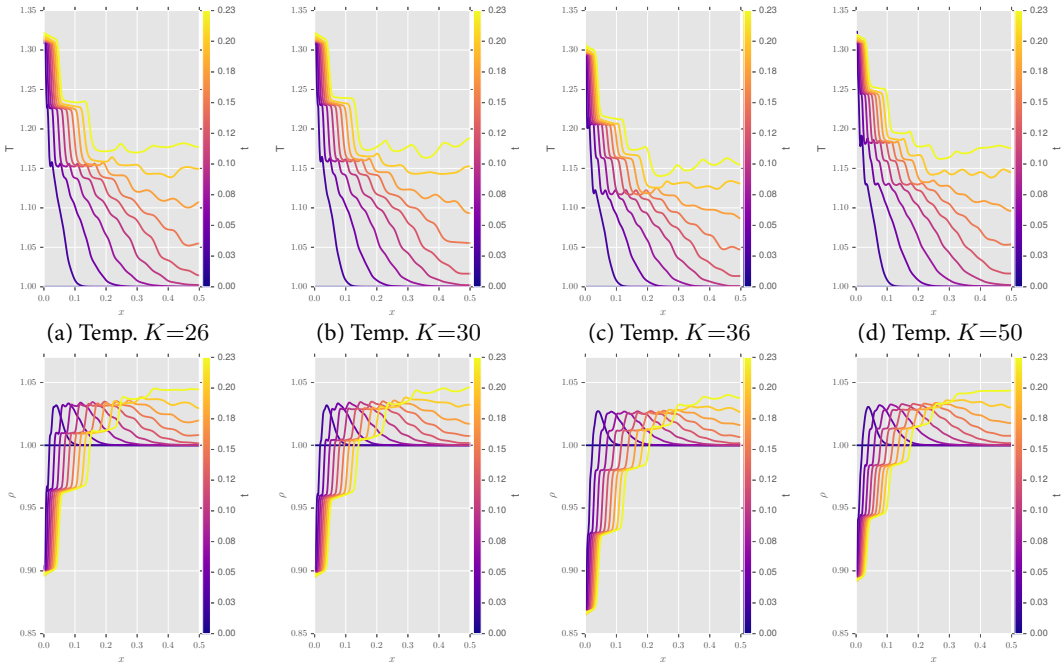
(a) Temp. $K{=}26$ (b) Temp. $K{=}30$ (c) Temp. $K{=}36$ (d) Temp. $K{=}50$

(e) Mass, $K{=}26$ (f) Mass, $K{=}30$ (g) Mass, $K{=}36$ (h) Mass, $K{=}50$

Figure 13: 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.

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: $h = 512^{-1}$.



(a) Temp. K=26 (b) Temp. K=30 (c) Temp. K=36 (d) Temp. K=50

(e) Mass, K=26 (f) Mass, K=30 (g) Mass, K=36 (h) Mass, K=50

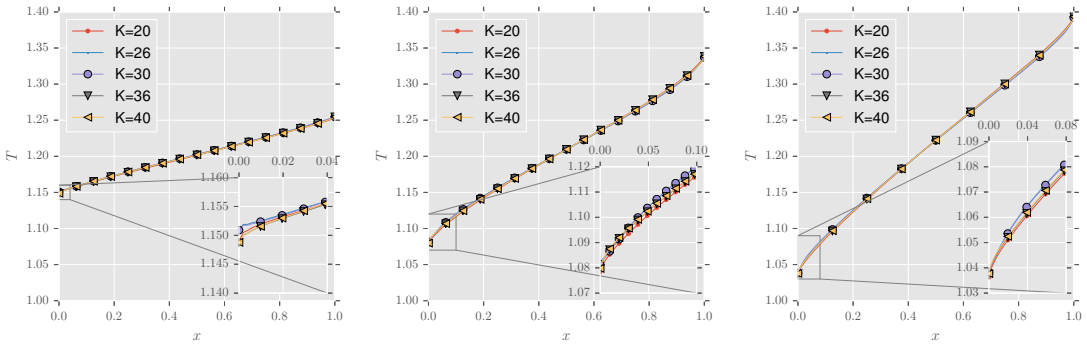
Figure 14: Sudden change in wall tempereature T_w=1.7: Evolution of the temperature $T(x,t)$ and mass ρ for $x \in [0,\ 0.5], t \in [0,\ 0.23]$. The time evolution is encoded in the color map.

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

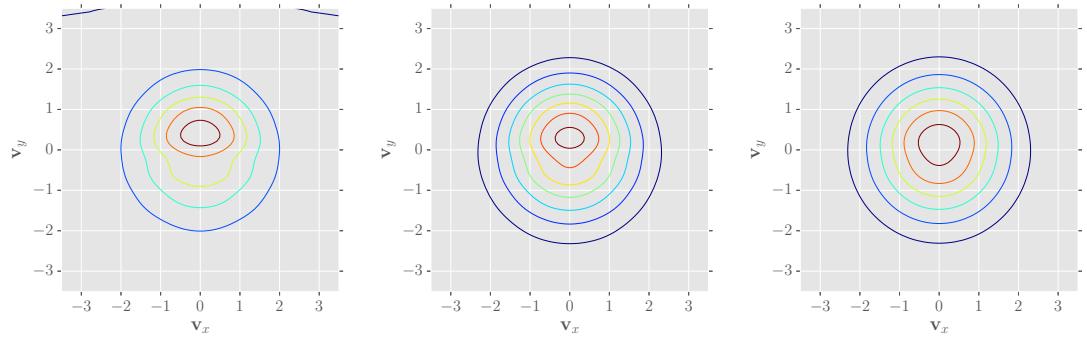
$$f(t=0,y,\mathbf{v}) = \frac{1}{2\pi T(y)} e^{-\frac{\|\mathbf{v}\|^2}{2T(y)}}$$
$$T(y) = 1 + 1.44y.$$

The simulations were carried out for Knudsen numbers kn=0.025, 0.1, 1 until a stationary state was reached with timestep $\Delta t = 0.001$. We observe good agreement in the temperature profiles for K=20, . . . , 40.



(a) $kn = 1$ (b) $kn = 0.1$ (c) $kn = 0.025$

Figure 15: Temperature profiles for the stationary states at $t = 6,\ 25,\ 75$ for $kn = 1,\ 0.1,\ 0.025$.



(a) $kn = 1$ (b) $kn = 0.1$ (c) $kn = 0.025$

Figure 16: 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 demonstrated 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 $C(\cos\theta)\,\|v - \mathbf{v}_\star\|$ can be tackled by our scheme. We have observed that the conservative Galerkin discretization with Lagrange multipliers usually yields smaller errors with respect to the L_2-norm than an approach relying on modified test functions. 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 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] Milton Abramowitz and Irene Stegun. *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*. Dover, June 1972. 358 pp.

[2] G. A. Bird, ed. *Molecular gas dynamics and the direct simulation of gas flows*. Oxford engineering science series 42. Oxford: Clarendon, 1994. 458 pp.

[3] 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.

[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] C. Cercignani. "Chapter 1 - The Boltzmann Equation and Fluid Dynamics". In: *Handbook of Mathematical Fluid Dynamics*. Ed. by S. Friedlander {and} D. Serre. Vol. 1. North-Holland, 2002, pp. 1–69.

[6] Carlo Cercignani. *Rarefied gas dynamics : from basic concepts to actual calculations*. Cambridge texts in applied mathematics. Cambridge: Cambridge University Press, 2000. 320 pp.

[7] 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.

[8] Matthieu H. Ernst. "Exact solutions of the nonlinear Boltzmann equation". In: *J Stat Phys* 34.5 (Mar. 1, 1984), pp. 1001–1017. DOI: 10.1007/BF01009454.

[9] 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.

[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] 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.

[12] 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.

[13] Michael A. Heroux et al. "An overview of the Trilinos project". In: *ACM Trans. Math. Softw.* 31.3 (2005), pp. 397–423. DOI: http://doi.acm.org/10.1145/1089014.1089021.

[14] 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.

[15] Kenichi Nanbu. "Direct Simulation Scheme Derived from the Boltzmann Equation. I. Monocomponent Gases". In: *J. Phys. Soc. Jpn.* 49.5 (Nov. 15, 1980), pp. 2042–2049. DOI: 10.1143/JPSJ.49.2042.

[16] 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.

[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] 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.

[19] W. Bangerth et al. *The deal.II Library, Version 8.3*.

[20] Paul Woodward and Phillip Colella. "The numerical simulation of two-dimensional fluid flow with strong shocks". In: *Journal of Computational Physics* 54.1 (Apr. 1, 1984), pp. 115–173. DOI: 10.1016/0021-9991(84)90142-6.

[21] 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.

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

$$\langle Q(\phi_{\tau_1,l_1}, \phi_{\tau_2,l_2}), \phi_{\tau,l} \rangle_{L^2(\mathbb{R}^2)} \tag{72}$$

with trial functions $\phi_{\tau_1,l_1},\ \phi_{\tau_2,l_2} \in V_{\mathcal{V}}^N$ and a test function $\phi_{\tau,l} \in \hat{V}_{\mathcal{V}}^N$.

The rotation invariance, cf. Theorem. 1.2, 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 $\phi_{\cdot,\cdot}$:

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

where l_1 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:

$$\begin{aligned} Q(\rho_\omega\phi_{c,l_1}, \rho_\omega\phi_{c,l_2}) &= \rho_\omega Q(\phi_{c,l_1}, \phi_{c,l_2}) \\ Q(\rho_\omega\phi_{c,l_1}, \rho_\omega\phi_{s,l_2}) &= \rho_\omega Q(\phi_{c,l_1}, \phi_{s,l_2}) \\ Q(\rho_\omega\phi_{s,l_1}, \rho_\omega\phi_{c,l_2}) &= \rho_\omega Q(\phi_{s,l_1}, \phi_{c,l_2}) \\ Q(\rho_\omega\phi_{s,l_1}, \rho_\omega\phi_{s,l_2}) &= \rho_\omega Q(\phi_{s,l_1}, \phi_{s,l_2}) \end{aligned} \tag{73}$$

Using the trigonometric identities

$$\begin{aligned} \rho_\omega \sin(l\varphi) &= \cos(l\varphi)\sin(l\omega) + \sin(l\varphi)\cos(l\omega) \\ \rho_\omega \cos(l\varphi) &= \cos(l\varphi)\cos(l\omega) - \sin(l\varphi)\sin(l\omega) \end{aligned} \tag{74}$$

and the bilinearity of Q, (73) is transformed into a 4×4 system of linear equations in the unknowns $Q(\phi_{c,l_1}, \phi_{c,l_2})$, $Q(\phi_{c,l_1}, \phi_{s,l_2})$, $Q(\phi_{s,l_1}, \phi_{c,l_2})$, $Q(\phi_{s,l_1}, \phi_{s,l_2})$:

$$\mathbf{A}(\omega)\mathbf{q}(\varphi) = \mathbf{q}_0(\omega), \tag{75}$$

where

$$\mathbf{A}(\omega) := \begin{pmatrix} \cos(l_2w)\cos(l_1w) & -\cos(l_1w)\sin(l_2w) & -\cos(l_2w)\sin(l_1w) & \sin(l_2w)\sin(l_1w) \\ \cos(l_1w)\sin(l_2w) & \cos(l_2w)\cos(l_1w) & -\sin(l_2w)\sin(l_1w) & -\cos(l_2w)\sin(l_1w) \\ \cos(l_2w)\sin(l_1w) & -\sin(l_2w)\sin(l_1w) & \cos(l_2w)\cos(l_1w) & -\cos(l_1w)\sin(l_2w) \\ \sin(l_2w)\sin(l_1w) & \cos(l_2w)\sin(l_1w) & \cos(l_1w)\sin(l_2w) & \cos(l_2w)\cos(l_1w) \end{pmatrix}$$

$$\mathbf{q}(\varphi) := \begin{pmatrix} Q(\phi_{c,l_1}, \phi_{c,l_2})(\varphi, r) \\ Q(\phi_{c,l_1}, \phi_{s,l_2})(\varphi, r) \\ Q(\phi_{s,l_1}, \phi_{c,l_2})(\varphi, r) \\ Q(\phi_{s,l_1}, \phi_{s,l_2})(\varphi, r) \end{pmatrix} \tag{76}$$

and

$$\mathbf{q}_0(\omega) := \begin{pmatrix} Q(\phi_{c,l_1}, \phi_{c,l_2})(\varphi+\omega, r) \\ Q(\phi_{c,l_1}, \phi_{s,l_2})(\varphi+\omega, r) \\ Q(\phi_{s,l_1}, \phi_{c,l_2})(\varphi+\omega, r) \\ Q(\phi_{s,l_1}, \phi_{s,l_2})(\varphi+\omega, r) \end{pmatrix}. \tag{77}$$

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:

$$\mathbf{A}(-\varphi)^{-1} = \begin{pmatrix} \cos(l_2\varphi)\cos(l_1\varphi) & -\cos(l_1\varphi)\sin(l_2\varphi) & -\cos(l_2\varphi)\sin(l_1\varphi) & \sin(l_2\varphi)\sin(l_1\varphi) \\ \cos(l_1\varphi)\sin(l_2\varphi) & \cos(l_2\varphi)\cos(l_1\varphi) & -\sin(l_2\varphi)\sin(l_1\varphi) & -\cos(l_2\varphi)\sin(l_1\varphi) \\ \cos(l_2\varphi)\sin(l_1\varphi) & -\sin(l_2\varphi)\sin(l_1\varphi) & \cos(l_2\varphi)\cos(l_1\varphi) & -\cos(l_1\varphi)\sin(l_2\varphi) \\ \sin(l_2\varphi)\sin(l_1\varphi) & \cos(l_2\varphi)\sin(l_1\varphi) & \cos(l_1\varphi)\sin(l_2\varphi) & \cos(l_2\varphi)\cos(l_1\varphi) \end{pmatrix}, \tag{78}$$

thus we have simplified (73) to

$$\begin{aligned} Q(\phi_{c,l_1}, \phi_{c,l_2}) &= \cos(l_2\varphi)\cos(l_1\varphi)\ Q(\phi_{c,l_1}, \phi_{c,l_2})(0, r) \\ Q(\phi_{c,l_1}, \phi_{s,l_2}) &= \cos(l_1\varphi)\sin(l_2\varphi)\ Q(\phi_{c,l_1}, \phi_{c,l_2})(0, r) \\ Q(\phi_{s,l_1}, \phi_{c,l_2}) &= \cos(l_2\varphi)\sin(l_1\varphi)\ Q(\phi_{c,l_1}, \phi_{c,l_2})(0, r) \\ Q(\phi_{s,l_1}, \phi_{s,l_2}) &= \sin(l_2\varphi)\sin(l_1\varphi)\ Q(\phi_{c,l_1}, \phi_{c,l_2})(0, r) \end{aligned} \tag{79}$$

We multiply (79) with the test function $\phi_{\tau,l}$ and integrate over φ.

$$\int_0^{2\pi} \begin{pmatrix} Q(\phi_{c,l_1}, \phi_{c,l_2}) \\ Q(\phi_{c,l_1}, \phi_{s,l_2}) \\ Q(\phi_{s,l_1}, \phi_{c,l_2}) \\ Q(\phi_{s,l_1}, \phi_{s,l_2}) \end{pmatrix} \phi_{\tau,l} \, \mathrm{d}\varphi = Q(\phi_{c,l_1}, \phi_{c,l_2})(0,r) \int_0^{2\pi} \begin{pmatrix} \cos(l_2\varphi)\cos(l_1\varphi) \\ \cos(l_1\varphi)\sin(l_2\varphi) \\ \cos(l_2\varphi)\sin(l_1\varphi) \\ \sin(l_2\varphi)\sin(l_1\varphi) \end{pmatrix} \phi_{\tau,l} \, \mathrm{d}\varphi \quad (80)$$

From (80) 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 $\tau = \cos$.

$$\langle Q(\phi_{\tau_1,l_1}, \phi_{\tau_2,l_2})(\mathbf{v})), \phi_{\tau,l}(\mathbf{v}) \rangle_{L^2(\mathbb{R}^2)} = \begin{cases} \neq 0 & ((l_1 + l_2) = l \vee |l_1 - l_2| = l) \bigwedge \tau_1 \neq \tau_2 \\ 0 & \text{otherwise} \end{cases} \quad (81)$$

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

$$\langle Q(\phi_{\tau_1,l_1}, \phi_{\tau_2,l_2})(\mathbf{v})), \phi_{\tau,l}(\mathbf{v}) \rangle_{L^2(\mathbb{R}^2)} = \begin{cases} \neq 0 & ((l_1 + l_2) = l \vee |l_1 - l_2| = l) \bigwedge \tau_1 = \tau_2 \\ 0 & \text{otherwise} \end{cases} \quad (82)$$

Recent Research Reports

Nr.	Authors/Title

2015-28 P. Chen and Ch. Schwab
Model Order Reduction Methods in Computational Uncertainty Quantification

2015-29 G. S. Alberti and S. Dahlke and F. De Mari and E. De Vito and S. Vigogna
Continuous and discrete frames generated by the evolution flow of the Schrödinger equation

2015-30 P. Grohs and G. Kutyniok and J. Ma and P. Petersen
Anisotropic Multiscale Systems on Bounded Domains

2015-31 R. Hiptmair and L. Scarabosio and C. Schillings and Ch. Schwab
Large deformation shape uncertainty quantification in acoustic scattering

2015-32 H. Ammari and P. Millien and M. Ruiz and H. Zhang
Mathematical analysis of plasmonic nanoparticles: the scalar

case

2015-33 H. Ammari and J. Garnier and L. Giovangigli and W. Jing and J.K. Seo
Spectroscopic imaging of a dilute cell suspension

2015-34 H. Ammari and M. Ruiz and S. Yu and H. Zhang
Mathematical analysis of plasmonic resonances for nanoparticles: the full Maxwell equations

2015-35 H. Ammari and J.K. Seo and T. Zhang
Mathematical framework for multi-frequency identification of thin insulating and small conductive inhomogeneities

2015-36 H. Ammari and Y.T. Chow and J. Zou
Phased and phaseless domain reconstruction in inverse scattering problem via scattering coefficients