Derivation and analysis of relaxation operators in kinetic theory

Stéphane Brull[1*], Vincent Pavan[2] and Jacques Schneider[3]

[1*]Institut de Mathématiques de Bordeaux, Bordeaux INP, Univ. Bordeaux, CNRS, 351, cours de la Libération, Talence, F-33400, France.
[2]IUSTI UMR 7343, Aix-Marseille Université, 5 rue Enrico Fermi,, Marseille, F-13013 Marseille, France.
[3]IMATH , Université de Toulon, Campus La Garde CS, Toulon, 83041, France.

*Corresponding author(s). E-mail(s): stephane.brull@math.u-bordeaux.fr;
Contributing authors: vincent.pavan@univ-amu.fr; jacques.schneider@univ-tln.fr;

Abstract

We aim to present a theory for the derivation of relaxation operators in kinetic theory. The construction is based on an approximation of the inverse Boltzmann linearized operator, on relaxation equations on the moments of the distribution function and finally on a variational problem to be solved. The theory comprises a characterization of the set of moments of non negative integrable functions, a study of those linear application whose range lies in this set and a generalization of the functional to be minimized under moment constraints. In particular we clarify but also modify some steps in the proof of Junk's theorem on the characterization of moments of non negative functions [30]. We also reestablish Csiszar's theorem [20] by different means on a class of functionals leading to well-posed variational problems. The present theory encompasses the derivation of known models and that of new models.

Keywords: relaxation operator, truncated moment problem, positive polynomials, ϕ-divergence

1 Introduction

Kinetic models are used to simulate rarefied gases in the context of atmospheric reentry, CVI deposition, micro-channels and other processes. The question arises whether the very fine description of the gas that is given by the original Boltzmann equation is required or not for such simulations. Depending on the Knudsen number, the collision operator may be replaced by simpler models that are easier to handle such as relaxation operators which seminal model is the BGK one [7]. Another reason for considering such operators is the difficulty to obtain the physical parameters that rule the interaction between molecules beyond the case of mono-atomic molecules. Contrarily, relaxation operators are often scaled by measurements obtained at the macroscopic level such as the diffusion coefficients. The idea that consists in taking some information available at this level is also used for the Boltzmann equation with continuous internal energy states - see e.g [10] - but this will not be our concern here. So, there exists plenty relaxation operators that are used in different context but no unified approach while many feature the same patterns. A linear or almost linear behavior with respect to the moments of the distribution function, most properties that are satisfied by the original equation such as positivity of solution or the existence of an entropy, and finally a correct hydrodynamic limit up to the Navier- Stokes equation.

In the present article, we aim to develop a theoretical approach that applies to existing models such as the BGK [7], ESBGK [29] or Shakhov model [41] and serves as a ground from which ongoing models can be constructed. The theory is presented only in the case of mono-atomic molecules for it requires much technical matters. The paper goes along the steps of what we name the method of moments relaxation which is a generalization of the work presented in [12, 13] together with some applications for modeling multicomponent gases (see for example [11, 14, 15, 40]). For short, the construction is based on relaxation equations that are relations between moments of the operator and moments of the distribution function with respect to a vector of weight functions $\boldsymbol{m}$. $\boldsymbol{m}$ together with relaxation parameters are suitably chosen in order to obtain for example the right transport coefficients in the hydrodynamic limit. The relaxation equations are restated in term of linear relations between the moments of the probability density function f and those of the target function to be found. So the question arises whether those relations have a range into the set of realizable moments $\mathcal{R}^+_{\boldsymbol{m}}$, that is the set of vectors which are moments of nonnegative functions. This implies in a first time to characterize $\mathcal{R}^+_{\boldsymbol{m}}$ and in a second time to specify which are the admissible relaxation equations. In one dimension and when the domain of velocity is $\mathbb{R}$ and $\boldsymbol{m} = \{1, \dots, v^{2N}\}^T$, the characterization of $\mathcal{R}^+_{\boldsymbol{m}}$ is known as the Hamburger moment problem. In this case, a moment $\boldsymbol{\rho}$ is realizable w.r.t. $\boldsymbol{m}$ iff the moment matrix $(\rho_{i+j})_{i=0,\dots,N;j=0,\dots,N}$ is symmetric positive definite (SPD) (see for example [3]). In higher dimension which is the case of the usual velocity domain, the problem has given rise to many research (see for example [21, 23, 24, 27, 28, 32] and also the survey in [25]). Most results deal with moments of Borel measures and eventually of moments of atomic measures. They are unfortunately abstract and do not lead to tractable conditions such as in dimension 1. Typically, those results are expressed in terms of : "a moment is realizable if and only if there exists an extension of it - i.e with respect to a basis of higher degree - satisfying some property". Such a remark has

lead Pichard [36, 37] to propose specific moment closures in kinetic theory. However no such an extension is required when one deals with moments of integrable functions as proved by Junk [30]. In this case realizable moments are related to nonnegative polynomials in the space spanned by the components of $\boldsymbol{m}$, and not those of an higher order. Characterizing realizable moments then amounts to characterizing nonnegative polynomials. This problem is known as the 17^{th} Hilbert problem (see [6] for a survey).

The next problem is the way to define G once its moments are known. For most relaxation operators, it is usually done by minimizing the natural entropy under moment constraints. Unfortunately, this functional is not suited neither to the simple case of Grad thirteen moments nor to moments of higher degree. Junk was the first to raise and characterize the problem that one may face with Levermore's closure [33]. He was then followed by different authors giving their own insight on the problem [26, 35, 39]. One way to bypass the problem related to the natural entropy is to approximate it in a proper way. To do so one has to consider optimization problems with linear constraints such as in the article of Borwein and Lewis [8]. Another article by Csiszar [18] presents a large set of convex functionals - named ϕ-divergence - to be minimized under moments constraints. He would establish later on [20] the conditions for those minimization problems to have a solution. Abdelmalik and Van Brummelen have then proposed to use Csiszar's approach to set well-posed closure relations for moment systems [2]. The generalization to different approximations of $\int f \ln(f)$ by ϕ-divergence can finally be found in [1].

This article is organized as follows. We set in section 2 the properties that must be satisfied by a well-posed relaxation operator. The method of moments relaxation is displayed in section 3 together with the mathematical problems that one must address. Section 4 is devoted to the characterization of realizable moments and of admissible linear relaxations. Firstly, we revisit and try to clarify Junk's theorem [30] on the characterization of $\mathcal{R}_{\boldsymbol{m}}^{+}$. Secondly, we display the known results on the 17^{th} Hilbert problems, how they can be used but also what is the limitation of the characterization of $\mathcal{R}_{\boldsymbol{m}}^{+}$ by using nonnegative polynomials. Thirdly we consider the case of Grad thirteen moments and fully describe admissible relaxations whose range are in $\mathcal{R}_{\boldsymbol{m}}^{+}$. We close this section by a studying a class of linear operators that let $\mathcal{R}_{\boldsymbol{m}}^{+}$ invariant and that are related to Galilean invariance of the models to be constructed. Section 5 deals with the optimization problem. The somehow formal derivation of the solution given in [20] leads us to reestablish his existence theorem. Another reason to do this is that his proof relies on probabilistic tools such as convergence in measure but also on duality in Orlicz spaces whereas convex analysis is more appropriate to the problem we are dealing with. In section 6, we show that the model which is constructed just basing on relaxations on the Grad thirteen moments is well-posed in the sense of section 2. We also address the general case and point out some problems to be solved. We then display some known models that can be derived in this framework just by using different functional in the variational problem [7, 29, 41]. We compare the present approach to Levermore's sum of relaxation operators [33]. Velocity dependent relaxation frequency models are not contained within the present approach even if they share a lot of common points with the present study [9, 34, 42, 43]. Finally, all proofs are presented separately in section 6.

2 Well-posed BGK models and method of relaxation

2.1 Well-posed BGK models

Consider the kinetic equation

$$\partial_t f + \mathbf{v} \cdot \nabla_{\mathbf{x}} f = Q(f, f), \tag{1}$$

where $Q(f, f)$ is the classical Boltzmann operator [16, 17]. We are looking for a family of relaxation operators

$$K(f) = \nu(G - f)$$

that we may substitute to $Q(f, f)$ in some physical regime. We require $K(f)$ to satisfy the properties:

1. Conservation laws. There must hold

$$\forall f, \int K(f)\phi(\mathbf{v})\, d\mathbf{v} = 0 \Leftrightarrow \phi \in \mathbb{K},$$

with

$$\mathbb{K} = \mathtt{span}\left\{\mathbf{1}, \mathbf{v}, \mathbf{v}^2\right\} \tag{2}$$

2. There exists an "entropy" $\mathcal{H}(f) = \int \eta(f) d\mathbf{v}$ satisfying

$$\int K(f)\partial_f \eta(f)\, d\mathbf{v} \leq 0. \tag{3}$$

$\mathcal{H}(f)$ with $\eta(x) = x \ln x$ is a Lyapunov functional for the non homogeneous equation. Unfortunately, the moment approach of relaxation operators is not suited to the Boltzmann entropy in most cases. The weakened property that is required here enlarges the choice in η and so may lead to well-posed variational problems which is part of the construction of $K(f)$. Also this may give some stability if the variations of f due to the transport of particles are smaller than those due to the relaxation operator. Property (3) must then be completed with the usual characterization of local equilibrium.

$$K(f) = 0 \Leftrightarrow \int \partial_f \eta(f) K(f) = 0 \Leftrightarrow \partial_f \eta(f) \in \mathbb{K}, \tag{4}$$

with

$$K(f) = 0 \Leftrightarrow \mathcal{M}(\mathbf{v}) = \frac{n}{(2\pi k_B T/m)^{3/2}} \exp\left(-\frac{m(\mathbf{v} - \mathbf{u})^2}{2k_B T}\right) \tag{5}$$

where $n, \mathbf{u}, T$ are the density, velocity and temperature. We refer (3),(4) and (5) as the extended H-theorem or H-theorem in short.

3. Preservation of positivity. Starting from nonnegative initial data $f(0, \mathbf{x}, \mathbf{v}) \geq 0$ the solution must remain nonnegative. Preservation of non negativity is ensured by the Boltzmann operator but also by $K(f)$ if $G(f) \geq 0$. Models such as the Shakhov model [41] which write $G[f] = \mathcal{M}[f] P[f]$ where $P[f]$ is just a polynomial function do not satisfy this property.
4. Galilean invariance. For any translation: $\tau \mathbf{v} = \mathbf{v} - \mathbf{u}$ and rotation $\tau \mathbf{v} = \Theta \mathbf{v}$ in the velocity space there holds
$$K([\tau f]) = [\tau K(f)]$$
where by definition
$$\forall \mathbf{v}, \quad [\tau f](\mathbf{v}) := f(\tau \mathbf{v}).$$
5. Hydrodynamic limit. The relaxation operator must be defined so as to obtain the right transport coefficients in the hydrodynamic limit. This essentially relies on the shape of the linearized operator $\mathcal{L}$ defined with
$$\mathcal{L}(g) := \lim_{\epsilon \mapsto 0^+} \frac{1}{\epsilon \mathcal{M}} K(\mathcal{M}(1 + \epsilon g)), \tag{6}$$
as well as its pseudo-inverse. In order to perform the Chapman-Enskog expansion, there must hold
   (a) $\ker \mathcal{L} = \mathbb{K}$
   (b) $\mathcal{L}$ id Fredholm, so invertible on the orthogonal of $\ker \mathcal{L}$
   (c) It is symmetrical negative on $(\ker \mathcal{L})^{\perp}$
   (d) The viscosity and thermal conductivity computed from $\mathcal{L}^{-1}$ must be the same as the ones derived from the Boltzmann equation.

2.2 Method of moments relaxation

Let us nlow consider the kinetic equation
$$\partial_t f + \mathbf{v} \cdot \nabla_{\mathbf{x}} f = K(f) \tag{7}$$
with
$$K(f) = \nu(G - f).$$
The question is : how to define ν and G in such a way that the solution equation (7) behaves as that of the Boltzmann equation (1). The range of validity in this comparison is understood in the following way. $\mathcal{M}$ being a local equilibrium function, $f = \mathcal{M} + \mathcal{M}g$ and the Boltzmann operator reads
$$Q(f, f) = \mathcal{M}\mathcal{L}_B(g) + Q(\mathcal{M}g, \mathcal{M}g),$$
where $\mathcal{L}_B$ is the linearized Boltzmann operator. So we assume that $Q(Mg, Mg)$ is negligible compared to $\mathcal{M}\mathcal{L}_B(g)$ and we are looking in a first time for a well-posed relaxation operator $K(f)$ approaching $\mathcal{M}\mathcal{L}_B(g)$. This writes in the weak form
$$\int_{\mathbb{R}^3} K(f)\phi d\mathbf{v} \approx \int_{\mathbb{R}^3} \mathcal{M}\mathcal{L}_B(g)\phi \, d\mathbf{v} = \int_{\mathbb{R}^3} \mathcal{M}g\mathcal{L}_B(\phi) \, d\mathbf{v}$$

$$= \int_{\mathbb{R}^3} (\mathcal{M} + \mathcal{M}g)\mathcal{L}_B(\phi)\, d\mathbf{v} = \int_{\mathbb{R}^3} f\mathcal{L}_B(\phi)\, d\mathbf{v}.$$

This approximation is of course impossible if one considers all test functions ϕ in $L^2(\mathcal{M})$ since this would imply $K(f) = \mathcal{M}\mathcal{L}_B(g)$. So we must restrict ourselves to a space of finite dimension $\mathbb{P}$. In the case lof Maxwell molecules, one just has to consider the space $\mathbb{P}$ spanned by the q first eigenfunctions of $\mathcal{L}_B$: $(m_i(v))_{i=1,\cdots,q}$. One then set the following relaxation equations

$$\int_{\mathbb{R}^3} K(f)\, m_i\, d\mathbf{v} = \int_{\mathbb{R}^3} f\mathcal{L}_B(m_i)\, d\mathbf{v} = -\nu_i \int_{\mathbb{R}^3} f m_i\, d\mathbf{v}.$$

The generalization to other types of molecular interaction can be done in two ways. Let $\mathbb{P}$ be a (polynomial) space with $\mathbb{K} \subset \mathbb{P}$ and $\mathcal{P}$ be the projection onto $\mathbb{P}$ in $L^2(\mathcal{M})$. One then considers the linear operator $\tilde{\mathcal{L}}$ which restriction to $\mathbb{P}$ is $\mathcal{P}\mathcal{L}_B\mathcal{P}$. Notice that this approximation corresponds to the one performed in [16] for the linearized and linear Boltzmann equation. $\mathcal{P}\mathcal{L}_B\mathcal{P}$ being self-adjoint and compact, there exists an orthogonal basis $(m_i(v))_{i=1,\cdots,q}$ such that

$$\forall g \in \mathbb{P}, \quad \tilde{\mathcal{L}}(g) = \mathcal{P}\mathcal{L}_B\mathcal{P}(g) = -\sum_{i=1}^{q} \tilde{\nu}_i \mathcal{P}_{m_i}(g), \tag{8}$$

with $\tilde{\nu}_i = 0$ for $m_i \in \mathbb{K}$ and $\tilde{\nu}_i > 0$ in $\mathbb{K}^\perp$. One then replace $\mathcal{L}_B$ with $\tilde{\mathcal{L}}$ in the above approximation. Unfortunately such a direct approximation of $\mathcal{L}_B$ does not give the right transport coefficients in the hydrodynamic limit as will be shown below. Instead, the idea consists in approximation $\mathcal{L}_B^{-1}$ by stating

$$\forall g \in \mathbb{P} \cap \mathbb{K}^\perp, \quad \mathcal{L}^{-1}(g) = \mathcal{P}_{\mathbb{K}^\perp}\mathcal{L}_B^{-1}\mathcal{P}_{\mathbb{K}^\perp}(g) = -\sum_{i=6}^{q} \nu_i^{-1} \mathcal{P}_{m_i}(g), \tag{9}$$

where $\mathcal{P}_{\mathbb{K}^\perp}$ is the restriction of $\mathcal{P}$ to $\mathbb{K}^\perp$. Here the eigenvalues $(-\nu_i^{-1})_{i=6,\ldots,q}$ corresponding to each eigenfunctions in $\mathbb{K}^\perp$ are strictly negative. It must denoted that $\mathcal{L}^{-1}$ is not the pseudo-inverse of $\tilde{\mathcal{L}}$ except in the case of Maxwell molecules. With the eigenfunctions $(m_i)_i$ and eigenvalues $(-\nu_i^{-1})_i$ defined in (9), one set the following relaxation equations

$$\int_{\mathbb{R}^3} K(f)\, (1, \mathbf{v}, |\mathbf{v}|^2) d\mathbf{v} \;=\; 0, \tag{10}$$

$$\int_{\mathbb{R}^3} K(f)\; m_i(\mathbf{v})\, d\mathbf{v} \;=\; -\nu_i \int_{\mathbb{R}^3} f m_i(\mathbf{v}) d\mathbf{v}, \;\; \forall m_i \in \mathbb{P} \cap \mathbb{K}^\perp. \tag{11}$$

One can then prove formally the following proposition.

Proposition 1. *Assume (4) and (5) hold. Assume moreover that $f \to G(f)$ is smooth, then if $K(f)$ satisfies (10, 11) there is*

$$\mathcal{L} = \nu \left((\mathcal{P}_{\mathbb{K}} - \mathcal{I}) + \sum_i \left(1 - \frac{\nu_i}{\nu}\right) \mathcal{P}_{m_i} \right). \tag{12}$$

Details of the proof are let to the last section of this article.

Example 1. *Relaxation on the Grad thirteen moments : for $\mathcal{M}$ defined in (5) we consider the polynomial space*

$$\mathbb{P} = \mathbb{K} \oplus^{\perp} \mathbb{A} \oplus^{\perp} \mathbf{b} := \mathbb{K} \oplus^{\perp} \mathbb{D}, \tag{13}$$

where $\mathbb{A}$ and $\boldsymbol{b}$ are the Sonine polynomials

$$\mathbb{A}(\boldsymbol{v} - \boldsymbol{u}) = \frac{m}{k_B T} \left[(\boldsymbol{v} - \boldsymbol{u}) \otimes (\boldsymbol{v} - \boldsymbol{u}) - \frac{1}{3} \|\boldsymbol{v} - \boldsymbol{u}\|^2 \mathbb{I} \right], \tag{14}$$

$$\boldsymbol{b}(\boldsymbol{v} - \boldsymbol{u}) = (\boldsymbol{v} - \boldsymbol{u}) \left[\frac{1}{2} m (\boldsymbol{v} - \boldsymbol{u})^2 - \frac{5}{2} k_B T \right]. \tag{15}$$

The orthogonality in (13) holds for the $L^2(\mathcal{M})$ scalar product with full contraction when applied to tensors. Then there exists functions $a(|\mathbf{V}|, T) > 0$ and $b(|\mathbf{V}|, T) > 0$ with $\mathbf{V} = (\mathbf{v} - \mathbf{u})/\sqrt{k_B T/m}$ (see e.g [17]) such that

$$\mathcal{L}_B^{-1}(\mathbb{A}) = -a(|\mathbf{V}|, T)\mathbb{A}, \quad \mathcal{L}_B^{-1}(\boldsymbol{b}) = -b(|\mathbf{V}|, T)\boldsymbol{b} \tag{16}$$

so that $\mathcal{L}_B^{-1}(\mathbb{A}) \perp \boldsymbol{b}$, $\mathcal{L}_B^{-1}(\boldsymbol{b}) \perp \mathbb{A}$. As a consequence $\mathcal{L}^{-1}$ in (9) satisfies

$$\mathcal{L}^{-1}(\mathbb{A}) = -\nu_{\mathbb{A}}^{-1}\mathbb{A}, \quad \mathcal{L}^{-1}(\boldsymbol{b}) = -\nu_{\boldsymbol{b}}^{-1}\boldsymbol{b}, \tag{17}$$

for some positive values $\nu_{\mathbb{A}}$ and $\nu_{\boldsymbol{b}}$. Those values are then related to the viscosity μ_B and the thermal conductivity κ_B obtained in the Navier-Stokes limit of the Boltzmann equation

$$\mu_B = -\frac{k_B T}{10} \left\langle \mathcal{L}_B^{-1}(\mathbb{A}), \mathbb{A} \right\rangle = -\frac{k_B T}{10} \left\langle \mathcal{L}^{-1}(\mathbb{A}), \mathbb{A} \right\rangle = \frac{n k_B T}{\nu_{\mathbb{A}}}, \tag{18}$$

$$\kappa_B = -\frac{1}{3k_B T^2} \left\langle \mathcal{L}_B^{-1}(\boldsymbol{b}), \boldsymbol{b} \right\rangle = -\frac{1}{3k_B T^2} \left\langle \mathcal{L}^{-1}(\boldsymbol{b}), \boldsymbol{b} \right\rangle = \frac{5}{2} \frac{n k_B^2 T}{m \nu_{\boldsymbol{b}}}. \tag{19}$$

With those eigenfunctions and eigenvalues at hand one set two relaxation constraints in addition to the conservation laws (10)

$$\int_{\mathbb{R}^3} \nu(G - f)\mathbb{A}\, d\mathbf{v} = -\nu_{\mathbb{A}} \int_{\mathbb{R}^3} f\mathbb{A}\, d\mathbf{v}, \tag{20}$$

$$\int_{\mathbb{R}^3} \nu(G-f)\boldsymbol{b}\,d\mathbf{v} \;=\; -\nu_{\boldsymbol{b}} \int_{\mathbb{R}^3} f\,\boldsymbol{b}\,d\mathbf{v}. \tag{21}$$

Under the assumptions of proposition 1, one can perform a Chapman-Enskog expansion in

$$\partial_t f + \mathbf{v}\cdot\nabla_{\mathbf{x}} f = \frac{1}{\varepsilon} K(f). \tag{22}$$

f then writes as $f = \mathcal{M} + \varepsilon \mathcal{M} g + \mathcal{O}(\varepsilon^2)$. This provides us with the Euler equation up to an order $\mathcal{O}(\varepsilon)$ while at the next order g satisfies

$$\mathcal{L}(g) = \mathbb{A} : \mathbb{D}(\boldsymbol{u}) \;+\; \boldsymbol{b}.\nabla_x \left(-\frac{1}{k_B T}\right), \tag{23}$$

where $\mathbb{D}(\boldsymbol{u})$ is the Reynolds tensor

$$\mathbb{D}(\boldsymbol{u}) = \left[\nabla_{\boldsymbol{x}}\boldsymbol{u} + (\nabla_{\boldsymbol{x}}\boldsymbol{u})^T\right] - \frac{2}{3}(\nabla_{\boldsymbol{x}}\cdot\boldsymbol{u})\,\mathbb{I}$$

Proposition 1 allows to solve (23) and one finds that the corresponding diffusion coefficients are those given in (18). Contrarily, if one considers the operator defined in (8) the diffusion coefficients in the Navier-Stokes equations are not correct. One first remark that $\langle \mathcal{L}(\mathbb{A}), \boldsymbol{b}\rangle = 0$ thanks to the even/odd symmetry. Thus $\mathbb{A}$ and $\boldsymbol{b}$ are eigenfunctions of $\tilde{\mathcal{L}}$ with corresponding eigenvalues $-\tilde{\nu}_{\mathbb{A}}$ and $-\tilde{\nu}_{\boldsymbol{b}}$ defined by

$$\tilde{\nu}_{\mathbb{A}} = -\frac{\langle \mathcal{L}_B(\mathbb{A}), \mathbb{A}\rangle}{\langle \mathbb{A}, \mathbb{A}\rangle}, \quad \tilde{\nu}_{\boldsymbol{b}} = -\frac{\langle \mathcal{L}_B(\boldsymbol{b}), \boldsymbol{b}\rangle}{\langle \boldsymbol{b}, \boldsymbol{b}\rangle}.$$

So, in the Chapman-Enskog expansion, the Navier-Stokes equations with the following diffusion coefficients

$$\mu_K = \frac{nk_BT}{\tilde{\nu}_{\mathbb{A}}} = -nk_BT\frac{\langle \mathbb{A}, \mathbb{A}\rangle}{\langle \mathcal{L}_B(\mathbb{A}), \mathbb{A}\rangle},$$
$$\kappa_K = \frac{5}{2}\frac{nk_B^2T}{m\tilde{\nu}_{\boldsymbol{b}}} = -\frac{5}{2}\frac{nk_B^2T}{m}\frac{\langle \boldsymbol{b}, \boldsymbol{b}\rangle}{\langle \mathcal{L}_B(\boldsymbol{b}), \boldsymbol{b}\rangle}.$$

By comparison with (18) one finds that $\mu_K = \mu_B$ and $\kappa_K = \kappa_B$ for Maxwell molecules. In the other cases, $\mu_K = \mu_B$ reads also $\langle \mathcal{L}_B^{-1}(\mathbb{A}), \mathbb{A}\rangle = C(n,T)\langle \mathcal{L}_B(\mathbb{A}), \mathbb{A}\rangle^{-1}$ where $C(n,T)$ does not depend on the interaction potential which is untrue (and likewise for the heat conductivity). Remark finally that such a problem arises in the hydrodynamic limit of moment system for the Boltzmann equation [33]. It is then found that $\mu_K < \mu_B$ and $\kappa_K < \kappa_B$.

The final step of the method consists in defining ν and G in a proper way. This is part of the following mathematical problems that we are going to study :

1. Characterization of the set of realizable moments $\mathcal{R}^+_{\mathbf{m}}$: the set of constraints writes

$$\int_{\mathbb{R}^3} G\,m_i d\mathbf{v} = (1-\frac{\nu_i}{\nu})\int_{\mathbb{R}^3} f\,m_i d\mathbf{v},\ i\in\{1,\dots,q\} \tag{24}$$

$$\Leftrightarrow \boldsymbol{\rho}_G := L(\boldsymbol{\rho}_f) \tag{25}$$

If one assumes that the moments $\boldsymbol{\rho}_f$ of the nonnegative function are bounded - $\boldsymbol{\rho}_f \in \mathcal{R}^+_{\mathbf{m}}$ - then the range of L must be in $\mathcal{R}^+_{\mathbf{m}}$ as well. The set $(m_i, \nu_i)_i$ being already defined in (9), ν must be defined so that this important property holds. This requires to have a tractable way to characterize $\mathcal{R}^+_{\mathbf{m}}$ and to study the range of L as a linear operator depending on ν.

2. How to define $G(f)$? Assume that

$$C_f = \left\{ h \geq 0, \int h(\mathbf{v})\, \mathbf{m}(\mathbf{v}) d\mathbf{v} = L(\boldsymbol{\rho}_f) \right\} \neq \emptyset. \tag{26}$$

The final step of the method is performed by solving a variational problem and defining G as

$$G = \underset{h \in C_f}{\arg\min} \int \eta(h)\, d\mathbf{v}, \tag{27}$$

where η is a given functional. Statistical physics state that η must be defined as $\eta_B(.) = .\ln(.)$. Unfortunately theoretical and numerical studies have revealed numerous problems as soon as there are more constraints than (10) and (20) [30]. In particular there is no solution to the variational problem under the constraints (10, 11) in the example of Grad relaxation on thirteen moments. Adding more constraints neither provides us with a well-posed variational problem in general. This explains why other choices of η must be found and why we only ask for some local stability through (3), (4) and (5).

3 The moment problem

3.1 Introducing the problem of realizable moment

Let $(\mathbf{m}_0(\mathbf{v})), \cdots, \mathbf{m}_n(\mathbf{v}))$ be a list of (symmetrical tensor) polynomial functions and $\boldsymbol{\rho}_0, \cdots, \boldsymbol{\rho}_n$ be a list of (symmetrical) tensors of same order. Let $\mathbb{R}^3$ be equipped with the Lebesgue measure $d\mathbf{v}$. The problem of realizability is stated as follows : is there an integrable (non null) non negative function $f : \mathbb{R}^3 \mapsto \mathbb{R}^+$ such that

$$\forall k \in [0, n], \quad \int \mathbf{m}_k(\mathbf{v})\, f(\mathbf{v})\, d\mathbf{v} = \boldsymbol{\rho}_k. \tag{28}$$

$\boldsymbol{\rho} := (\boldsymbol{\rho}_0, \cdots, \boldsymbol{\rho}_n)$ is said realizable when (28) holds. In whole generality, this problem is stated in a more general framework. That is: is there is a Borel measure μ such that

$$\int_{\mathbb{R}^3} \mathbf{m}_k(\mathbf{v})\ d\mu = \rho_k. \tag{29}$$

The existence of such a measure is related to nonnegative polynomials according to the following remark. If μ exists, then for any nonnegative polynomial writing as

$P(\mathbf{v}) = \boldsymbol{\alpha} \cdot \boldsymbol{m}(\mathbf{v})$ there is

$$\int_{\mathbb{R}^3} \boldsymbol{\alpha} \cdot \boldsymbol{m}(\mathbf{v})\, d\mu = \boldsymbol{\alpha} \cdot \boldsymbol{\rho} \geq 0. \tag{30}$$

Definition 1. *$L_{\boldsymbol{\rho}}(P) = \boldsymbol{\rho} \cdot \boldsymbol{\alpha}$ is the Riesz functional at P. It is positive when $L_{\boldsymbol{\rho}}(P) \geq 0$ for all $P \geq 0$ in $\mathbb{P}$.*

The converse statement of (30), that is $L_{\boldsymbol{\rho}}(P) \geq 0$ for any nonnegative polynomials in $\mathbb{P}$ implies the existence of a Borel measure, is not true in general. More precisely, if $\mathbb{P}_{2n} = \mathtt{span}\ (\mathbf{m}_0, \cdots, \mathbf{m}_n)$, the Riesz-Haviland theorem states that $\boldsymbol{\rho}$ is the moment of a Borel measure iff there exists an extension to it in $\tilde{\rho} \in \mathbb{R}^{dim(\mathbb{P}_{2n+2})}$ whose $dim(\mathbb{P}_{2n})$ first components are those of $\boldsymbol{\rho}$ and such that $\tilde{L}_{\tilde{\boldsymbol{\rho}}}$ is positive in $\mathbb{P}_{2n+2}$ [24]. Contrarily, the existence of such an extension is not necessary when addressing the case of realizable moments as proved by Junk [30]. The characterization of realizable moments then turns into the characterization of nonnegative polynomials in $\mathbb{P} = span\ (\mathbf{m}_0, \cdots, \mathbf{m}_n)$ which is known as the 17^{th} Hilbert problem.

In the sequel we study Junk's theorem by clarifying but also modifying some steps in the proof. Then we display known results on the characterization of positive polynomials, give their link with the criteria of realizability in one dimension and point out the limitation of those results in higher dimension. Next we address the case of Grad thirteen moments where the characterization by positive polynomials provides us with a tractable criteria for realizability. Finally, the last section is devoted to a special class of linear maps onto the set of realizable moments which are compatible with the Galilean invariance of the model to be constructed.

3.2 Junk's theorem

Let us first assume that $\mathbf{m} := (\mathbf{m}_0, \cdots, \mathbf{m}_k, \cdots, \mathbf{m}_n)^T$ is pseudo-Haar with the following meaning.

Definition 2. *Let $\mathbf{m} := (\mathbf{m}_0, \cdots, \mathbf{m}_n)$ be a list of (tensor structured) functions defined on $\mathbb{R}^3$. The list is pseudo-Haar when the following property is satisfied:*

$$\forall \boldsymbol{\alpha}, \quad [\boldsymbol{\alpha} \neq \mathbf{0} \Rightarrow \boldsymbol{\alpha} \cdot \mathbf{m}(\mathbf{v}) \neq 0], \quad a.e\ \mathbf{v} \in \mathbb{R}^d$$

We denote by q the total dimension of the space generated by $\mathbf{m}$.

Definition 3. *We denote with $\mathbb{L}^1(\mathbf{m})$ the set of integrable functions $f : \mathbb{R}^3 \mapsto \mathbb{R}$ such that*

$$\forall k \in [0, n], \quad \int \|\mathbf{m}_k(\mathbf{v})\|\, |f(\mathbf{v})|\, d\mathbf{v} < +\infty.$$

When $f \in \mathbb{L}^1(\mathbf{m})$ is non negative, we note $f \in \mathbb{L}^{1,+}(\mathbf{m})$, and when it is not zero we note $f \in \mathbb{L}^{1,*}(\mathbf{m})$.

Now we introduce the moment map

Definition 4. *Let $\mathbf{m}(\mathbf{v})$ be pseudo-Haar. Then we define the map $R : \mathbb{L}^1(\mathbf{m}) \mapsto \mathbb{R}^q$ as follows*

$$\forall f \in \mathbb{L}^1(\mathbf{m}), \quad R[f] = \int \mathbf{m}(\mathbf{v})\, f(\mathbf{v})\, d\mathbf{v}$$

we adopt for the sequel the following notations

$$\mathcal{R}_{\mathbf{m}}^{+}=\left\{\mathcal{R}[f],f\in\mathbb{L}^{1,+}(\mathbf{m})\right\},\quad\mathcal{R}_{\mathbf{m}}^{*+}=\left\{\mathcal{R}[f],f\in\mathbb{L}^{1,*,+}(\mathbf{m})\right\}.$$

In principle, most results are concerned with finding a Borel measure satisfying (29) and not necessarily represented by L^1 functions. Apart of this extensive literature, the theorem stated by Junk addresses the case of integrable functions. So we recall the theorem together with the steps of the proof.

Theorem 2. *(Junk [30]) A vector $\boldsymbol{\rho}\in\mathcal{R}_{\mathbf{m}}^{*+}$ iff all $\boldsymbol{\alpha}\neq 0\in\mathbb{R}^q$ which satisfies $\boldsymbol{\alpha}\cdot\mathbf{m}(\mathbf{v})\leq 0$ on $\mathbb{R}^3$, the relation $\boldsymbol{\alpha}\cdot\boldsymbol{\rho}<0$ holds.*
In particular, $\mathcal{R}_{\mathbf{m}}^{*+}$ is an open set. Moreover each $\boldsymbol{\rho}\in\mathcal{R}_{\mathbf{m}}^{*+}$ is a moment vector of bounded $f\in\mathbb{L}^{1,*}(\mathbf{m})$ which is compactly supported.

The proof relies essentially on a duality argument. The right implication is quite obvious. Consider indeed a moment $\boldsymbol{\rho}\in\mathcal{R}_{\mathbf{m}}^{*+}$, that is a vector such that there exists a nonnegative and non null function $f\in\mathbb{L}^{1,+}(\mathbf{m})$ satisfying $\boldsymbol{\rho}=\int\mathbf{m}(\mathbf{v})f(\mathbf{v})d\mathbf{v}$. Let $\boldsymbol{\alpha}\cdot\mathbf{m}(\mathbf{v})$ be a non positive and non null polynomial, then

$$\boldsymbol{\alpha}\cdot\boldsymbol{\rho}=\boldsymbol{\alpha}\cdot\int\mathbf{m}(\mathbf{v})f(\mathbf{v})d\mathbf{v}=\int\boldsymbol{\alpha}\cdot\mathbf{m}(\mathbf{v})f(\mathbf{v})d\mathbf{v}<0 \tag{31}$$

since $f\neq 0$ on a set of non zero measure and $\boldsymbol{\alpha}\cdot\mathbf{m}(\mathbf{v})\neq 0$ almost everywhere. So, up to a sign, this equation is very similar to (30) except that the inequality is strict. $\mathcal{R}_{\mathbf{m}}^{+}$ is a positive cone and the above computation proves that its polar cone is the set of coefficients of non positive polynomials :

$$\begin{aligned}(\mathcal{R}_{\mathbf{m}}^{+})^{\circ}&=\left\{\boldsymbol{\alpha}\in\mathbb{R}^q,\ \forall\boldsymbol{\rho}\in\mathcal{R}_{\mathbf{m}}^{+},\ \ \boldsymbol{\alpha}\cdot\boldsymbol{\rho}\leq 0\right\}\\&=\left\{\boldsymbol{\alpha}\in\mathbb{R}^q,\ \forall f\in\mathbb{L}^{1,+}(\mathbf{m}),\ \int\boldsymbol{\alpha}\cdot\mathbf{m}(\mathbf{v})f(\mathbf{v})d\mathbf{v}\leq 0\right\}\\&=\left\{\boldsymbol{\alpha}\in\mathbb{R}^q,\ \boldsymbol{\alpha}\cdot\mathbf{m}(\mathbf{v})\leq 0\right\}:=C^{\circ}.\end{aligned}$$

If the convex cone $\mathcal{R}_{\mathbf{m}}^{+}$ is of non empty interior, one finds that $\mathcal{R}_{\mathbf{m}}^{+*}$ is characterized by the right statement in theorem 2 according to (31) and the following characterization of cone's interior

Theorem 3. *Assume C is a convex cone of $\mathbb{R}^q$ with non empty interior, then*

$$\forall\mathbf{y}\in\mathbb{R}^q,\quad\left[\mathbf{y}\in\boldsymbol{int}(C)\Leftrightarrow\forall\boldsymbol{\alpha}\in C^{\circ},\ \ \boldsymbol{\alpha}\neq\mathbf{0}\Rightarrow\boldsymbol{\alpha}\cdot\mathbf{y}<0\right].$$

It is possible to prove that $\mathcal{R}_{\mathbf{m}}^{+*}$ is an open set in the case where the space of velocity is a bounded set in one dimension [36] but such a construction is not possible in other cases. A more general argument consists in considering the set

$$C=\left\{\sum_i\lambda_i\mathbf{m}(\mathbf{v}_i),\lambda_i\geq 0,\ \ \forall\mathbf{v}_i\in\mathbb{R}^3\right\} \tag{32}$$

which is the set of moments of all atomic measures. C should contain $\mathcal{R}_{\mathbf{m}}^{+}$ according to Tchakaloff's theorem on quadrature rules [22]. Indeed, this theorem implies that a compactly supported Borel measure has the same moment as the one of an atomic measure. Moreover, the moments in C are those of all atomic measures without any restriction in their support.
It is easily seen that the polar cone of C is the same as the one of $\mathcal{R}_{\mathbf{m}}^{+}$

$$C^{\circ} = \left\{\boldsymbol{\alpha} \in \mathbb{R}^{q},\ \ \boldsymbol{\alpha} \cdot \mathbf{m}(\mathbf{v}) \leq 0,\ \forall \mathbf{v} \in \mathbb{R}^{3}\right\} \tag{33}$$

$$= (\mathcal{R}_{\mathbf{m}}^{+})^{\circ}. \tag{34}$$

So, one proves in a first time
Proposition 4. *For C defined in (32) $\mathtt{int}(C) \neq \varnothing$.*

This implies that $\mathcal{R}_{\mathbf{m}}^{+*} \subset \mathtt{int}\, C$ according to theorem 3. Then remark that for $\Psi_{\varepsilon} \in C_{c}^{\infty}(\mathbb{R}^{3})$, $\Psi_{\varepsilon} \geq 0$ such that $\Psi_{\varepsilon} \to \delta_{0}$ as $\varepsilon \to 0$, one has

$$\forall \mathbf{v} \in \mathbb{R}^{d}, \qquad \int \mathbf{m}(x)\, \Psi_{\varepsilon}(\mathbf{v} - \mathbf{x}) d\mathbf{x} \longrightarrow \mathbf{m}(\mathbf{v})$$

This proves that the set of moments of nonnegative functions in $C_{c}^{\infty}(\mathbb{R}^{d})$ is dense in C ($C \subset \mathtt{cl}\, \mathcal{R}_{\mathbf{m}}^{+}$). To summarize, one has

$$\mathcal{R}_{\mathbf{m}}^{+*} \subset \mathtt{int}\, C \subset \mathtt{int}\left(\mathtt{cl}\, \mathcal{R}_{\mathbf{m}}^{+}\right)$$

$\mathcal{R}_{\mathbf{m}}^{+*}$ is a convex set so that $\mathcal{R}_{\mathbf{m}}^{+*} = \mathtt{int}\left(\mathtt{cl}\, \mathcal{R}_{\mathbf{m}}^{+}\right)$ according to a Caratheodory theorem in finite dimension [38] : each $\boldsymbol{\rho} \in \mathtt{int}\left(\mathtt{cl}\, \mathcal{R}_{\mathbf{m}}^{+}\right)$ is the convex combination of $q+1$ affinely independent points in $\mathtt{int}(\mathtt{cl}\, \mathcal{R}_{\mathbf{m}}^{+})$ and by density of $q+1$ affinely independent points in $\mathcal{R}_{\mathbf{m}}^{+*}$. So finally there is

$$\mathcal{R}_{\mathbf{m}}^{+*} = \mathtt{int}\, C = \mathtt{int}\left(\mathtt{cl}\, \mathcal{R}_{\mathbf{m}}^{+}\right)$$

which proves the first assertion in theorem 2 together with "$\mathcal{R}_{\mathbf{m}}^{+*}$ is an open set". Finally, the $q+1$ affinely independent points in the above reasoning can be chosen as moments of nonnegative functions in $C_{c}^{\infty}(\mathbb{R}^{d})$ by density which ends the proof.
Remark 1. *Let $\mathbb{P}$ be the space generated by the components of the set of tensors $(\mathbf{m}_{k}(\mathbf{v}))_{k}$. Then the first statement of theorem 2 also writes as :*

$$\boldsymbol{\rho} \in \mathcal{R}_{\mathbf{m}}^{*+} \Longleftrightarrow \left[L_{\boldsymbol{\rho}}(P) > 0 \quad \forall P \in \mathbb{P} \quad \textit{with} \quad P(\mathbf{v}) \geq 0\ (P(\mathbf{v}) \neq 0)\right].$$

Thus, when $\mathbb{P} = \mathbb{P}_{2n}$, the characterization of a realizable moment does not require any property related to polynomials of higher degree contrarily to moments of Borel measure [24].

3.3 Moment matrices and sum of square polynomials

According to Theorem 2, characterizing realizable moments is equivalent to characterizing those $\boldsymbol{\alpha}$ for which $\boldsymbol{\alpha} \cdot \boldsymbol{m}(\mathbf{v}) \leq 0$ on $\mathbb{R}^{3}$. As we are working with a polynomial

pseudo Haar family, the problem is usually stated in term of nonnegative polynomials. In one dimension and for polynomial spaces spanned by $\boldsymbol{m}(v) = \{\mathbf{1}, v, v^2, \dots, v^{2p}\}^t$, the problem is known as the Hamburger moment problem. A well known theorem states that $\boldsymbol{\rho} = \{\rho_0, \rho_1, \dots, \rho_{2p}\} \in \mathcal{R}^{+,*}_{\mathbf{m}}$ iff the moment matrix $(\rho_{i+j})_{i=0,\cdots,p,\, j=0,\cdots,p}$ is positive definite. While the original proof does not make mention to it, the result may be obtained by the characterization of positive polynomials in one dimension. That is:

- In one dimension, every non negative polynomials in $\mathbb{P} = \mathtt{span}(\boldsymbol{m})$ writes as a sum of square polynomials

$$p(v) = \sum_i (\beta_0^i + \beta_1^i v + \ldots + \beta_p^i v^p)^2 = \sum_i (\boldsymbol{\beta}_i \cdot \widetilde{\mathbf{m}}(v))^2, \tag{35}$$

  where $\boldsymbol{\beta}_i = (\beta_0^i, \beta_1^i \mathbf{v}, \cdots, \beta_p^i)^t \in \mathbb{R}^{p+1}$ and $\widetilde{\mathbf{m}}(\mathbf{v}) = (1, \mathbf{v}, \cdots, \mathbf{v}^p)^t$. This result is a consequence of the fundamental theorem of algebra (d'Alembert-Gauss).
- According to Theorem 2, $\boldsymbol{\rho} \in \mathcal{R}^+_{\mathbf{m}}$ iff for every $\boldsymbol{\alpha}$ s.t $\boldsymbol{\alpha} \cdot \mathbf{m}(v) \geq 0$ and $\boldsymbol{\alpha} \neq 0$, there holds $\boldsymbol{\alpha} \cdot \boldsymbol{\rho} > 0$. Equivalently for any $f \in \mathbb{L}^{1,*,+}(\boldsymbol{m})$ such that $\int f \boldsymbol{m}(v) dv = \boldsymbol{\rho}$ then there is also

$$\int \boldsymbol{\alpha} \cdot \mathbf{m}(v)\, f(v)\, dv = \boldsymbol{\alpha} \cdot \boldsymbol{\rho} > 0 \tag{36}$$

  Using (35) this also writes

$$\sum \boldsymbol{\beta}_i^t \left(\int_{\mathbb{R}^3} \widetilde{\mathbf{m}}(v) \widetilde{\mathbf{m}}(v)^t\, f(v) dv \right) \boldsymbol{\beta}_i > 0.$$

  Thus $\boldsymbol{\rho}$ is realizable iff the moment matrix

$$\boldsymbol{H} = \left(\int_{\mathbb{R}} \widetilde{\mathbf{m}}(v) \widetilde{\mathbf{m}}(v)^t f(v) dv \right)$$

  is symmetric positive definite.

In higher dimension, the problem may be generalized as follows.

Definition 5. *Let $\widetilde{\mathbf{m}} = \{1, \widetilde{m_1}, \dots, \widetilde{m_p}\}$ be a pseudo-Haar family composed of monomials in $\mathbb{R}[v_1, \cdots, v_d]$, $d > 1$ and $\widetilde{\mathbb{P}}$ be the polynomial space generated by this family. We denote with*

$$\mathbb{P} = \mathtt{span}(\widetilde{m}_i \widetilde{m}_j),\ 0 \leq i \leq j \leq p) \tag{37}$$

the space generated by the family of pairwise multiplication of any element in $\widetilde{\mathbf{m}}(\mathbf{v})$. We say that $\mathbb{P}$ is the quadratic space over $\widetilde{\mathbb{P}}$. The dimension of $\mathbb{P}$ being at most $\frac{(p+1)(p+2)}{2}$, one may extract from $(\tilde{m}_i \tilde{m}_j)_{i,j}$ a pseudo-Haar family which we denote as before with $\mathbf{m}(\mathbf{v})$.

Example 1: Each polynomial space $\mathbb{P}_{2p}$, $p \geq 1$ is quadratic over $\mathbb{P}_p$.

Example 2: The space of collisional invariant $\mathbb{K}$ is not quadratic over another space if

$d > 1$ but the space generated by $\{\mathbf{1},\ \mathbf{v},\ \mathbf{v}\otimes\mathbf{v},\ \mathbf{v}\mathbf{v}^2,\ \mathbf{v}^4\}$ is quadratic over $\mathbb{K}$ (here we have used the tensorial notation for convenience.) It is a space strictly contained in $\mathbb{P}_4$.

Let us write now the condition $\boldsymbol{\rho}\cdot\boldsymbol{\alpha} > 0$ for square polynomials that is $p(\mathbf{v}) = \boldsymbol{\alpha}\cdot\boldsymbol{m}(\mathbf{v}) = (\boldsymbol{\beta}\cdot\tilde{\boldsymbol{m}}(\mathbf{v}))^2$. For all $k \in \{1,\dots,dim(\mathbb{P})\}$, denote with $I_k = \{(i,j) \in \{1;p\}$ s.t. $\tilde{m}_i\tilde{m}_j = m_k\}$. Then

$$\alpha_k = \sum_{(i,j)\in I_k} \beta_i\beta_j. \tag{38}$$

Correspondingly, for any $(i,j) \in I_k$, we denote $\rho_{ij} = \rho_k$. There is

$$\sum_{i,j=1}^{p} \rho_{ij}\beta_i\beta_j = \sum_{k=1}^{dim(\mathbb{P})} \sum_{(i,j)\in I_k} \rho_{ij}\beta_i\beta_j = \sum_{k=1}^{dim(\mathbb{P})} \rho_k\alpha_k = \boldsymbol{\rho}\cdot\boldsymbol{\alpha},$$

because of relation (38). So if we write the condition $\boldsymbol{\rho}\cdot\boldsymbol{\alpha} > 0$ for all square polynomials (and thus for all sum of square polynomials), we end up with an equivalent condition which is: the moment matrix $\mathbf{H} = (\rho_{ij})$ is symmetric positive definite (SPD). This means that those moments whose moment matrix is symmetric definite positive constitute a space which is $\mathcal{R}^+_{\mathbf{m}}$ or contains it. The condition on $\mathbf{H}$ becomes sufficient only if all polynomials are sum of squares (SOS). This question is part of the celebrated 17^{th} Hilbert problem. We refer to [6] for a bibliography on the topic. It turns out that in dimension bigger than 1, there exists positive polynomials which are not sum of square. As an example, for $d > 3$, there exists polynomials in $\mathbb{P}_4$ and of higher degree which are not SOS. Questions relating SPD moment matrix, SOS polynomials and realizable moments are addressed for example in [23, 24].

To conclude with, we always start in the method of moments relaxation with a moment $\boldsymbol{\rho}_f$ of a nonnegative function f. So, the criteria that is given by the moment matrix allows to discard most relaxations (24) that lead to non admissible moments, but not all of them.

3.4 Application to relaxation of the Grad moments

So far, we have focused on pretty general consideration on realizable moment. In practice, some usual basis in kinetic theory are

- "Euler" basis $\{\mathbf{1},\mathbf{v},\mathbf{v}^2\}$
- Gauss basis $\{\mathbf{1},\mathbf{v},\mathbf{v}\otimes\mathbf{v}\}$
- Grad basis $\{\mathbf{1},\mathbf{v},\mathbf{v}\otimes\mathbf{v},\mathbf{v}\,\mathbf{v}^2\}$
- Levermore basis $\{\mathbf{1},\mathbf{v},\mathbf{v}\otimes\mathbf{v},\mathbf{v}\mathbf{v}^2,\mathbf{v}^4\}$

The last one has no particular physical interpretation and was just introduced for solving variational problems with the usual entropy $\int f\,\ln(f)d\mathbf{v}$. Among those basis, the Grad one's is the most important because it contains all physically meaningful moments (mass, momentum, energy, pressure tensor, heat flux). In the sequel, we use the decomposition of the Grad space defined in (13). The polynomial family of interest

is the pseudo-Haar families denoted as:

$$\mathbf{a}(\mathbf{v}-\mathbf{u}) = \left(\mathbf{1}, (\mathbf{v}-\mathbf{u}), (\mathbf{v}-\mathbf{u})^2 - 3\frac{k_B T}{m}, \frac{k_B T}{m}\mathbb{A}(\mathbf{v}-\mathbf{u}), \mathbf{b}(\mathbf{v}-\mathbf{u})\right) \quad (39)$$

$$:= \left(a_0(\mathbf{v}-\mathbf{u}), \mathbf{a}_1(\mathbf{v}-\mathbf{u}), a_2(\mathbf{v}-\mathbf{u}), a_3(\mathbf{v}-\mathbf{u}), \mathbf{a}_4(\mathbf{v}-\mathbf{u})\right). \quad (40)$$

It is still composed of tensors of even or odd ranks in the variable $\mathbf{v}-\mathbf{u}$. The polar cone of $\mathcal{R}_{\mathbf{a}}^{*+}$ is given by:

$$\mathcal{C}_{\mathbf{a}}^{\circ} = \left\{\boldsymbol{\alpha} \in \mathbb{R}^q, \quad \forall \mathbf{v} \in \mathbb{R}^d, \quad \boldsymbol{\alpha} \cdot \mathbf{a}(\mathbf{v}-\mathbf{u}) \leq 0\right\}.$$

It is immediate that any $\boldsymbol{\alpha} \in \mathcal{C}_{\mathbf{a}}^{\circ}$ has a null component on $(\mathbf{v}-\mathbf{u})(\mathbf{v}-\mathbf{u})^2$ and as a consequence on $\mathbf{b}(\mathbf{v}-\mathbf{u})$. Then we have

$$\mathcal{C}_{\mathbf{a}}^{\circ} = \left\{(\boldsymbol{\beta}, \mathbf{0}), \forall \mathbf{v} \in \mathbb{R}^d, \ \boldsymbol{\beta} \cdot \left(1, (\mathbf{v}-\mathbf{u}), (\mathbf{v}-\mathbf{u})^2 - 3\frac{k_B T}{m}, a_3(\mathbf{v}-\mathbf{u})\right) \leq 0\right\}.$$

The study of realizable moments on $\mathbf{a}(\mathbf{v}-\mathbf{u})$ then simplifies dramatically since we just need to know which moments are realizable in the Gauss basis defined as $\mathbb{P}_{\text{Gauss}} = \texttt{span}\{\mathbf{1}, \mathbf{v}, \mathbf{v}^2, \mathbb{A}(\mathbf{v})\}$. Using Theorem 2, the characterization of $\mathcal{R}_{\mathbf{a}}^{*+}$ is equivalent to the characterization of the positive polynomials in $\mathbb{P} = \texttt{span}\{\mathbf{1}, \mathbf{v}, \mathbf{v}\otimes\mathbf{v}\}$. From Hilbert's theorems, any positive polynomial in $\mathbb{P}$ can be written as a sum of square of polynomials, that is

$$g(\mathbf{v}) = \sum_i (p_i + \mathbf{q}_i \cdot \mathbf{v})^2,$$

with $p_i \in \mathbb{R}$ and $q_i \in \mathbb{R}^3$. Let $\boldsymbol{\rho} = (n, n\mathbf{u}, n\mathbb{D})$ be a moment w.r.t. this $(\mathbf{1}, \mathbf{v}, \mathbf{v}\otimes\mathbf{v})$. According to Theorem 2, $\boldsymbol{\rho}$ is realizable if and only if for any non negative, non null polynomial $g(\mathbf{v})$ defined by its list of coefficients $\boldsymbol{\gamma}$ there holds $\boldsymbol{\gamma}\cdot\boldsymbol{\rho} > 0$. Rearranging the components of $g(\mathbf{v})$ in the Gauss basis, we find after some algebra that this condition can be written as

$$n\sum_i \left([p_i \ \mathbf{q}_i]\begin{bmatrix}\mathbf{1} & \mathbf{u}^t \\ \mathbf{u} & \mathbb{D}\end{bmatrix}\begin{bmatrix}p_i \\ \mathbf{q}_i\end{bmatrix}\right) > 0 := n\sum_i \left([p_i \ \mathbf{q}_i]\,\mathbf{H}\begin{bmatrix}p_i \\ \mathbf{q}_i\end{bmatrix}\right).$$

This is equivalent to $n > 0$ and to the positivity of the moment matrix $\mathbf{H}$. But $\mathbf{H}$ is positive if and only $\mathbb{D} - \mathbf{u}\otimes\mathbf{u}$ is positive as proved in the following lemma:

Lemma 1. *The matrix*

$$\begin{bmatrix}1 & \boldsymbol{u}^t \\ \boldsymbol{u} & \mathbb{D}\end{bmatrix}$$

is positive if and only if the matrix $\mathbb{D} - \boldsymbol{u}\otimes\boldsymbol{u}$ *is positive.*

It is interesting to remark that for a given realizable moment $\boldsymbol{\rho} := \{n, n\mathbf{u}, 3ne, n\Pi, n\mathbf{Q}\}$, the heat flux $n\mathbf{Q}$ can take any value in $\mathbb{R}^3$ because of the characterization given by theorem 2 which does not include moments of order 3. This is

due to the fact that one can always add to a positive function having this moment a perturbation that let it nonnegative, keeps all of its moments in the Gauss basis but not that with respect to $\mathbf{b}\,(\mathbf{v}-\mathbf{u})$. The Grad basis being conveniently written in (39), that is starting from a nonnegative function f, there is

$$\int_{\mathbb{R}^3} f\,\mathbf{a}(\mathbf{v}-\mathbf{u})d\mathbf{v} = (n,\,0,\,0,\,n\Pi,\,n\mathbf{Q})\,.$$

whete $n\Pi$ is the traceless pressure tensor and $n\mathbf{Q}$ is the heat flux. A straightforwards consequence of lemma 1 is the following proposition

Proposition 5. *[Grad relaxation] If $(n,\mathbf{0},0,n\Pi,n\mathbf{Q})$ is realizable, then for any $\lambda_{\mathbb{A}} \in [-\frac{1}{2},1]$, and $\lambda_b \in \mathbb{R}$ the moment $(n,\mathbf{0},0,n\lambda_{\mathbb{A}}\Pi,n\lambda_b\mathbf{Q})$ is still realizable.*

Coming back to the relaxation constraints, the above result states that the relaxed moment $(n,\,0,\,0,\,(1-\frac{\nu_{\mathbb{A}}}{\nu})n\Pi,\,(1-\frac{\nu_{\mathbf{b}}}{\nu})n\mathbf{Q})$ is realizable when $0 \leq 1-\frac{\nu_{\mathbb{A}}}{\nu} \leq 1$ and for all $\nu_{\mathbf{b}} \in \mathbb{R}$. Also, the admissible relaxation on $n\Pi$ is just the one that is found in the study of the ESBGK model [4, 12].

3.5 Stability under Galilean transformations

We may now address the problem of finding linear application that let $\mathcal{R}^+_{\boldsymbol{m}}$ invariant. However, it is important to set this question in the context of kinetic theory. In the sequel we are going to focus on the relation between such linear maps and Galilean invariance. The method developed in section 2.2 for constructing relaxation operators $K(f)$ must satisfy

$$\tau_{\mathbf{u}}(K(f)) = K(\tau_{\mathbf{u}}f) \quad \forall \mathbf{u}\in\mathbb{R}^3 \quad \textbf{and} \quad \tau_\Theta K(f) = K(\tau_\Theta f)\ \forall \Theta \in SO(3), \tag{41}$$

where

$$\tau_{\mathbf{u}}f(\mathbf{v}) = f(\mathbf{v}+\mathbf{u}) \quad \textbf{and} \quad \tau_\Theta f(\mathbf{v}) = f(\tau_\Theta \mathbf{v}).$$

Recall that $K(f) = \nu(G-f)$. The setting of ν and construction of G just depends on the moments of f - $\boldsymbol{\rho} = \int f(\mathbf{v})\mathbf{m}d\mathbf{v} \in \mathcal{R}^{*+}_{\mathbf{m}}$ - and not on f itself. We may write $\nu = \nu(\boldsymbol{\rho}(f))$ and $G = G(\boldsymbol{\rho}(f))$. So, starting from a function $f \in \mathbb{L}^{1,+}(\mathbf{m})$, $\tau_{\mathbf{u}}f$ and $\tau_\Theta f$ must themselves be function of $\mathbb{L}^{1,+}(\boldsymbol{m})$ to make the construction $G(\boldsymbol{\rho}(\tau_{\mathbf{u}}f))$ (likewise $G(\boldsymbol{\rho}(\tau_\Theta f))$) possible. This writes

$$\boldsymbol{\rho}(\tau_{\mathbf{u}}f) = \int (\tau_{\mathbf{u}}f)\mathbf{m}(\mathbf{v})d\mathbf{v} = \int f(\mathbf{w})\mathbf{m}(\tau_{-\mathbf{u}}(\mathbf{w}))d\mathbf{w} \in \mathcal{R}^{*+}_{\mathbf{m}}.$$

A sufficient condition for integrability of $\tau_{\mathbf{u}}f$ in $\mathbb{L}^1(\mathbf{m})$ is: $\mathbb{P} = \mathtt{span}(\mathbf{m})$ is invariant under the action of $\tau_{-\mathbf{u}}$, which can be expressed as

$$\exists \Lambda(-\mathbf{u}) \in \mathbb{R}^q \times \mathbb{R}^q \text{ such that } \mathbf{m}(\tau_{-\mathbf{u}}(\mathbf{w})) = \Lambda(-\mathbf{u})\mathbf{m}(\mathbf{w}). \tag{42}$$

Positivity is then satisfied since $\tau_{\mathbf{u}}f \geq 0$ and there is $\boldsymbol{\rho}(\tau_{\mathbf{u}}f) \in \mathcal{R}^{*+}_{\mathbf{m}}$. As a consequence $\mathcal{R}^{*+}_{\mathbf{m}}$ is stable under the action of the linear map $\Lambda(-\mathbf{u})$ which inverse is $\Lambda(\mathbf{u})$. Likewise,

if there exists $\Lambda(\Theta^t) \in \mathbb{R}^q \times \mathbb{R}^q$ such that

$$\mathbf{m}(\Theta^t \mathbf{w}) = \Lambda(\Theta^t)\mathbf{m}(\mathbf{w}), \tag{43}$$

the same conclusion holds. The above conditions relate polynomial spaces $\mathbb{P}$ which are invariant under the action of translations and rotations to set of moments $\mathcal{R}^{*+}_{\mathbf{m}}$ which are invariant under the above mapping. Such polynomial spaces are named Galilean invariant and are necessarily polynomial as proved by Junk and Unterreiter [31].

Proposition 6. *Assume that the space $\mathbb{P} = \mathtt{span}\{\mathbf{m}\}$ is invariant under the translations and the rotations. Then $\mathcal{R}^{+}_{\mathbf{m}}$ is invariant under the action of $\Lambda(\mathbf{u})$ and $\Lambda(\Theta)$ for any $\mathbf{u} \in \mathbb{R}^3$ and any $\Theta \in SO(3)$, where $\Lambda(\mathbf{u})$ and $\Lambda(\Theta)$ are the matrices defined in (42) and (43).*

Instead of proving the equivalence in the above conclusion, we prefer to look at the construction of the model and state a necessary condition on $\mathbb{P}$.

Proposition 7. *Let $G : \mathcal{R}^{+}_{\mathbf{m}} \to \mathbb{L}^{1,+}(\mathbf{m})$ s.t. $G(\boldsymbol{\rho}_1) \neq G(\boldsymbol{\rho}_2)$ for any $\boldsymbol{\rho}_1 \neq \boldsymbol{\rho}_2 \in \mathcal{R}^{+}_{\mathbf{m}}$. Then if G satisfies*

$$(\forall f \in \mathbb{L}^{1,+}(\mathbf{m})),\ (\forall \mathbf{u} \in \mathbb{R}^3),\ (\forall \Theta \in SO(3)),\ G(R[\tau_{\mathbf{u}} f]) = \tau_{\mathbf{u}} G(R(f)),\quad G(R[\tau_{\Theta} f]) = \tau_{\Theta} G(R[f]),$$

the space $\mathbb{P} = \mathtt{span}\{\mathbf{m}\}$ is invariant under Galilean transforms.

4 Solving the variational problem

From now on, we assume that a vector $\boldsymbol{\rho} \in \mathcal{R}^{+}_{\mathbf{m}}$ has been obtained after some relaxation like (25). We are looking for a function G satisfying $\int G \boldsymbol{m}\, d\mathbf{v} = \boldsymbol{\rho}$ together with other properties that are summarized here.

- Nonnegativess: the Duhamel formulation of the kinetic equation (7) shows that f remains nonnegative when G is also nonnegative.
- Entropy and convexity: the natural property related to an entropy defined as $\mathcal{H}(f) = \int \eta(f)\, d\mathbf{v}$ is convexity. If it holds the H-theorem is satisfied as soon as $\mathcal{H}(G) \leq \mathcal{H}(f)$ with equality iff $\mathcal{H}(G) = \mathcal{H}(\mathcal{M})$ since

$$\int \partial_f \eta'(f)\, K(f)\, d\mathbf{v} \leq \nu\left(\mathcal{H}(G) - \mathcal{H}(f)\right).$$

  This suggests that G should minimize $\mathcal{H}(h)$ under the moment constraints $\int h \boldsymbol{m} d\mathbf{v} = \boldsymbol{\rho}$. This also requires that when $\boldsymbol{\rho}$ are the moment of the local Maxwellian function then the minimum is obtained for $G = \mathcal{M}$.
- Analytic form: There is a need to to identify as much us possible and the exact shape of G. If we formally apply Kuhn and Tucker theorem for constrained optimization then there is

$$\eta'(G) = \sum_i \alpha_i[\boldsymbol{\rho}] \cdot m_i(\mathbf{v}) = \boldsymbol{\alpha} \cdot \mathbf{m}(\mathbf{v}),$$

where $(\alpha_i)_i$ are the Lagrange multipliers related to the constraints $\int G\boldsymbol{m}\, d\mathbf{v} = \boldsymbol{\rho}$. Moreover, if η is a strictly convex functions, then $(\eta')^{-1} = \eta^{*\prime}$ where η^* is the Legendre dual function to η, and there holds formally:

$$G(\mathbf{v}) = \eta^{*\prime}(\boldsymbol{\alpha} \cdot \mathbf{m}(\mathbf{v})). \tag{44}$$

This analytic expression provides us with a relation between $\boldsymbol{\alpha}$ and $\boldsymbol{\rho}$ through the moment constraints. $\boldsymbol{\alpha}$ is then found either explicitly or using a Newton algorithm depending on the choice of η.

Let us start with a simple example. Assume that the basis $\mathbf{m}(\mathbf{v})$ is orthonormal for the $L^2(\mathcal{M})$ scalar product. Then for $\boldsymbol{\rho} \in \mathcal{R}^+_{\mathbf{m}}$ we set

$$G(v) = \mathcal{M} \sum \rho_i m_i \tag{45}$$

which naturally satisfies $\int G\mathbf{m}d\mathbf{v} = \boldsymbol{\rho}$. $\sum \rho_i m_i \in L^2(\mathcal{M})$ is the classical orthogonal projection onto $\mathbb{P} = \mathtt{span}(\boldsymbol{m})$ of any $h \in L^2(\mathcal{M})$ satisfying $\int \mathbf{m}\, h\, \mathcal{M}\, d\mathbf{v} = \boldsymbol{\rho}$. It is also well defined for those functions $h = f/\mathcal{M} \notin L^2(\mathcal{M})$ as soon as $\int \mathbf{m}\, f\, d\mathbf{v} = \boldsymbol{\rho}$ and there is

$$\|\boldsymbol{\rho}\|^2 = \|\sum \rho_i m_i\|^2_{L^2(\mathcal{M})} = \mathtt{inf}_{D(\boldsymbol{\rho})} \int \mathcal{M}(f/\mathcal{M})^2 d\mathbf{v}$$

with $D(\boldsymbol{\rho}) = \{f / \int f\mathbf{m} = \boldsymbol{\rho}\}$.
Finally the function $\mathcal{H}(f) = \int \mathcal{M}(f/\mathcal{M})^2 d\mathbf{v}$ is an entropy for the relaxation operator if

$$\int G(\mathbf{1},\, \mathbf{v},\, \mathbf{v}^2) d\mathbf{v} = \int f(\mathbf{1},\, \mathbf{v},\, \mathbf{v}^2) d\mathbf{v}$$

and

$$\int G m_i d\mathbf{v} = \lambda_i \int f m_i d\mathbf{v}$$

for any $0 \leq \lambda_i < 1$. The only problem is that G has no sign!

In what follows, we intent to clarify the above consideration. Firstly by analyzing known results on this topic and then by stating a theorem proving that an equation similar to (44) holds.

4.1 The framework of convex analysis

Apart from the $\mathbb{L}^2(\mathcal{M})$ case which is quite simple but does not provide us with a non-negative solution, one should come back to the general setting of variational problem under linear constraints (27) and to the earlier work of Borwein and Lewis [8]. While their article addresses a wide class of problems, one may outline some reasonable conditions for the problem to be solved. We display them as follows:

1. $dom(\eta) = [0, +\infty[$, in order to have nonnegative solutions. η is convex and even strictly convex in order to ensure uniqueness of the solution if it exists. η is proper and closed and finally η is super linear at infinity.
2. In this framework, there always hold the following: There exists a weak topology for which the functional $g \mapsto \int \eta(g)\, d\mathbf{v}$ is semi-lower continuous and which is suited to the continuity of the constraints with respect to g.
3. The next condition may stated in simple words as follows. There exists a feasible function g, that is a function satisfying at the same time $\int \eta(g)\, d\mathbf{v} \in \mathbb{R}$ and $\int g\, \boldsymbol{m}\, d\mathbf{v} = \boldsymbol{\rho}$.

With this set of conditions, it may be proved that (44) is satisfied. As we will see below, the second condition is hardly compatible with kinetic theory, especially when one considers the entropy functional: $g \mapsto \int g \ln(g)\, d\mathbf{v}$ in an unbounded domain. We know want to make use of the simple analysis of the $L^2(\mathcal{M})$ case, the above framework of Borwein and Lewis and an other approach by Csiszar.

4.2 ϕ-divergence

In 1970, Csiszar introduced general distances between measures defined in the following way [18].

Definition 6. *Let ϕ with domain on $[0, +\infty[$, which is strictly convex at $x = 1$. The ϕ-divergence of two distribution function is defined as*

$$I(p\|q) = \int_{\mathbb{R}^3} q(\mathbf{v})\phi(\frac{p(\mathbf{v})}{q(\mathbf{v})})d\mathbf{v}.$$

This definition was first intended to generalize Shannon's entropy. When $\phi(x) = x\ln(x)$ it is known as the Kullback-Leibler divergence between p and q or relative entropy of p with respect to q. This definition allows to consider more general functions ϕ with the assumption that $\phi(1) = \phi'(1) = 0$ and $\phi''(1) \geq 0$. However $\int f \ln(f) d\mathbf{v}$ is in principle the only meaningful entropy for kinetic equations.

We are now interested in solving a variational problem when $q = \mathcal{M}$.

Definition 7 (Entropy and primal problem)**.** *Let ϕ be a ϕ divergence. Let $\boldsymbol{\rho} \in \mathbb{R}^q$ and consider the convex domain*

$$D(\boldsymbol{\rho}) = \left\{ g \in \mathbb{L}^1(\mathbf{m}),\ \int \mathbf{m}\, g = \boldsymbol{\rho} \right\}. \tag{46}$$

Define the entropy as $\mathcal{H}(f) = I(f\|\overline{\mathcal{M}})$ and the real extended value function $h\colon \mathbb{R}^q \to \overline{\mathbb{R}}$ for any $\boldsymbol{\rho} \in \mathbb{R}^q$ by

$$h(\boldsymbol{\rho}) = inf_{g \in D(\boldsymbol{\rho})} \mathcal{H}(g).$$

The primal problem consists in finding if possible a function G s.t.
1) $G \in D(\boldsymbol{\rho})$
2) $\mathcal{H}(G) = h(\boldsymbol{\rho})$

Thus solving the primal problem requires firstly to define the domain of h and secondly to find (in the sense of existence) a function satisfying 1) and 2).

4.2.1 Analysis of the case $\phi(x) = x\ln(x)$

It has been analyzed in series of papers and we want to summarize here some key point related to that case. Remark first that changing the Lebesgue measure with $\mathcal{M}\,d\mathbf{v}$ the problem is almost equivalent to the class of problem that were addressed in Borwein and Lewis. Also ϕ is easily seen to satisfy the first set of conditions of subsection 4.1. The condition 3 is addressed in the following proposition.

Proposition 8. *The domain of h is $\mathcal{R}^+_{\mathbf{m}}$.*

Proof. $(\forall f \geq 0),\ f\ln(f/\mathcal{M}) \geq -\frac{1}{e}\mathcal{M}$ as a simple consequence of the inequality $x\ln(x) \geq -\frac{1}{e}$. Thus $(\forall f \geq 0),\ \mathcal{H}(f) \geq -\frac{n}{e}$.
Next, $\forall \boldsymbol{\rho} \in \mathcal{R}^+_{\mathbf{m}}$, $\exists \Psi_{\boldsymbol{\rho}} \in \mathcal{C}^\infty_c(\mathbb{R}^3)$ such that $\Psi_{\boldsymbol{\rho}} \in D(\boldsymbol{\rho})$. Hence $f/\mathcal{M}$ is compactly supported, bounded and $\mathcal{H}(\Psi_{\boldsymbol{\rho}}) < +\infty$. And there is

$$D^+(\boldsymbol{\rho}) = \{g \text{ s.t. } \int_{\mathbb{R}^3} \mathbf{m}\, g\, d\mathbf{v} = \boldsymbol{\rho} \text{ and } \mathcal{H}(g) < +\infty\} \neq \emptyset.$$

As a consequence $h(\boldsymbol{\rho})$ is well defined for all $\boldsymbol{\rho} \in \mathcal{R}^+_{\mathbf{m}}$.
Assume now that $\boldsymbol{\rho} \notin \mathcal{R}^+_{\mathbf{m}}$. Then for any f s.t. $\int f\mathbf{m}\, d\mathbf{v} = \boldsymbol{\rho}$, there is an open set ω_f of non zero measure s.t. $f < 0$ on ω_f. ϕ being equal to $+\infty$ when $x < 0$, we have $h(\boldsymbol{\rho}) = +\infty$. □

Utilizing theorem 2, for any $\boldsymbol{\rho} \in \mathcal{R}^+_{\mathbf{m}}$ we may restrict $D^+(\boldsymbol{\rho})$ to

$$D^+_\Psi(\boldsymbol{\rho}) = \{g \text{ s.t. } \int_{\mathbb{R}^3} \mathbf{m}\, g\, d\mathbf{v} = \boldsymbol{\rho} \text{ and } \mathcal{H}(g) \leq \mathcal{H}(\Psi_{\boldsymbol{\rho}})\} \neq \emptyset.$$

The superlinearity of ϕ together with the boundedness of moments of order more than 1 show that D^+_Ψ is weakly relatively compact in $\mathbb{L}^1$ by Dunford-Pettis lemma. In principle, it is not possible to prove more than this and condition 2 of Borwein and Lewis is not satisfied.

As a consequence any minimizing sequence $f_n \in D^+(\boldsymbol{\rho})$ converges weakly in $\mathbb{L}^1$ to a function G but it is not sufficient to ensure that $\int G\mathbf{m}\, d\mathbf{v} = \boldsymbol{\rho}$. Junk has shown in a famous paper that the constraint of highest degree might drop in when looking at the infimum in $D^+(\boldsymbol{\rho})$. We want here to give a rapid hint into that problem. Consider the dual function h^* of h defined on its domain Λ by

$$h^*(\boldsymbol{\alpha}) = \int \exp(\boldsymbol{\alpha}\cdot\boldsymbol{m})d\mathbf{v}$$

and assume that $\Lambda \cap \partial\Lambda \neq \emptyset$. For $\boldsymbol{\alpha} \in \Lambda \cap \partial\Lambda$. Clearly h^* has only sided derivative at $\boldsymbol{\alpha}$ in the directions pointing into the domain and there is only a subdifferential at $\boldsymbol{\alpha}$. One can prove the following. Firstly, each moment

$$\boldsymbol{\rho}_+ = (\rho_0,\, \rho_1,\, \ldots,\, \rho_q + t),\ t > 0,$$

where

$$\boldsymbol{\rho} = \int \boldsymbol{m}(v)\, \exp(\boldsymbol{\alpha}\cdot\boldsymbol{m}) d\mathbf{v}$$

belongs to $\mathcal{R}^+_{\mathbf{m}}$ when $\boldsymbol{m} = \{\mathbf{1}, \mathbf{v}, \mathbf{v}^2, \ldots, |\mathbf{v}|^N\}$ (here N is the maximal degree of the component). Indeed, $\boldsymbol{\rho}$ being a realizable moment, the necessary and sufficient condition $\boldsymbol{\rho}\cdot\boldsymbol{\alpha} > 0$ for any positive polynomial implies $\boldsymbol{\rho}_+\cdot\boldsymbol{\alpha} > 0$ for any $\boldsymbol{\rho}_+$. One can then prove that [26, 35]

$$h^{**}(\boldsymbol{\rho}_+) = \max_{\tilde{\boldsymbol{\alpha}}}\{\tilde{\boldsymbol{\alpha}}\cdot\boldsymbol{\rho}_+ - h^*(\tilde{\boldsymbol{\alpha}})\} = \boldsymbol{\alpha}\cdot\boldsymbol{\rho} - h^*(\boldsymbol{\alpha}) = h(\boldsymbol{\rho}).$$

h being semi lower continuous in $\mathcal{R}^{+*}_{\mathbf{m}}$ (implying $h^{**} = h$) this proves that the subdifferential $\partial h^*(\boldsymbol{\alpha})$ is the whole half-line $\boldsymbol{\rho}_+$. Therefore $\inf_{g\in D(\boldsymbol{\rho}_+)} \mathcal{H}(g) = h(\boldsymbol{\rho})$ is attained at the function $\exp(\boldsymbol{\alpha}\cdot\boldsymbol{m}) \notin D(\boldsymbol{\rho}_+)$. To summarize, the existence of a solution to the primal problem is subjected to the shape of the domain of definition of the dual function h^*. However, the problem can still be seen as a problem of approximation theory.

Remark 2. *We have seen that under suitable assumption on ϕ, any minimizing sequence $f_n \in D^+(\boldsymbol{\rho})$ converges weakly in $\mathbb{L}^1$ to a function G which satisfies*

$$\mathcal{H}(f) = \inf_{f\in D(\boldsymbol{\rho})} \mathcal{H}(f) = \inf_{f\in D(\boldsymbol{\rho})} I(f|\mathcal{M})$$

Csiszar has defined the later value as the "distance"' $d(\mathcal{M}, D(\boldsymbol{\rho}))$ between $\mathcal{M}$ and $D(\boldsymbol{\rho})$. G was called later on the generalized projection of $\mathcal{M}$ on $D(\boldsymbol{\rho})$ [20]. If $G \in D(\boldsymbol{\rho})$, G is named projection and it is the solution to the primal problem.
d is not a metric but it plays the role of the square Euclidean distance when $\phi(x) = x\ln x$ and there is Pythagorus theorem

$$(\forall f \in D(\boldsymbol{\rho})), \qquad d(\mathcal{M}, f) = d(\mathcal{M}, G) + d(G, f).$$

The question that will be addressed in the next section is when does $d(\mathcal{M}, f_n) \to d(\mathcal{M}, G)$ imply weak convergence of f_n to G in $\mathbb{L}^1(\mathbf{m})$.

4.2.2 Csiszar assumption

In a remarkable paper, Csizsár [20] has shown existence of generalized projection onto convex subset D of some finite measure space for a wide class of problem. If we restrict ourselves to the case of constraints like $D(\boldsymbol{\rho})$, sufficient assumptions on ϕ to the existence of generalized projection are: ϕ is a strictly convex differentiable function defined on $]0, +\infty[$,

$$\phi(1) = \phi'(1) = 0, \qquad \lim_{p\to+\infty} \phi'(p) = +\infty. \tag{47}$$

When one turns to the problem of the existence of a projection, one needs an additional assumption. Let us assume for simplicity that the condition at $x = 1$ occurs at $x = 0$.

The idea consists in considering the Orlicz (Banach) space L_ϕ whose norm is related (but not equal) to $\mathcal{H}(|g|)$ and its dual space $L_\phi^* = L_{\phi^*}$. If the dual space contains $\mathbb{P} = \mathtt{span}(\boldsymbol{m})$, a theorem states that any bounded sequence g_n in Orlicz morm which converges in measure to a given measure G satisfies on one hand $G \in L_\phi$ and on the other hand converges $\mathbb{P}$ weakly, that is $\int g_n\, p(\mathbf{v})\, d\mathbf{v} \to \int G\, p(\mathbf{v})\, d\mathbf{v}\ \forall p \in \mathbb{P}$. So, after one has proved that G is the generalized projection of $\mathcal{M}$ on $D(\boldsymbol{\rho})$, this proves that $G \in D(\boldsymbol{\rho})$. In other words G is the (unique) solution to the primal problem.

4.3 Assumption and main result

Without too much restriction, Csiszar assumption's can be restated as follows.

Definition 8 (ϕ-divergence functions)**.** *In this article, we assume that ϕ has the following properties:*

1. *The function $\phi : \mathbb{R} \mapsto \mathbb{R} \bigcup \{+\infty\}$ is strictly convex on its domain $\mathtt{dom}(\phi) = [0, +\infty)$ and differentiable on $[0, +\infty)$*
2. *There holds the following properties:*

$$\phi(0) = 0, \quad p_0 := \inf_{y>0} \frac{\phi(y)}{y} \in \mathbb{R}, \quad \sup_{y>0} \frac{\phi(y)}{y} = +\infty$$

3. *For any polynomial $\pi(\mathbf{v}) = \boldsymbol{\gamma} \cdot \mathbf{m}(\mathbf{v})$ then $\phi^*(\pi) \in \mathbb{L}^1(\mathcal{M}(\mathbf{v})\, d\mathbf{v})$ where ϕ^* is the Legendre transform of ϕ*

Let us address the following remarks:

1. From the first assumption, $\mathtt{dom}(\phi) = [0, +\infty)$ enables to grant that the primal problem will have solutions that will be non negative. Also, it is quite natural to assume that 0 is in the domain of ϕ unless the p.d.f 0 cannot appear as a solution. So we will be able to produce non negative modeling function for the BGK construction. Differentiation on $[0, +\infty)$ implies semi-lower continuity on $[0, +\infty)$ (which is essential in convex optimization) and will enable one to one property of ϕ'. Strict convexity will have, as often, many implications in uniqueness consideration.
2. The third assumption is essential for integrations properties. In particular, it breaks when $\phi(x) = x \ln(x)$ since there holds then $\phi^*(y) = \exp(y-1)$ and only few polynomial $\exp(\pi(\mathbf{v}) - 1)$ have proper integration properties. This is one of the most difficult aspect of the Levermore program which was based first on a function like $x \ln(x)$.

As usual in convex analysis, the theorem that will be established is based on the analysis of the primal problem and of the dual problem.

Problem 1. *The dual problem consists in defining the Legendre dual function $h_{\boldsymbol{m}}^*$ from $\mathbb{R}^q$ to the extended reals $\overline{\mathbb{R}}$ - that is $\mathbb{R} \bigcup \{+\infty, -\infty\}$- as follows:*

$$\forall \boldsymbol{\alpha} \in \mathbb{R}^q, \quad h_{\boldsymbol{m}}^*(\boldsymbol{\alpha}) = \sup_{\boldsymbol{\rho} \in \mathbb{R}^q} \left(\boldsymbol{\rho} \cdot \boldsymbol{\alpha} - h_{\mathbf{m}}(\boldsymbol{\rho})\right). \tag{48}$$

The theorem that we are going to prove is the following

Theorem 9. *1. For any $\boldsymbol{\rho} \in \mathcal{R}_{\mathbf{m}}^{+*}$ there exists a unique $\boldsymbol{\alpha} \in \mathbb{R}^q$ such that*

$$\boldsymbol{\rho} = \int \phi^{*\prime} \left(\boldsymbol{\alpha} \cdot \mathbf{m} \left(\mathbf{v} \right) \right) \mathbf{m} \left(\mathbf{v} \right) \mathcal{M} d\mathbf{v}$$

and the moments $\boldsymbol{\rho}$ and its conjugate moment $\boldsymbol{\alpha}$ are linked thanks to the sub-differential equation:

$$h_{\mathbf{m}} \left(\boldsymbol{\rho} \right) + h_{\mathbf{m}}^{*} \left(\boldsymbol{\alpha} \right) = \boldsymbol{\alpha} \cdot \boldsymbol{\rho} \tag{49}$$

2. Moreover the function $G = \mathcal{M} \phi^{*\prime} \left(\boldsymbol{\alpha} \cdot \mathbf{m} \left(\mathbf{v} \right) \right)$ is the unique minimizer of the primal problem and satisfies

$$h_{\mathbf{m}} \left(\boldsymbol{\rho} \right) = \int \phi \left(\frac{G}{\mathcal{M}} \right) \mathcal{M} d\mathbf{v} \tag{50}$$

3. $h_{\mathbf{m}}$ is strictly convex in its domain and ∇h^{*} is a bijection from C° to $\mathcal{R}_{\mathbf{m}}^{+} \setminus \{0\}$ where

$$C^{\circ} = \left\{ \boldsymbol{\alpha} \in \mathbb{R}^q, \quad \boldsymbol{\alpha} \cdot \mathbf{m}(\mathbf{v}) > p_0, \text{ on a set of } \omega_{\boldsymbol{\alpha}} \text{ of non 0 measure} \right\}. \tag{51}$$

The proof goes along different steps we are going to develop. The first one consists in generalizing Proposition 8 concerning the primal problem.

Proposition 10. *The following properties are satisfied*

1. *The moment entropy function $h_{\mathbf{m}} \left(\boldsymbol{\rho} \right)$ has returned values in $\mathbb{R} \bigcup \{+\infty\}$*
2. *Its domain is exactly $\mathtt{dom}(h_{\mathbf{m}}) = \mathcal{R}_{\mathbf{m}}^{+} = \left\{ \int \mathbf{m} g, g \in \mathbb{L}^{1+} \left(\mathbf{m} \right) \right\}$.*
3. *The interior of the domain is given by $\mathtt{int} \left(\mathtt{dom}(h_{\mathbf{m}}) \right) = \mathcal{R}_{\mathbf{m}}^{+*} = \mathtt{dom}(h_{\mathbf{m}}) \setminus \{0\}$*
4. *The function $h_{\mathbf{m}}$ is convex.*

From this we may show using Dunford-Pettis Theorem that there exists a generalized projection in the sense of Remark 2. However, it is not sufficient to prove on one hand that G is a projection and on the other hand to exhibit the shape of the solution. Both can be proved by using duality arguments. We start with some properties of the Legendre conjugate ϕ^{*} of ϕ.

Lemma 2. *Let ϕ be a ϕ-divergence function. Then the following properties hold*

1. *There holds $\forall x \in (-\infty, p_0], \quad \phi^{*} \left(x \right) = 0$*
2. *There holds $\forall x \in (p_0, +\infty), \quad \phi^{*} \left(x \right) > 0$*
3. *The function $\phi^{*} : \mathbb{R} \mapsto \mathbb{R}^{+}$ is C^1 smooth and thus semi-lower continuous.*

The link between strict convexity of a given function and smoothness of its conjugate is well known in convex analysis (see [38]). But we prefer to detail the whole proof of this lemma for the sake of consistency. Smoothness of the function ϕ^{*} will have many important implications on the smoothness of the dual function h^{*}, which, in turn, will have important consequences on the primal minimization problem and on existence and uniqueness of the minimizer.

At this point, it is important to recall the noteworthy example of Abdelmalik and Van Brummelen [2].

Example 2. *For $N \in \mathbb{N}$, define ϕ_N with*

$$\forall x < 0, \quad \phi_N(x) = +\infty, \quad \forall x \geq 0, \quad \phi_N(x) = N\left(x^{1+1/N} - x\right)$$

is a ϕ-divergence function. Moreover, there holds respectively

$$p_0(N) = -N, \ \forall y \geq p_0(N), \ \phi_N^*(y) = \left(1 + \frac{y-1}{N+1}\right)^{N+1}$$

It is really worth noticing the following pointwise convergence aspect:

$$\forall x \geq 0, \quad \lim_N \phi_N(x) = x\ln(x), \quad \forall y \in \mathbb{R}, \quad \lim_N \phi_N^*(y) = \exp(y-1)$$

that is, the entropy ϕ_N is a pointwise approximation of the Boltzmann entropy density $x\ln(x)$ and ϕ_N^* is an approximation of $\exp(y-1)$ which is the Legendre transform of $x\ln(x)$. Moreover we have the important limits:

$$\forall x \in \mathbb{R}^{*+}, \ \partial_x\phi_N(x) = (N+1)x^{1/N} - N, \ \forall y \geq -N, \quad \partial_y\phi_N^*(y) = \left(1 + \frac{y-1}{N+1}\right)^N$$
$$\forall x \in \mathbb{R}^{*+}, \quad \lim_N \partial_x\phi_N(x) = \ln(x) + 1, \ \forall y \in \mathbb{R} \ \lim_N \partial_y\phi_N^*(y) = \exp(y-1)$$

This means that the derivatives of ϕ_N and ϕ_N^* are respectively good approximation of the $\partial_x(x\ln x)$ and $\partial_y(\exp(1-y))$.

We can now give an explicit expression of the dual function h^* and analyze its properties.

Theorem 11. *1. For any $\boldsymbol{\alpha} \in \mathbb{R}^q$ there is*

$$h_{\mathbf{m}}^*(\boldsymbol{\alpha}) = \sup_{g \in \mathbb{L}^1(\mathbf{m})} \int \left[\boldsymbol{\alpha} \cdot \mathbf{m}(\mathbf{v})\, g(\mathbf{v}) - \phi\left(\frac{g}{\mathcal{M}}\right)\mathcal{M}(\mathbf{v})\right] d\mathbf{v} \in \mathbb{R}. \tag{52}$$

2. Moreover one can also compute for any $\boldsymbol{\alpha} \in \mathbb{R}^q$

$$h_{\mathbf{m}}^*(\boldsymbol{\alpha}) = \int \phi^*(\boldsymbol{\alpha} \cdot \mathbf{m}(\mathbf{v}))\, \mathcal{M}(\mathbf{v})\, d\mathbf{v}. \tag{53}$$

3. The function $h_{\mathbf{m}}^*$ is continuously differentiable on $\mathbb{R}^q$ and there holds for any $\boldsymbol{\alpha} \in \mathbb{R}^q$

$$h_{\mathbf{m}}^{*\prime}(\boldsymbol{\alpha}) = \int \phi^{*\prime}(\boldsymbol{\alpha} \cdot \mathbf{m}(\mathbf{v}))\, \mathbf{m}(\mathbf{v})\, \mathcal{M}(\mathbf{v})\, d\mathbf{v}. \tag{54}$$

Now we are able to finish the proof of Theorem 9. It is based on important properties. In few words, it is based on well known properties of convex analysis.

1. Any convex function $f : \mathbb{R}^q \to \mathbb{R}\bigcup\{+\infty\}$ is continuous on the interior of its domain
2. Any convex function $f : X \to \mathbb{R}\bigcup\{+\infty\}$ which is continuous at $\mathbf{x} \in \mathtt{dom}(f)$ has a non void sub-differential at $\mathbf{x}$, that is $\partial f(\mathbf{v})) \neq \emptyset$.
3. For a proper convex function f, closed at $\boldsymbol{\rho}$, there is $\boldsymbol{\alpha} \in \partial f(\boldsymbol{\rho}) \Leftrightarrow \boldsymbol{\rho} \in \partial f^*(\boldsymbol{\alpha})$.

5 Application to the construction of BGK models using ϕ-divergence

5.1 Relaxation on the Grad thirteen moments

Let $f(t,x,v)$ be a nonnegative function at (t,x). Denote with n, $\mathbf{u}$ and T the corresponding density, velocity temperature and with $\mathcal{M}$ the local Maxwellian associated to f. Let $\mathbf{a}(\mathbf{v}-\mathbf{u})$ be the local Grad basis and denote

$$\boldsymbol{\rho}_f = (n, \mathbf{0}, 0, n\Pi, n\mathbf{Q}) \tag{55}$$

$$= \int \left(\mathbf{1}, (\mathbf{v}-\mathbf{u}), (\mathbf{v}-\mathbf{u})^2 - 3\frac{k_B T}{m}, \mathbb{A}(\mathbf{v}-\mathbf{u}), \mathbf{b}(\mathbf{v}-\mathbf{u})\right) f(\mathbf{v})\, d\mathbf{v}. \tag{56}$$

Remark that there is by definition $\boldsymbol{\rho}_{\mathcal{M}} = (n, \mathbf{0}, 0, 0, 0)$.

5.1.1 Principle of construction

We just recall here the steps in the derivation of a relaxation operator in the framework of example 1.

1. The relaxation frequencies $\nu_{\mathbb{A}}$ and $\nu_{\mathbf{b}}$ being defined in (18), one may take any value for ν with the condition $\nu > \nu_{\mathbb{A}}, \nu_{\mathbf{b}}$.
2. The relaxed moment $\mathtt{L}(\boldsymbol{\rho}_f)$ writes

$$\mathtt{L}(\boldsymbol{\rho}_f) = (n, \mathbf{0}, 0, \lambda_{\mathbb{A}} n\mathbb{A}, \lambda_{\mathbf{b}} n\mathbf{b}), \quad \lambda_{\mathbb{A}} = 1 - \frac{\nu_{\mathbb{A}}}{\nu}, \quad \lambda_{\mathbf{b}} = 1 - \frac{\nu_{\mathbf{b}}}{\nu}. \tag{57}$$

   With the above value of ν, $\mathtt{L}(\boldsymbol{\rho}_f)$ is still realizable from proposition 5
3. Choose a ϕ-divergence function satisfying the properties of Definition 8. Then replace in Theorem 9, $\boldsymbol{\rho}$ with $L(\boldsymbol{\rho}_f)$. Then G is defined as $\mathcal{M}(\phi^*)'(\boldsymbol{\alpha}\cdot\boldsymbol{m})$
4. The BGK operator reads as $K(f) = \nu(G - f)$.

5.1.2 Properties of the model

Remark firstly that $G = \mathcal{M}(\phi^*)'(\boldsymbol{\alpha}\cdot\boldsymbol{m})$ is nonnegative. Thus the solution f to (7), if it exists, is nonnegative as well. Next the Grad space satisfies the necessary conditions of proposition 7 as concerns Galilean invariance of the modeling equation (7). Then Galilean invariance holds according to

Proposition 12. *For $\tau \in \{\tau_{\mathbf{u}};\, \tau_\theta\}$ $\tau G(f) = G(\tau f)$*

Let us now prove the (full) H-theorem.

Theorem 13. *[H-theorem] Recall that*

$$\mathcal{H}(f) = \int_{\mathbb{R}^3} \mathcal{M}\phi(f/\mathcal{M})\, d\mathbf{v}.$$

Then there hold

$$\forall f \geq 0 \in \mathbb{L}^1(\mathbf{a}), \quad \left\langle K(f)\, \partial_x \phi\left(\frac{f}{\mathcal{M}}\right)\right\rangle \leq 0.$$

together with the characterization of equilibrium

$$K(f) = 0 \Leftrightarrow \left\langle K(f)\, \partial_x \phi \left(\frac{f}{\mathcal{M}} \right) \right\rangle = 0 \Leftrightarrow f = \mathcal{M}.$$

It must be denoted that the condition $\nu > \nu_{\mathbb{A}}, \nu_{\mathbf{b}}$ is necessary in order to obtain the above results. This may be easily understood if one considers the relations between the moments of G and those of f (24) which only equate at $\boldsymbol{\rho}_G = \boldsymbol{\rho}_f = \boldsymbol{\rho}_{\mathcal{M}}$. Let us now derive proposition 1 in the case of relaxation in Grad space together with other properties stated in section 2.1 (the proof easily extends to the general case).

Proposition 14. *For $K(f)$ derived in section 5.1.1, the linearized operator $\mathcal{L}$ defined in (6) reads as*

$$\mathcal{L}(g) = \nu \left(\mathcal{P}_{\mathbb{K}} - \mathcal{I} + \left(1 - \frac{\nu_{\mathbb{A}}}{\nu}\right) \mathcal{P}_{\mathbb{A}} + \left(1 - \frac{\nu_{\mathbf{b}}}{\nu}\right) \mathcal{P}_{\mathbf{b}} \right). \tag{58}$$

As a consequence there holds

1. *The kernel of the operator $\mathcal{L}$ is exactly $\mathbb{K}$ and there is also*

$$\forall f, \quad \left[\int K(f)\, \phi = 0 \right] \Leftrightarrow \phi \in \mathbb{K}$$

2. *The operator is Fredholm, self-adjoint and negative on $\mathbb{K}^{\perp}$*
3. *The diffusion coefficients in the Navier-Stokes limit of (7) are given by (18).*

Remark that the second item is easily seen from (58). The third one is already proved in example 1.

5.2 The general case

We now consider a polynomial space $\mathbb{P}$ satisfying the condition in proposition 7 and containing $\mathbb{P}_{Grad} \subset \mathbb{P}$. $\mathbb{P}$ being invariant under translations and rotations, we may write

$$\mathbb{P} = \mathbb{K} \oplus^{\perp} m_6 \oplus^{\perp} \cdots \oplus^{\perp} m_q. \tag{59}$$

where the polynomials $(m_i)_i$ are defined in (9). If the collision invariants are written in an orthogonal basis for the $L^2(\mathcal{M})$ dot product, then the vectors $\boldsymbol{\rho}_f$, $\boldsymbol{\rho}_{\mathcal{M}}$ and $\boldsymbol{\rho}_G = L(\boldsymbol{\rho}_f)$ in the above decomposition read respectively

$$\begin{aligned}
\boldsymbol{\rho}_f &= (n, \mathbf{0}, 0, \rho_6, \cdots, \rho_q)^T, \\
\boldsymbol{\rho}_{\mathcal{M}} &= (n, \mathbf{0}, 0, 0, \cdots, 0)^T, \\
\boldsymbol{\rho}_G &= \left(n, \mathbf{0}, 0, (1 - \frac{\nu_6}{\nu})\rho_6, \cdots, (1 - \frac{\nu_q}{\nu})\rho_q\right)^T.
\end{aligned}$$

The criteria of realizability of $\boldsymbol{\rho}_G$ through symmetric positive definite moment matrix (see section 3.3) is unsatisfactory for two reasons : 1 - if $\mathbb{P}$ is a quadratic space (definition 5), it is difficult to express the moment matrix corresponding to $\boldsymbol{\rho}_G$ without

knowing explicitly the eigenfunctions $(m_i)_i$ and the corresponding eigenvalues $(\nu_i)_i$, 2 - even if it would be possible this criteria is not sufficient if all positive polynomials in $\mathbb{P}$ are not sum of square polynomials.
So, for want of anything better we make the following assumption : the solution to (7) is such that the ball $B(\boldsymbol{\rho}_{\mathcal{M}}, r)$ of radius $r = ||\boldsymbol{\rho}_f - \boldsymbol{\rho}_{\mathcal{M}}||$ stays in $\mathcal{R}^{*+}_{\mathbf{m}}$. In this case, it is easily seen that $\boldsymbol{\rho}_G \in \mathcal{R}^{*+}_{\mathbf{m}}$ if $\nu > \nu_i,\ \forall i$.
Let us consider again a ϕ-divergence function satisfying the properties of Definition 8. The question whether

$$\int K(f)\partial_f\phi\left(\frac{f}{\mathcal{M}}\right) d\mathbf{v} \leq 0 \tag{60}$$

holds or not is for the moment an open problem. However the characterization of local equilibrium is easily found since $K(f) = 0$ if and only if $\boldsymbol{\rho}_f = \boldsymbol{\rho}_G$ which occurs only at $G = \mathcal{M}$. All other properties in the preceding section are satisfied. In particular, in the Chapman-Enskog expansion, one still finds that the solution satisfies the Euler equation in $O(\varepsilon)$ while the Navier-Stokes equation is obtained with the right viscosity and heat conductivity just by using the definition of (9) in (18).

5.3 Some known models

5.3.1 BGK and ESBGK models

Let $\phi(x) = x\ln(x)$. If one just considers the conservation laws (10) together with the relaxation equation (20), then the variational problem is well-posed for $-\frac{1}{2} \leq 1 - \frac{\nu_{\mathbb{A}}}{\nu} \leq 1$. Indeed on one hand the of constraints (26) is non empty according to proposition 5 and on the other hand the domain of h^* is non empty and open [30].
$\nu = \nu_{\mathbb{A}}$ with $\nu_{\mathbb{A}}$ defined in (18) gives the well-known BGK operator [7]. Remark in this case that (21) is satisfied for $\nu_{\mathbf{b},BGK} = \nu$. So while the right viscosity is recovered in the hydrodynamic limit, the heat conductivity κ_{BGK} is such that the Prandtl number

$$Pr = \frac{5}{2}R\frac{\mu_B}{\kappa_{BGK}} = \frac{\nu_{\mathbf{b},BGK}}{\nu_{\mathbb{A}}} = 1.$$

More generally, for $0 \leq \frac{\nu_{\mathbb{A}}}{\nu} \leq \frac{3}{2}$, the solution to the variational problem always satisfies $\int G\,\mathbf{b}\,(\mathbf{v}-\mathbf{u})\,d\mathbf{v} = 0$. The ESBGK model is then found by letting $\nu = \nu_{\mathbf{b}}$ with $\nu_{\mathbf{b}}$ defined in (18) which corresponds to the limit $\frac{\nu_{\mathbb{A}}}{\nu} = \frac{3}{2}$ and $Pr = \frac{2}{3}$.

5.3.2 Shakhov model

Let now $\phi(x) = \frac{1}{2}(x-1)^2$. In the Grad space, one considers the system (10, 20, 21). Assume that μ_B and κ_B are either given by the exact computations in (18, 19) or by using some approximations of them. Let $\nu = \nu_{\mathbb{A}} = \frac{nk_BT}{\mu_B}$ and $\nu_{\mathbf{b}} = \frac{5}{2}\frac{nk_B^2T}{m\kappa_B}$. Remark that

$$Pr = \frac{5}{2}R\frac{\mu_B}{\kappa_B} = \frac{\nu_{\mathbf{b}}}{\nu_{\mathbb{A}}} = \frac{\nu_{\mathbf{b}}}{\nu}.$$

Then the system (10, 20, 21) together with the minimization problem give

$$G_S = \mathcal{M}\left(1 + \frac{1-Pr}{5}\frac{m}{n(k_BT)^2}\mathbf{q}\cdot(\mathbf{v}-\mathbf{u})\left(m\frac{(\mathbf{v}-\mathbf{u})^2}{k_BT} - 5\right)\right),$$

where $\mathbf{q}$ is the heat flux defined by

$$\mathbf{q} = \frac{1}{2}M\int_{\mathbb{R}^3} f\,(\mathbf{v}-\mathbf{u})(\mathbf{v}-\mathbf{u})^{\mathbf{2}}\,d\mathbf{v}.$$

Originally G_S was computed in such a way that

$$\int_{\mathbb{R}^3} \mathbf{a}(\mathbf{v}-\mathbf{u})\nu(G_S - f)d\mathbf{v} = \int_{\mathbb{R}^3} \mathbf{a}(\mathbf{v}-\mathbf{u})Q(f,f)d\mathbf{v} \tag{61}$$

for Maxwell molecules and then adapted to other types of molecular interaction by introducing Pr into the definition of G_S. In the later case, the above equation is not valid.
The generalization through the diagonalization in (9) is easily performed by letting

$$G = \mathcal{M}\left(1 + \sum_i \left(1 - \frac{\nu_i}{\nu}\right)\mathcal{P}_{m_i}(g)\right),$$

where $g = f/\mathcal{M} - 1$ and $\nu > \nu_i$, $\forall i$. In both cases G is not nonnegative. However one must point out that $H(f) = \int \mathcal{M}\phi(f/\mathcal{M})\,d\mathbf{v}$ is the natural entropy related to the whole method since

$$\int_{\mathbb{R}^3} \nu(G_S - f)\phi'(\frac{f}{\mathcal{M}} - 1)\,d\mathbf{v} = \langle\mathcal{L}(g), g\rangle \leq 0$$

where $g = f/\mathcal{M} - 1$ and $\mathcal{L}$ is defined in (12). It might happen that $g \notin L^2(\mathcal{M})$ in which case the above value is $-\infty$. However $\langle\mathcal{L}(g), g\rangle = 0$ only for $g = 0$ or equivalently $f = \mathcal{M}$. Every other properties of section 2.1 are satisfied except the nonnegativness of G.

5.3.3 Levermore's operator

The analysis of the Chapman-Enskog expansion for moment system of the Boltzmann equation shows that wrong diffusion coefficients are obtained at the Navier-Stokes level [33]. Levermore has then proposed to substitute to the collision operator $Q(f,f)$ a sum of relaxation operators constructed as follows. Let $\mathbb{K} = \mathbb{M}_1 \subset \mathbb{M}_2 \subset \ldots \subset \mathbb{M}_N$ and $0 < \eta_1 < \eta_2 < \ldots < \eta_N$. Set

$$\mathcal{M}_k = Argmin\left\{\int g\ln(g) \,/\, \int gp(\mathbf{v})\,d\mathbf{v} = \int fp(\mathbf{v})\,d\mathbf{v},\ \forall p \in \mathbb{M}_k\right\}, \tag{62}$$

then $K_{Lev}(f)$ writes

$$K_{Lev}(f) = \eta_1(\mathcal{M} - f) + \sum_{k=2}^{N}(\eta_k - \eta_{k-1})(\mathcal{M}_k - f)$$

Due to the assumption on each relaxation frequencies ν_i, it is clear that $K_{Lev}(f)$ preserves positivity, together with conservation laws. Also $\int f \ln(f) d\mathbf{v}$ is the entropy in the non homogeneous equation (7). The linearized operator reads as

$$\mathcal{L}_{Lev} = -\sum_{k=1}^{N-1} \eta_k \left(\mathcal{P}_{k+1} - \mathcal{P}_k\right) + \eta_N \left(\mathcal{P}_N - \mathcal{I}\right)$$

where $\mathcal{P}_k$ is the orthogonal projection onto $\mathbb{M}_k$ in $L^2(\mathcal{M})$. Denoting with $(m_{i,k})_{i=1,\dots,D_k}$ an orthogonal basis of $\mathbb{M}_k \cap \mathbb{M}_{k-1}^{\perp}$, there is

$$\begin{aligned}
\mathcal{L}_{Lev} &= \eta_N \left(\mathcal{P}_N - \mathcal{I}\right) - \sum_{k=1}^{N-1} \eta_k \sum_{i=1}^{D_{k+1}} \mathcal{P}_{m_{i,k+1}} \\
&= \eta_N \left(\left(\mathcal{P}_{\mathbb{K}} - \mathcal{I}\right) + \sum_{k=1}^{N-1} \left(1 - \frac{\eta_k}{\eta_N}\right) \sum_{i=1}^{D_{k+1}} \mathcal{P}_{m_{i,k+1}} \right)
\end{aligned}$$

which has a form similar to (12). However there are many problems related to this construction. Junk was the first to point out that the solution to the variational problem might not satisfy all constraints [30]. Also, if $\mathcal{L}_{Lev} = \mathcal{L}$ in (12), the solution to the variational problem in (62) may not exist as shows the simple case of Maxwell molecules. Indeed, some spaces $\mathbb{M}_k$ have a maximal odd degree in which case there exists no solution in (62). Remark finally that $K_{Lev}(f)$ does not satisfy relaxation equations such as (11). Indeed, for $p \geq 3$ and $1 \leq i \leq D_p$, there is

$$\begin{aligned}
\int K_{Lev}(f)\, m_{i,p}\, d\mathbf{v} &= \eta_1 \int (\mathcal{M} - f)\, m_{i,p}\, d\mathbf{v} + \sum_{k=2}^{p-1} (\eta_k - \eta_{k-1}) \int (\mathcal{M}_k - f)\, m_{i,p}\, d\mathbf{v} \\
&= -\eta_{p-1} \int f\, m_{i,p}\, d\mathbf{v} + \sum_{k=2}^{p-1} (\eta_k - \eta_{k-1}) \int \mathcal{M}_k\, m_{i,p}\, d\mathbf{v}.
\end{aligned}$$

But for $2 \leq k \leq p-1$, $\int \mathcal{M}_k\, m_{i,p}\, d\mathbf{v}$ is not related to $\int f\, m_{i,p}\, d\mathbf{v}$ in the minimization problem (62). Also, it does not vanish except if $\mathcal{M}_k = \mathcal{M}\, q(\mathbf{v})$ for some polynomial $q(\mathbf{v}) \in \mathbb{M}_k$ in which case the functional to be minimized in (62) is the one of the previous section.
The minimization problem in (62) can be fixed by using ϕ-divergence as in definition 8 instead of $\phi(x) = x \ln x$ since then the solution exists whatever the parity of the highest degree of the polynomials in the constraints. The operator satisfies by construction (60) and the characterization (4) follows under the sufficient condition $\mathbb{M}_1 = \mathbb{K}$. Thus,

in the general case of section 5.2, the model is well-posed. Notice again that relations (11) still not hold so that it cannot be used in practice, especially in the context of moment systems for which it was originally designed.

6 Proofs

6.1 Proofs of the section 3

Proof. (Theorem 3) Let $x \in \texttt{int}(C)$. Then there exists $\varepsilon > 0$ s.t. $B(x, \varepsilon) \subset \texttt{int}(C)$. So $\forall y \in C^0,\ y \neq 0,\ x \cdot y \leq 0$. But if there exists $y \neq 0$ s.t. $x \cdot y = 0$, then by introducing $z = x + \varepsilon \frac{y}{\|y\|}$, we get $z \cdot y > 0$. But as $z \in B(x, \varepsilon)$, we get a contradiction.
Conversely, let $x_0 \in C$ s.t. $(\forall y \in C^0,\ y \neq 0)$, $x_0 \cdot y < 0$. Consider the linear form: $y \mapsto x_0 \cdot y$. Hence, by compactness of the unit sphere, we get

$$\sup_{y \in C^0,\, \|y\|=1} x_0 \cdot y = -\alpha < 0.$$

Therefore, $(\forall y \in C^0)$, $x_0 \cdot y \leq -\alpha \|y\|$. Then $\forall x \in B(x_0, \frac{\alpha}{2})$ and $\forall y \in C^0$,

$$x \cdot y \leq (x - x_0) \cdot y + x_0 \cdot y \leq \frac{\alpha}{2}\|y\| - \alpha\|y\| \leq -\frac{\alpha}{2}\|y\|.$$

Then $\forall x \in B(x_0, \frac{\alpha}{2})$, $\forall y \in C^0 \backslash \{0\}$, $x \cdot y < 0$. By recalling that $(C^0)^0 = \overline{C}$ and that $(C^0)^0 = \{x \in \mathbb{R}^q,\ s.t.\ \forall\, y \in C^0,\ x \cdot y \leq 0\}$, we deduce that $B(x_0, \frac{\alpha}{2}) \subset \overline{C}$ i.e. $x_0 \in \texttt{int}(C)$. □

Proof. (Proposition 4) Let q be the dimension of the space generated by $\mathbf{m}$. Then let us prove that there exists $\mathbf{x}_1, \cdots, \mathbf{x}_q$ such that the family $\mathbf{m}(\mathbf{x}_k), k \in [1, q]$ is independent. It is obvious first that there exists $\mathbf{x}_1$ such that $\mathbf{m}(\mathbf{x}_1) \neq \mathbf{0}$. Otherwise, for any $\boldsymbol{\gamma}$ and for any $\mathbf{x}$ there is $\boldsymbol{\gamma} \cdot \mathbf{m}(\mathbf{v})) = 0$ and the family $\mathbf{m}(\mathbf{v}))$ cannot be pseudo-haar. That being said, there exists $\mathbf{x}_2$ such that $\mathbf{m}(\mathbf{x}_1), \mathbf{m}(\mathbf{x}_2)$ is independent. If we assume the contrary, this means that for any $\mathbf{x}$ there is $\lambda(\mathbf{x})$ such that $\mathbf{m}(\mathbf{x}) = \lambda(\mathbf{x})\mathbf{m}(\mathbf{x}_1)$. Choose now $\boldsymbol{\gamma}$ not zero orthogonal to $\mathbf{m}(\mathbf{x}_1)$. Then we have for any $\mathbf{x}$, $\boldsymbol{\gamma} \cdot \mathbf{m}(\mathbf{x}) = \lambda(\mathbf{x})\boldsymbol{\gamma} \cdot \mathbf{m}(\mathbf{x}_1) = 0$. But this is in contradiction with the pseudo-haar assumption. By induction, if $\mathbf{m}(\mathbf{x}_1), \cdots, \mathbf{m}(\mathbf{x}_k)$ is independent such that $k < q$, we can always find $\boldsymbol{\gamma}$ not zero in the orthogonal part and so justify, by the pseudo-haar property, that there exists a $\mathbf{x}_{k+1}$ such that the family $\mathbf{m}(\mathbf{x}_1), \cdots, \mathbf{m}(\mathbf{x}_{k+1})$ is independent.
If C has an empty topological interior, then it is contained in an hyperplane and the subtraction of 2 elements of C is contained in an hyperplane containing 0. Here we just prove that $\mathbb{R}^q = C - C$. For this let us consider the family $\mathbf{m}(\mathbf{x}_k), k \in [1, q]$. By the former lemma it is (independent so also) generating. Then any $\boldsymbol{\rho} \in \mathbb{R}^q$ can be written by a linear combination:

$$\boldsymbol{\rho} = \sum_{k=1}^{k=q} \lambda_k \mathbf{m}(\mathbf{x}_k)$$

Just let then write $\lambda_k = \lambda_k^+ - \lambda_k^-$ with x^+, x^- the positive and negative part of any real x. Then we have immediately:

$$\boldsymbol{\rho} = \sum_{k=1}^{k=q} \lambda_k^+ m(\mathbf{x}_k) - \sum_{k=1}^{k=q} \lambda_k^- m(\mathbf{x}_k)$$

so proving $\mathbb{R}^q = C - C$. □

Proof. (Lemma 1) Let $\mathbf{v} \neq \mathbf{0}$ and α such that $\alpha + \mathbf{u}\cdot\mathbf{v} = 0$. A direct calculation gives

$$[\alpha\ \mathbf{v}] \begin{bmatrix} 1 & \mathbf{u} \\ \mathbf{u} & \mathbb{D} \end{bmatrix} \begin{bmatrix} \alpha \\ \mathbf{v} \end{bmatrix} = \alpha^2 + 2\alpha\mathbf{u}\cdot\mathbf{v} + \mathbb{D} : \mathbf{v}\otimes\mathbf{v} > 0$$

By factorization there holds

$$\alpha^2 + 2\alpha\mathbf{u}\cdot\mathbf{v} + \mathbb{D} : \mathbf{v}\otimes\mathbf{v} = (\alpha + \mathbf{u}\cdot\mathbf{v})^2 + (\mathbb{D} - \mathbf{u}\otimes\mathbf{u}) : \mathbf{v}\otimes\mathbf{v} > 0. \tag{63}$$

Then as $\alpha + \mathbf{u}\cdot\mathbf{v} = 0$, it comes that $\mathbb{D} - \mathbf{u}\otimes\mathbf{u}$ is positive.
The converse statement is straightforwards. If $\mathbb{D} - \mathbf{u}\otimes\mathbf{u}$ is positive, choose $\alpha, \mathbf{v} \neq 0$. Since $(\mathbb{D} - \mathbf{u}\otimes\mathbf{u})$ is positive, then RHS of (63) is always non negative. It is zero if and only if both $\alpha + \mathbf{u}\cdot\mathbf{v}$ and $(\mathbb{D} - \mathbf{u}\otimes\mathbf{u}) : \mathbf{v}\otimes\mathbf{v}$ are zero. From positiveness of $\mathbb{D} - \mathbf{u}\otimes\mathbf{u}$ we get $\mathbf{v} = \mathbf{0}$. So $\alpha = 0$. □

Finally we prove quickly Proposition 5.

Proof. (Proposition 5). If $(n, \mathbf{0}, ne, n\Pi, n\mathbf{Q})$ is realizable, then

$$\int_{\mathbb{R}^3} (\mathbf{v} - \mathbf{u}) \otimes (\mathbf{v} - \mathbf{u}) f \, d\mathbf{v} = n\Pi + n\frac{k_B}{m} T\, I_d$$

is SPD. We proceed as in [12]. Let Θ_1, Θ_2, Θ_3 the eigenvalues of Π. Then the eigenvalues of $\lambda_{\mathbb{A}}\Pi + (1-\lambda_{\mathbb{A}})n\frac{k_B T}{m} I_d$ are

$$\frac{1+2\lambda_{\mathbb{A}}}{3}\Theta_1 + (1-\lambda_{\mathbb{A}})\frac{\Theta_2}{3} + (1-\lambda_{\mathbb{A}})\frac{\Theta_3}{3}, \quad (1-\lambda_{\mathbb{A}})\frac{\Theta_1}{3} + \frac{1+2\lambda_{\mathbb{A}}}{3}\Theta_2 + (1-\lambda_{\mathbb{A}})\frac{\Theta_3}{3},$$
$$(1-\lambda_{\mathbb{A}})\frac{\Theta_1}{3} + (1-\lambda_{\mathbb{A}})\frac{\Theta_2}{3} + \frac{1+2\lambda_{\mathbb{A}}}{3}\Theta_3.$$

So $\lambda_{\mathbb{A}}\Pi + (1-\lambda_{\mathbb{A}})n\frac{k_B T}{m} I_d$ is SPD for $\lambda_{\mathbb{A}} \in [-\frac{1}{2}, 1]$ Hence, the relaxed moment $(n, \mathbf{0}, ne, n\lambda_{\mathbb{A}}\Pi, n\lambda_{\mathbf{b}}\mathbf{Q})$ is realizable. □

Proof. (Proposition 7). Let $\boldsymbol{\rho} \in \mathcal{R}_{\mathbf{m}}^+$ and $f \in \mathbb{L}^{1,*}(\mathbf{m})$, $f \geq 0$ s.t. $R[f] = \boldsymbol{\rho}$. Let $\mathbf{u} \in \mathbb{R}^3$ and assume that

$$G(R[\tau_{\mathbf{u}} f]) = \tau_{\mathbf{u}} G(R[f]) = \tau_{\mathbf{u}} G(\boldsymbol{\rho}). \tag{64}$$

Hence there exists a relation between $\boldsymbol{\rho}$ and $R[\tau_u f]$. Remark that the relation defined by (64) does not depend on f as soon $f \in R^{-1}(\boldsymbol{\rho})$. Thus, the application $L = R \circ \tau_{\mathbf{u}} \circ$

$R^{-1}: \mathcal{R}^+_{\mathbf{m}} \to \mathcal{R}^+_{\mathbf{m}}$ is well defined as soon as $\tau_u f \in \mathbb{L}^{1,+}(\mathbf{m})$ if $f \in \mathbb{L}^{1,+}(\mathbf{m})$.
Under this condition, R^{-1} defines a linear map from $\mathcal{R}^+_{\mathbf{m}}$ into subsets of $\mathbb{L}^{1,*}(\mathbf{m})$ as follows:

$$(\forall\, \lambda \geq 0), R(R^{-1}(\boldsymbol{\rho}_1) + \lambda R^{-1}(\boldsymbol{\rho}_2)) = \int (f + \lambda g)\mathbf{m}(\mathbf{v}) d\mathbf{v} = \boldsymbol{\rho}_1 + \lambda \boldsymbol{\rho}_2$$

which is equivalent to $R^{-1}(\boldsymbol{\rho}_1 + \lambda\boldsymbol{\rho}_2) = R^{-1}(\boldsymbol{\rho}_1) + \lambda R^{-1}(\boldsymbol{\rho}_2))$. Hence $L = R \circ \tau_{\mathbf{u}} \circ R^{-1}$: $\mathcal{R}^+_{\mathbf{m}} \to \mathcal{R}^+_{\mathbf{m}}$ is linear.
Now remark that this relation can be extended to $\mathbb{R}^q$. Indeed, $\mathcal{R}^{+*}_{\mathbf{m}}$ is an open and non void set in $\mathbb{R}^q$ and contains q independent vectors $(\rho_1, \ldots, \rho_q)$ which form a basis of $\mathbb{R}^q$. Otherwise $\mathcal{R}^{+*}_{\mathbf{m}}$ would be contained in an hyperplane. Thus, the linear application L is well defined on $\mathbb{R}^q$ entirely.
$\forall \boldsymbol{\rho} \in \mathbb{R}^q$, $\forall f \in R^{-1}(\boldsymbol{\rho})$, there is $L\boldsymbol{\rho} = \int f\, L\mathbf{m}(\mathbf{v})\, d\mathbf{v}$ and $L\boldsymbol{\rho} = R \circ \tau_{\mathbf{u}} f$ at the same time i.e.

$$\int_{\mathbb{R}^3} \tau_{\mathbf{u}} f(\mathbf{v}) \mathbf{m}(\mathbf{v})\, d\mathbf{v} = \int_{\mathbb{R}^3} f(\mathbf{w})\, \mathbf{m}(\mathbf{w} - \mathbf{u})\, d\mathbf{w} = \int_{\mathbb{R}^3} f(\mathbf{w})\, L\mathbf{m}(\mathbf{v})\, d\mathbf{v}.$$

As $Im(R^{-1}(\mathbb{R}^q) = \mathbb{L}^1(\mathbf{m})$, the previous relation must be true for any $f \in \mathbb{L}^1(\mathbf{m})$. Hence $\mathbb{P} = \mathtt{span}(\mathbf{m})$ must be invariant under the action of $\tau_{-\mathbf{u}}$ and $L = \Lambda(-\mathbf{u})$ as defined in (42). We can proceed in the same way for any $\mathbf{u} \in \mathbb{R}^3$ and $\Theta \in SO(3)$, we deduce that $\mathbb{P}$ is invariant by Galilean transforms. □

6.2 Proof of section 4

Proof. (Lemma 2). Let us prove any of the claimings:

1. First let us remark that from the definition of ϕ-divergence, there holds the following property: since ϕ is strictly convex and that there is $\phi(0) = 0$ then the following function
$$\forall y \in (0, +\infty), \quad y \mapsto \frac{\phi(y)}{y} \quad \in (p_0, +\infty)$$
is strictly increasing and one to one from $(0, +\infty)$ onto $(p_0, +\infty)$. For any $p \leq p_0$ and for any $y > 0$ we have
$$p \leq p_0 < \frac{\phi(y)}{y}.$$
Then using $\phi(0)$ we have $\forall p \leq p_0,\ \forall y \geq 0,\quad py - \phi(y) \leq 0$. Hence, taking the supremum on $y \in \mathtt{dom}(\phi) = [0, +\infty)$ there holds $\forall p \leq p_0,\ \phi^*(p) \leq 0$. But on the other hand, since the function ϕ is convex and semi lower continuous, there holds $\phi^{**} = \phi$. As a consequence,
$$\phi(0) = -\inf_{p \in \mathbb{R}} \phi^*(p) = 0.$$
This means that $\inf_{p\in\mathbb{R}} \phi^*(p) = 0$. So we have $\forall p \in \mathbb{R},\ \phi^*(p) \geq 0,\ \forall p \leq p_0,\ \phi^*(p) \leq 0$. So $\forall p \leq p_0,\ \phi^*(p) = 0$.

2. Now let us prove the non negativity of ϕ^*. Since the function $\phi(y)/y$ is strictly increasing and one to one, it is also continuous (characterization of bijection on intervals). For $p \in (p_0, +\infty)$ there is just one element $y_p \in (0, +\infty)$ such that $py_p - \phi(y_p) = 0$. Then for any $y \in (0, y_p)$ we have $py_p - \phi(y_p) > 0$ and for any $y > y_p$ there holds $py_p - \phi(y_p) < 0$. In particular, there holds

$$\forall p > p_0, \quad \sup_{y \in \mathtt{dom}(\phi)} yp - \phi(y) = \sup_{y \in [0, y_p]} yp - \phi(y) \geq 0.$$

But any semi-upper continuous function gets its supremum on a compact set (and $yp - \phi(y)$ is semi-upper continuous). Then there exists $z_p \in [0, y_p]$ such that

$$\forall p > p_0, \quad \phi^*(p) = z_p p - \phi(z_p) \geq 0, \quad \phi^*(p) = z_p p - \phi(z_p) \in \mathbb{R}.$$

Finally, since for all $y \in (0, y_p)$, $yp - \phi(y) > 0$ the supremum is of course > 0.

3. From the former property, $\mathtt{dom}(\phi^*) = \mathbb{R}$. Then ϕ^* is continuous on $\mathbb{R}$ and at any point $p \in \mathbb{R}$, its sub-differential is not void. Assume that $y_1 < y_2$ are in its sub-differential. Then, by the characterization of the sub-differential for ϕ^*, and since $\phi = \phi^{**}$ (semi-lower continuity)

$$\phi(y_1) = py_1 - \phi^*(p), \quad \phi(y_2) = py_2 - \phi^*(p)$$

this also means that $p \in \partial\phi(y_1)$ and $p \in \partial\phi(y_2)$. So, $\phi(y) - \phi(y_i) \geq p(y - y_i)$. So

$$p(y_2 - y) \geq \phi(y_2) - \phi(y), \quad \phi(z) - \phi(y_1) \geq p(z - y_1).$$

For $y = y_1$ and $z = y_2$, there holds $p(y_2 - y_1) \geq \phi(y_2) - \phi(y_1) \geq p(y_2 - y_1)$. So $\phi(y_2) - \phi(y_1) = p(y_2 - y_1)$. In particular, since ϕ is convex, then for any $\alpha \in [0, 1]$ and $y = \alpha y_1 + (1 - \alpha) y_2$

$$\phi(y) \leq \alpha\phi(y_1) + (1 - \alpha)\phi(y_2) = \phi(y_2) + \alpha p(y_1 - y_2)$$

having $\phi(y_2) = -\phi^*(p) + py_2$ we get $\phi(y) \leq -\phi^*(p) + (1 - \alpha) py_2 + \alpha py_1 = -\phi^*(p) + py$. This proves that $p \in \partial\phi(y)$. Using the same inequalities as above we have then

$$\forall y \in [y_1, y_2], \quad \phi(y) - \phi(y_1) = p(y - y_1).$$

So ϕ is affine on $[y_1, y_2]$ (with $y_1 < y_2$) which contradicts that ϕ it is strictly convex. Then for any p the sub-differential $\partial\phi^*(p)$ has only one element. Then ϕ^* is differentiable. Finally any convex function on $\mathbb{R}$ which is differentiable is C^1 smooth.

□

Proof. (Proposition 10)

1. Define the entropy by

$$\forall g \in \mathbb{L}^1(\mathbf{m}), \quad \mathcal{H}(g) = \int \phi\left(\frac{g}{\mathcal{M}}\right)\mathcal{M} \in \mathbb{R}\bigcup\{+\infty\}.$$

First let us begin by assuming that $g \geq 0$ almost everywhere. ϕ is differentiable on $[0, +\infty[$ strictly, $\phi(0) = 0$ and $\phi(p) \to +\infty$ imply that ϕ is bounded from below. So

$$\mathcal{H}(g) \geq \min(\phi) \int \mathcal{M} d\mathbf{v}.$$

This proves that $\mathcal{H}$ is bounded form below independently of g.
Now assume that g takes negative values on a non zero measure set of $\mathbb{R}_+$ denoted by Ω. So $\phi = +\infty$ on Ω. Then

$$\mathcal{H}(g) = \int \phi\left(\frac{g}{\mathcal{M}}\right)\mathcal{M} = +\infty.$$

2. Finally $\mathcal{H}$ is strictly convex on its domain comes thanks to the strict convexity of ϕ.
3. Now let us prove the rest of the proposition.

- Let $\boldsymbol{\rho} \in \mathcal{R}_{\mathbf{m}}^+$. Then there exists $\Psi_{\boldsymbol{\rho}} \in \mathcal{C}_c^\infty(\mathbb{R}^3)$, $\Psi_{\boldsymbol{\rho}} \geq 0$ s.t. $\int \Psi_{\boldsymbol{\rho}}\, \boldsymbol{m}\, d\mathbf{v} = \boldsymbol{\rho}$. Thus the set $D^+(\boldsymbol{\rho})$ defined as

$$D^+(\boldsymbol{\rho}) = \{g \geq 0,\ \int g\, \boldsymbol{m} = \boldsymbol{\rho},\ \mathcal{H}(g) \leq \mathcal{H}(\Psi_{\boldsymbol{\rho}})\}$$

is non empty and convex. Moreover, $\inf \mathcal{H}$ exists on $D^+(\boldsymbol{\rho})$ but is not necessarily attained by a function in $D^+(\boldsymbol{\rho})$. Then $\mathcal{R}_{\mathbf{m}}^+ \subset \ \mathtt{dom}(h_{\mathbf{m}})$. If $\boldsymbol{\rho} \notin \mathcal{R}_{\mathbf{m}}^+$, then there are no nonnegative function that realizes $\boldsymbol{\rho}$. Hence, $h_{\mathbf{m}}(\boldsymbol{\rho}) = +\infty$ by definition of $h_{\mathbf{m}}$. So $\mathtt{dom}(h_{\mathbf{m}}) = \mathcal{R}_{\mathbf{m}}^+$.
- It is very clear that any ball of $\mathbb{R}^q$ which contains $\mathbf{0}$ contains moment $\boldsymbol{\rho} \notin \mathcal{R}_{\mathbf{m}}^+$. Then any subset of $\mathcal{R}_{\mathbf{m}}^+$ which contains $\mathbf{0}$ is not open. On the other hand the set $\mathcal{R}_{\mathbf{m}}^{+*}$ is open. Then it is obviously the biggest open set (in sense of inclusion) which is included in $\mathcal{R}_{\mathbf{m}}^+$. Then we have straightforwards: $\mathtt{int}(\mathtt{dom}(h_{\mathbf{m}})) = \mathtt{int}(\mathcal{R}_{\mathbf{m}}^+) = \mathcal{R}_{\mathbf{m}}^{+*}$.
- Now we need to prove that the function $h_{\mathbf{m}} : \mathbb{R}^d \mapsto \mathbb{R}\bigcup\{+\infty\}$ is convex. Let $\rho_1,\ \rho_2 \in \mathcal{R}_{\mathbf{m}}^{+*}$. Then $(\forall \varepsilon > 0)$, $\exists g_1,\ g_2 \in \mathbb{L}^1(\mathbf{m})$ with $\int g_i\, \boldsymbol{m} d\mathbf{v} = \boldsymbol{\rho}_i$ such that $h(\rho_i) > \mathcal{H}(g_i) - \varepsilon$ for $i \in \{1;\ 2\}$. Thus,

$$(\forall \lambda \in [0,1])\ \lambda h_{\mathbf{m}}(\rho_1) + (1-\lambda) h_{\mathbf{m}}(\rho_2) > \lambda \mathcal{H}(g_1) + (1-\lambda)\mathcal{H}(g_2) - \varepsilon.$$

$\mathcal{H}$ being strictly convex it comes that

$$(\forall \lambda \in [0,1])\ \lambda h_{\mathbf{m}}(\rho_1) + (1-\lambda) h_{\mathbf{m}}(\rho_2) > \mathcal{H}(\lambda g_1 + (1-\lambda) g_2) - \varepsilon.$$

By definition of $h_{\mathbf{m}}$ it holds that

$$(\forall \lambda \in [0,1])\ \mathcal{H}(\lambda g_1 + (1-\lambda) g_2) \geq h_{\mathbf{m}}(\lambda \rho_1 + (1-\lambda)\rho_2)$$

and the convexity of $h_{\mathbf{m}}$ follows.

□

6.3 Proofs of section 4

Proof. (Theorem 11)

1. Let $\boldsymbol{\rho} \in \mathbb{R}^q$ and recall that $D(\boldsymbol{\rho}) = \left\{ g \in \mathbb{L}^1(\mathbf{m}),\ \int \mathbf{m} g = \boldsymbol{\rho} \right\}$ (which is never the empty set). There is by definition:

$$h_{\mathbf{m}}(\boldsymbol{\rho}) = \inf_{g \in D(\boldsymbol{\rho})} \int \phi\left(\frac{g}{\mathcal{M}}\right) \mathcal{M}$$

Noting that for any $g \in D(\boldsymbol{\rho})$ there holds $\boldsymbol{\rho} = \int \mathbf{a} g$ we get

$$\boldsymbol{\alpha} \cdot \boldsymbol{\rho} + \sup_{g \in D(\boldsymbol{\rho})} \left[- \int \phi\left(\frac{g}{\mathcal{M}}\right) \mathcal{M} \right] = \sup_{g \in D(\boldsymbol{\rho})} \left(\int \left[\mathbf{m} \cdot \boldsymbol{\alpha} g - \phi\left(\frac{g}{\mathcal{M}}\right) \mathcal{M} \right] \right)$$

As $\boldsymbol{\alpha} \cdot \boldsymbol{\rho} - h_{\boldsymbol{m}}(\boldsymbol{\rho}) = h^*_{\boldsymbol{m}}(\boldsymbol{\alpha})$ we have

$$h^*_{\boldsymbol{m}}(\boldsymbol{\alpha}) = \sup_{\boldsymbol{\rho} \in \mathbb{R}^q} \sup_{g \in D(\boldsymbol{\rho})} \left(\int \left[\mathbf{m} \cdot \boldsymbol{\alpha}\, g - \phi\left(\frac{g}{\mathcal{M}}\right) \mathcal{M} \right] \right).$$

The next step consists to show that $\sup_{\boldsymbol{\rho} \in \mathbb{R}^q}$ and $\sup_{g \in D(\boldsymbol{\rho})}$ can be permuted in the previous formula.
Let $\boldsymbol{\rho}$ be fixed. It is clear that $\forall \boldsymbol{\rho} \in \mathbb{R}^q$

$$\sup_{g \in D(\boldsymbol{\rho})} \left(\int \left[\mathbf{m} \cdot \boldsymbol{\alpha} g - \phi\left(\frac{g}{\mathcal{M}}\right) \mathcal{M} \right] \right) \leq \sup_{g \in \mathbb{L}^1} \left(\int \left[\mathbf{m} \cdot \boldsymbol{\alpha} g - \phi\left(\frac{g}{\mathcal{M}}\right) \mathcal{M} \right] \right)$$

then obviously we have:

$$\sup_{\boldsymbol{\rho} \in \mathbb{R}^q} \sup_{g \in D(\boldsymbol{\rho})} \left(\int \left[\mathbf{m} \cdot \boldsymbol{\alpha} g - \phi\left(\frac{g}{\mathcal{M}}\right) \mathcal{M} \right] \right) \leq \sup_{g \in \mathbb{L}^1} \left(\int \left[\mathbf{m} \cdot \boldsymbol{\alpha} g - \phi\left(\frac{g}{\mathcal{M}}\right) \mathcal{M} \right] \right)$$

On the other hand, let $g \in \mathbb{L}^1(\mathbf{m})$ and note by $\boldsymbol{\rho}(g) = \int \mathbf{m} g \in \mathbb{R}^q$. Then

$$\int \left[\mathbf{m} \cdot \boldsymbol{\alpha} g - \phi\left(\frac{g}{\mathcal{M}}\right) \mathcal{M} \right] \leq \sup_{\psi \in D(\boldsymbol{\rho}(g))} \int \left[\mathbf{m} \cdot \boldsymbol{\alpha} \psi - \phi\left(\frac{\psi}{\mathcal{M}}\right) \mathcal{M} \right].$$

We have then

$$\sup_{g\in\mathbb{L}^1(\mathbf{a})}\int\left[\mathbf{m}\cdot\boldsymbol{\alpha}g-\phi\left(\frac{g}{\mathcal{M}}\right)\mathcal{M}\right]\leq\sup_{g\in\mathbb{L}^1(\mathbf{a})}\sup_{\psi\in D(\boldsymbol{\rho}(g))}\int\left[\mathbf{m}\cdot\boldsymbol{\alpha}\psi-\phi\left(\frac{\psi}{\mathcal{M}}\right)\mathcal{M}\right].$$

But there is $\mathbb{R}^q=\left\{\int\mathbf{m}g,g\in\mathbb{L}^1\right\}$ (see for example the construction given in (45)). Then

$$\sup_{g\in\mathbb{L}^1(\mathbf{a})}\sup_{\psi\in D(\boldsymbol{\rho}(g))}=\sup_{\boldsymbol{\rho}\in\mathbb{R}^q}\sup_{\psi\in D(\boldsymbol{\rho})}$$

and finally we have

$$\sup_{g\in\mathbb{L}^1(\mathbf{m})}\int\left[\mathbf{m}\cdot\boldsymbol{\alpha}g-\phi\left(\frac{g}{\mathcal{M}}\right)\mathcal{M}\right]\leq\sup_{\boldsymbol{\rho}\in\mathbb{R}^q}\sup_{\psi\in D(\boldsymbol{\rho})}\int\left[\mathbf{m}\cdot\boldsymbol{\alpha}\psi-\phi\left(\frac{\psi}{\mathcal{M}}\right)\mathcal{M}\right]$$

and we get formula (52).

2. Pick $\boldsymbol{\alpha}\in\mathbb{R}^3$ and consider the polynomial $\pi_{\boldsymbol{\alpha}}:=\boldsymbol{\alpha}\cdot\mathbf{m}$ and $G_{\boldsymbol{\alpha}}=\phi^{*\prime}\left(\pi_{\boldsymbol{\alpha}}\right)\mathcal{M}$. From the characterization of sub-differential of ϕ^* at (real) point $G_{\boldsymbol{\alpha}}/\mathcal{M}$, and taking into account that ϕ is semi-lower continuous, we have

$$\phi\left(\frac{G_{\boldsymbol{\alpha}}}{\mathcal{M}}\right)+\phi^*\left(\pi_{\boldsymbol{\alpha}}\right)=\pi_{\boldsymbol{\alpha}}\frac{G_{\boldsymbol{\alpha}}}{\mathcal{M}}$$

Using Young inequality for ϕ^* we have straightforwards

$$\phi^*\left(2\pi_{\boldsymbol{\alpha}}\right)-\phi^*\left(\pi_{\boldsymbol{\alpha}}\right)\geq\pi_{\boldsymbol{\alpha}}\phi^{*\prime}\left(\pi_{\boldsymbol{\alpha}}\right)=\pi_{\boldsymbol{\alpha}}\frac{G_{\boldsymbol{\alpha}}}{\mathcal{M}}.$$

Recall that we have for any convex function $\phi:\mathbb{R}\mapsto\overline{\mathbb{R}}$

$$-\phi^*\left(0\right)=\inf_{y\in\mathbb{R}}\phi\left(y\right)$$

Then there holds immediately

$$-\phi^*\left(0\right)\leq\phi\left(\frac{G_{\boldsymbol{\alpha}}}{\mathcal{M}}\right)=\pi_{\boldsymbol{\alpha}}\frac{G_{\boldsymbol{\alpha}}}{\mathcal{M}}-\phi^*\left(\pi_{\boldsymbol{\alpha}}\right)\leq\phi^*\left(2\pi_{\boldsymbol{\alpha}}\right)-2\phi^*\left(\pi_{\boldsymbol{\alpha}}\right)$$

Since $\pi_{\boldsymbol{\alpha}}$ is a polynomial function, and since by the assumption 3 any $\phi^*\left(\pi\right)$ is in $\mathbb{L}^1\left(\mathcal{M}d\mathbf{v}\right)$, then the former inequalities prove that $\phi\left(\frac{G_{\boldsymbol{\alpha}}}{\mathcal{M}}\right)\in\mathbb{L}^1\left(\mathcal{M}d\mathbf{v}\right)$. Using then

$$\phi\left(\frac{G_{\boldsymbol{\alpha}}}{\mathcal{M}}\right)+\phi^*\left(\pi_{\boldsymbol{\alpha}}\right)=\pi_{\boldsymbol{\alpha}}\frac{G_{\boldsymbol{\alpha}}}{\mathcal{M}}$$

we deduce finally that $\frac{G_{\boldsymbol{\alpha}}}{\mathcal{M}}\pi_{\boldsymbol{\alpha}}$ is also in $\mathbb{L}^1\left(\mathcal{M}d\mathbf{v}\right)$. By Young inequality we have for any g

$$\frac{g}{\mathcal{M}}\pi_{\boldsymbol{\alpha}}-\phi\left(\frac{g}{\mathcal{M}}\right)\leq\phi^*\left(\pi_{\boldsymbol{\alpha}}\right).$$

Multiplying by $\mathcal{M}$ and integrating (any term can be computed) there holds:

$$\int \left[\pi_{\boldsymbol{\alpha}} g - \phi\left(\frac{g}{\mathcal{M}}\right)\mathcal{M}\right] \leq \int \phi^*\left(\pi_{\boldsymbol{\alpha}}\right)\mathcal{M}.$$

By having the infinimum:

$$\sup_{g\in\mathbb{L}^1(\mathbf{a})} \int \left[\pi_{\boldsymbol{\alpha}} g - \phi\left(\frac{g}{\mathcal{M}}\right)\mathcal{M}\right] \leq \int \phi^*\left(\pi_{\boldsymbol{\alpha}}\right)\mathcal{M}$$

which gives us finally:

$$h^*_{\mathbf{m}}\left(\boldsymbol{\alpha}\right) \leq \int \phi^*\left(\pi_{\boldsymbol{\alpha}}\right)\mathcal{M} = \int \phi^*\left(\boldsymbol{\alpha}\cdot\mathbf{m}\right)\mathcal{M}.$$

On the other hand, since we have

$$\phi^*\left(\pi_{\boldsymbol{\alpha}}\right)\mathcal{M} = \left(\phi\left(\frac{G_{\boldsymbol{\alpha}}}{\mathcal{M}}\right) - \frac{G_{\boldsymbol{\alpha}}}{\mathcal{M}}\pi_{\boldsymbol{\alpha}}\right)\mathcal{M}.$$

Any term can be integrated and by having integration we have:

$$\int \phi^*\left(\pi_{\boldsymbol{\alpha}}\right)\mathcal{M} = \int \left(\phi\left(\frac{G_{\boldsymbol{\alpha}}}{\mathcal{M}}\right) - \frac{G_{\boldsymbol{\alpha}}}{\mathcal{M}}\pi_{\boldsymbol{\alpha}}\right)\mathcal{M} \leq \sup_{g\in\mathbb{L}^1(\mathbf{a})} \int \left[\pi_{\boldsymbol{\alpha}} g - \phi\left(\frac{g}{\mathcal{M}}\right)\mathcal{M}\right]$$

Then we have exactly (53)

3. Finally, consider for $\epsilon \in (0,1]$ the function

$$f_\epsilon\left(\boldsymbol{\beta}\right) = \left(\frac{\phi^*\left(\pi_{\boldsymbol{\alpha}} + \epsilon\pi_{\boldsymbol{\beta}}\right) - \phi^*\left(\pi_{\boldsymbol{\alpha}}\right)}{\epsilon} - \phi^{*\prime}\left(\pi_{\boldsymbol{\alpha}}\right)\pi_{\boldsymbol{\beta}}\right)\mathcal{M}.$$

Young inequality gives $\epsilon\phi^{*\prime}\left(\pi_{\boldsymbol{\alpha}}\right)\pi_{\boldsymbol{\beta}}\mathcal{M} \leq \left[\phi^*\left(\pi_{\boldsymbol{\alpha}} + \epsilon\pi_{\boldsymbol{\beta}}\right) - \phi^*\left(\pi_{\boldsymbol{\alpha}}\right)\right]\mathcal{M}$. So $f_\epsilon\left(\boldsymbol{\beta}\right) \geq 0$. On the other hand, by using $\pi_{\boldsymbol{\alpha}} + \epsilon\pi_{\boldsymbol{\beta}} = \epsilon\left(\pi_{\boldsymbol{\alpha}} + \pi_{\boldsymbol{\beta}}\right) + \left(1-\epsilon\right)\pi_{\boldsymbol{\alpha}}$, the convexity of ϕ^* gives

$$\phi^*\left(\pi_{\boldsymbol{\alpha}} + \epsilon\pi_{\boldsymbol{\beta}}\right) \leq \epsilon\phi^*\left(\pi_{\boldsymbol{\alpha}} + \pi_{\boldsymbol{\beta}}\right) + \left(1-\epsilon\right)\phi^*\left(\pi_{\boldsymbol{\alpha}}\right).$$

Hence we get

$$\frac{\phi^*\left(\pi_{\boldsymbol{\alpha}} + \epsilon\pi_{\boldsymbol{\beta}}\right) - \phi^*\left(\pi_{\boldsymbol{\alpha}}\right)}{\epsilon} \leq \phi^*\left(\pi_{\boldsymbol{\alpha}} + \pi_{\boldsymbol{\beta}}\right) - \phi^*\left(\pi_{\boldsymbol{\alpha}}\right).$$

So, after multiplication by $\mathcal{M}$

$$f_\epsilon\left(\boldsymbol{\beta}\right) \leq \left[\phi^*\left(\pi_{\boldsymbol{\alpha}} + \pi_{\boldsymbol{\beta}}\right) - \phi^*\left(\pi_{\boldsymbol{\alpha}}\right) - \phi^{*\prime}\left(\pi_{\boldsymbol{\alpha}}\right)\pi_{\boldsymbol{\beta}}\right]\mathcal{M} = f_1$$

that is $0 \leq f_\epsilon(\boldsymbol{\beta}) \leq f_1(\beta)$. Since $\lim_{\epsilon \to 0} f_\epsilon(\boldsymbol{\beta}) = 0$ pointwise, using Lebesgues dominated convergence theorem there holds:

$$\forall \boldsymbol{\beta}, \lim_{\epsilon \mapsto 0^+} \frac{h^*(\boldsymbol{\alpha} + \epsilon \boldsymbol{\beta}) - h^*(\boldsymbol{\alpha})}{\epsilon} = \int \phi^*(\pi_{\boldsymbol{\alpha}}) \pi_{\boldsymbol{\beta}} \mathcal{M} = \boldsymbol{\beta} \cdot \left[\int \phi^*(\boldsymbol{\alpha} \cdot \mathbf{a}) \mathbf{a} \mathcal{M} \right].$$

The convex function $\boldsymbol{\alpha} \in \mathbb{R}^q \mapsto h^*(\boldsymbol{\alpha})$ has partial derivatives in any direction at any point. Then, using a classical result of convex analysis, it is C^1 smooth and we have (54).

□

Proof. (Theorem 9) Let us finally prove Theorem 9 step by step.

1. We first prove that $h_{\mathbf{m}}$ is closed in its domain. $h_{\mathbf{m}}$ being convex, this amounts to prove that $h_{\mathbf{m}}^{**} = h_{\mathbf{m}}$ in $\mathcal{R}_{\mathbf{m}}^+$. The delicate point is to prove this relation at $\mathbf{0}$. Let us observe the following: there is $h_{\mathbf{m}}(\mathbf{0}) = 0$. This is just because the only non negative function which is able to realize $\mathbf{0}$ is $g = 0$. Since $\phi(0) = 0$ then we deduce immediately that $h_{\mathbf{m}}(\mathbf{0}) = 0$. Let us now compute $h_{\mathbf{m}}^{**}(\mathbf{0}) := \sup_{\boldsymbol{\alpha} \in \mathbb{R}^q} (\mathbf{0} \cdot \boldsymbol{\alpha} - h_{\mathbf{m}}^*(\boldsymbol{\alpha})) = -\inf_{\boldsymbol{\alpha} \in \mathbb{R}^q} h_{\mathbf{m}}^*(\boldsymbol{\alpha}) \leq 0$. Remark then that $h_{\mathbf{m}}^{**}(\mathbf{0}) \leq 0$ from the expression given of h^* (theorem 11). Recall from Lemma 2 that $\phi^*(y) \geq 0$ and for $y \leq p_0, \phi^*(y) = 0$. Chose now $\boldsymbol{\alpha}_0 = (y_0, \mathbf{0}, \cdots, \mathbf{0})$. There is $\phi^*(\boldsymbol{\alpha}_0 \cdot \mathbf{m}(\mathbf{v})) = 0$. So $h_{\mathbf{m}}^*(\boldsymbol{\alpha}_0) = 0$ and finally $h_{\mathbf{m}}^{**}(\mathbf{0}) = 0 = h_{\mathbf{m}}(\mathbf{0})$.
   Finally, $h_{\mathbf{m}}$ being convex, $h_{\mathbf{m}}$ is continuous on $\mathtt{int}(\mathtt{dom}(h_{\mathbf{m}}))$. As a consequence there is $h_{\mathbf{m}}^{**} = h_{\mathbf{m}}$ in $\mathtt{dom}(h_{\mathbf{m}}) = \mathcal{R}_{\mathbf{m}}^+$.
2. Let $\boldsymbol{\rho} \in \mathtt{int}(\mathtt{dom}\ (h_{\mathbf{m}}))$. $h_{\mathbf{m}}$ being continuous at this point, there is $\partial h_{\mathbf{m}}(\boldsymbol{\rho}) \neq \emptyset$ [38]. Pick some $\boldsymbol{\alpha} \in \partial h(\boldsymbol{\rho})$. $h_{\mathbf{m}}$ being proper convex and closed at $\boldsymbol{\rho}$, there is $\boldsymbol{\rho} \in \partial h_{\mathbf{m}}(\boldsymbol{\alpha})$. But $h_{\mathbf{m}}^*$ is $\mathcal{C}^1$ in $\mathbb{R}^q$ so $\partial h_{\mathbf{m}}(\boldsymbol{\alpha}) = \nabla h_{\mathbf{m}}^*(\boldsymbol{\alpha})$ and $\boldsymbol{\rho} = \nabla h_{\mathbf{m}}^*(\boldsymbol{\alpha})$.
3. Let $\boldsymbol{\rho} \in \mathtt{int}(\mathtt{dom}\ (h_{\mathbf{m}}))$ and $\boldsymbol{\alpha} \in \partial h_{\mathbf{m}}(\boldsymbol{\rho})$. We prove that the function $G = \mathcal{M} \phi^{*\prime}(\boldsymbol{\alpha} \cdot \mathbf{m}(\mathbf{v}))$ is the unique solution to the primal problem. We firstly have

$$\boldsymbol{\rho} = \nabla h_{\mathbf{m}}^*(\boldsymbol{\alpha}) = \int \phi^{*\prime}(\boldsymbol{\alpha} \cdot \mathbf{m}(\mathbf{v})) \mathbf{m}(\mathbf{v}) \mathcal{M} d\mathbf{v}.$$

   Next ϕ^* is a $\mathcal{C}^1$ convex function in $\mathbb{R}$ and thus

$$(\forall y \in \mathbb{R}),\ \phi^*(y) + \phi^{**}((\phi^*)'(y)) = y(\phi^*)'(y).$$

   ϕ is also convex, proper and semi lower continuous. So $\phi^{**} = \phi$ and

$$(\forall y \in \mathbb{R}),\ \phi^*(y) + \phi((\phi^*)'(y)) = y(\phi^*)'(y).$$

   Put $y = \boldsymbol{\alpha} \cdot \mathbf{m}(\mathbf{v})$ in the above equation, multiply by $\mathcal{M}$ and integrate w.r.t v gives

$$h_{\mathbf{m}}^*(\boldsymbol{\alpha}) + \int \mathcal{M} \phi((\phi^*)'(\boldsymbol{\alpha} \cdot \mathbf{m})) = \int \mathcal{M}(\boldsymbol{\alpha} \cdot \mathbf{m})(\phi^*)'(\boldsymbol{\alpha} \cdot \mathbf{m}) = \boldsymbol{\alpha} \cdot \boldsymbol{\rho}.$$

One then deduces from the subdifferential equation (49) that

$$h_{\mathbf{m}}(\boldsymbol{\rho}) = \int \mathcal{M}\phi\left((\phi^*)'(\boldsymbol{\alpha}\cdot\mathbf{m})\right) = \mathcal{H}\left(G\right).$$

Recall that $\mathcal{H}$ is strictly convex and thus G is the unique solution to the primal problem. From the form $G = \mathcal{M}\phi^{*\prime}\left(\boldsymbol{\alpha}\cdot\mathbf{m}\left(\mathbf{v}\right)\right)$ which is necessarily strictly positive on a set of non-zero measure - that is for those velocities $\mathbf{v}$ for which $\boldsymbol{\alpha}\cdot\mathbf{m}(v) > p_0$ - $\boldsymbol{\alpha}$ is found to be unique. This in turn proves that the subdifferential of h at interior point of $\mathcal{R}^+_{\mathbf{m}}$ is reduced to one point. As a consequence ∇h^* is a bijection from from C° to $\mathcal{R}^+_{\mathbf{m}}$ where C° is defined in (51). And there is $\nabla h^*(\boldsymbol{\alpha}) = 0$ in the complementary set of C°.

4. Let us finally prove that $h_{\mathbf{m}}$ is strictly convex on $\mathcal{R}^+_{\mathbf{m}}$. Let $\boldsymbol{\rho}_1 \neq \boldsymbol{\rho}_2 \in \mathcal{R}^+_{\mathbf{m}}$. Consider $\boldsymbol{\rho}(t) = (1-t)\boldsymbol{\rho}_1 + t\boldsymbol{\rho}_2$, $t \in [0,1]$ and $\boldsymbol{\alpha}(t)$ s.t. $\nabla h^*_{\mathbf{m}}(\boldsymbol{\alpha}(t)) = \boldsymbol{\rho}(t)$. In particular, we have $\boldsymbol{\rho}(0) = \boldsymbol{\rho}_1$ and $\boldsymbol{\rho}(1) = \boldsymbol{\rho}_2$. Define f_1 and f_2 by $f_1 = \mathcal{M}(\phi^*)'(\boldsymbol{\alpha}(0)\cdot\mathbf{m})$ and $f_2 = \mathcal{M}(\phi^*)'(\boldsymbol{\alpha}(1)\cdot\mathbf{m})$. They satisfy the relation

$$h_{\mathbf{m}}(\boldsymbol{\rho}_1) = \mathcal{H}(f_1),\ h_{\mathbf{m}}(\boldsymbol{\rho}_2) = \mathcal{H}(f_2). \tag{65}$$

Moreover, $(1-t)f_1 + tf_2$ is a nonnegative function which moment is $\boldsymbol{\rho}(t)$. Then

$$(\forall t \in]0,1[) \quad h_{\mathbf{m}}(\boldsymbol{\rho}(t)) \leq \mathcal{H}((1-t)f_1 + tf_2).$$

$\mathcal{H}$ being strictly convex, we get from (65) for any $t \in]0,1[$,

$$h_{\mathbf{m}}(\boldsymbol{\rho}(t)) < (1-t)\mathcal{H}(f_1) + t\mathcal{H}(f_2) = (1-t)h_{\mathbf{m}}(\boldsymbol{\rho}_1) + th_{\mathbf{m}}(\boldsymbol{\rho}_2).$$

□

6.4 Proof of section 5

Proof. (Proposition 12). We may first consider the rotation around the mean velocity u since they play an important role to obtain the right hydrodynamic limit. So we define $\tau = \tau_{\mathbf{u}^{-1}\theta\mathbf{u}} = \tau_{\mathbf{u}^{-1}}\tau_\theta\tau_{\mathbf{u}}$ and let us prove the result for this τ. By definition of G, $G(f) = \mathcal{M}(\phi^*)'(\boldsymbol{\alpha}\cdot\boldsymbol{a}(\mathbf{v}-\mathbf{u}))$, where $\boldsymbol{\alpha}$ is the polar variable of $L(\boldsymbol{\rho}_f)$. In other words, $\nabla h^*(\boldsymbol{\alpha}) = L(\boldsymbol{\rho}_f)$ By definition of $\mathcal{M}$, $\mathcal{M}$ remains unchanged with the transformation τ. So $\tau(G(f)) = \mathcal{M}(\phi^*)'(\boldsymbol{\alpha}\cdot\boldsymbol{a}(\tau(\mathbf{v})-\mathbf{u})$, with

$$\boldsymbol{a}(\tau(\mathbf{v})-\mathbf{u}) = (1,\, \theta(\mathbf{v}-\mathbf{u}),\, \frac{(\mathbf{v}-\mathbf{u})^2}{2} - \frac{3}{2}k_BT,\, \theta(\mathbf{v}-\mathbf{u})\otimes\theta(\mathbf{v}-\mathbf{u}) - \frac{1}{3}(\mathbf{v}-\mathbf{u})^2\, I_d,$$
$$\theta(\mathbf{v}-\mathbf{u})(\frac{(\mathbf{v}-\mathbf{u})}{2} - \frac{5}{2}).$$

This last vector can be written as

$$\boldsymbol{a}(\tau(\mathbf{v})-\mathbf{u}) = \big(1,\, \theta(\mathbf{v}-\mathbf{u}),\, a_2(\mathbf{v}-\mathbf{u}),\, \theta\mathbb{A}(\mathbf{v}-\mathbf{u})\theta^t,\, \theta\mathbf{b}(\mathbf{v}-\mathbf{u})\big) \tag{66}$$

A simple computation leads to $\boldsymbol{\alpha} : \boldsymbol{a}(\tau(\mathbf{v}-\mathbf{u})) = \Theta(\boldsymbol{\alpha}) : \boldsymbol{a}(\mathbf{v}-\mathbf{u})$, with

$$\Theta(\boldsymbol{\alpha}) = (\boldsymbol{\alpha}_0,\, \theta^t\boldsymbol{\alpha}_1,\, \alpha_2,\, \theta^t\boldsymbol{\alpha}_3\theta,\, \theta^t\boldsymbol{\alpha}_4). \tag{67}$$

And thus $\tau(G(f)) = \mathcal{M}(\phi^*)'(\Theta(\boldsymbol{\alpha}) \cdot \boldsymbol{a}(\mathbf{v}-\mathbf{u})$. Remark that $\tau G(f)$ has the form of the solution of Theorem 9 for some moment. Let us compute this moment. $\boldsymbol{\rho}_{\tau G(f)}$ is defined by

$$\begin{aligned}\boldsymbol{\rho}_{\tau G(f)} &= \int_{\mathbb{R}^3} \tau G(f)(\mathbf{v})\, \boldsymbol{a}(\mathbf{v}-\mathbf{u}) d\mathbf{v} \\ &= \int_{\mathbb{R}^3} G(\tau\mathbf{v}) \boldsymbol{a}(\mathbf{v}-\mathbf{u}) d\mathbf{v}.\end{aligned}$$

By using the change of variable $\boldsymbol{w} = \tau(\boldsymbol{v}) = \theta(\boldsymbol{v}-\boldsymbol{u}) + \boldsymbol{u}$,

$$\boldsymbol{\rho}_{\tau G(f)} = \int_{\mathbb{R}^3} G(\mathbf{w}) \boldsymbol{a}(\theta^t(\mathbf{w}-\mathbf{u})) d\mathbf{w}.$$

But as $\Theta(\boldsymbol{a}(\mathbf{v}-\mathbf{u})) = \boldsymbol{a}(\theta^t(\mathbf{v}-\mathbf{u}))$. Then

$$\boldsymbol{\rho}_{\tau G(f)} = \Theta\left(\int_{\mathbb{R}^3} G(w) \boldsymbol{a}(\mathbf{w}-\mathbf{u}) d\mathbf{w}\right) = \Theta\left(\boldsymbol{\rho}_G\right).$$

Likewise with the same computations gives $\boldsymbol{\rho}_{\tau(f)} = \Theta(\boldsymbol{\rho}_{(f)})$. Thus $G(\tau(f))$ is the solution of Theorem 9 for $\boldsymbol{\rho} = L(\boldsymbol{\rho}_f)$. Now remark for the definition of L and Θ that $\Theta L(\boldsymbol{\rho}_f) = L(\Theta(\boldsymbol{\rho}_f))$. As a conclusion $\tau G(f) = G(\tau f)$.
It remains to prove the result for the translations i.e. $\tau_{\mathbf{z}} G(f) = G(\tau_{\mathbf{z}} f)$. There is $G = \mathcal{M}(\phi^*)'\left(\boldsymbol{\alpha} \cdot \boldsymbol{a}(\mathbf{v}-\mathbf{u})\right)$. Moreover

$$\tau_{\mathbf{z}} G(f) = \frac{n}{(2\pi T)^{\frac{3}{2}}} \exp\left(-\frac{(\mathbf{v}-\mathbf{u}-\mathbf{z})}{2T}\right) \phi^*\left(\boldsymbol{\alpha} \cdot \boldsymbol{a}((\mathbf{v}-\mathbf{u}-\mathbf{z}))\right)$$

$G(\tau_{\mathbf{z}} f)$ is the solution of the minimization problem when changing the Grad basis into the framework related to $\tau_{\mathbf{z}} f$ which is precisely moving at velocity $\mathbf{u}+\mathbf{z}$.

$$\int_{\mathbb{R}^3} \tau_{\mathbf{z}} f(1,\, \mathbf{v},\, \mathbf{v}^2)\, d\mathbf{v} = (n,\, n(\mathbf{u}+\mathbf{z}),\, \frac{1}{2} n(\mathbf{u}+\mathbf{z})^2 + \frac{3}{2} nT).$$

In the corresponding Grad basis $\tau_z f$ has the same macroscopic value $\boldsymbol{\rho}_f$ as f in $\boldsymbol{a}(\mathbf{v}-\mathbf{u}-\mathbf{z})$. Namely

$$\boldsymbol{\rho}_f = \int_{\mathbb{R}^3} f(\mathbf{v}) \boldsymbol{a}(\mathbf{v}-\mathbf{u}) d\mathbf{v} = \int_{\mathbb{R}^3} \tau_{\mathbf{z}} f(\mathbf{v}) \boldsymbol{a}(\mathbf{v}-\mathbf{u}) d\mathbf{v}$$

So

$$L(\boldsymbol{\rho}_f) = \int_{\mathbb{R}^3} \mathcal{M}(\mathbf{v}-\mathbf{z})(\phi^*)'(\boldsymbol{\alpha}\cdot\boldsymbol{a}((\mathbf{v}-\mathbf{u}-\mathbf{z})\boldsymbol{a}(\mathbf{v}-\mathbf{u}-\mathbf{z})d\mathbf{v}.$$

This means that the relation between $\boldsymbol{\alpha}$ and $L(\boldsymbol{\rho}_f)$ is valid whatever is framework. The minimisation problem remains unchanged by changing both $\mathcal{M}$ and f in $\tau_{\mathbf{z}}\mathcal{M}$ and $\tau_{\mathbf{z}}f$. □

Proof. (Theorem 13). Remark that $\mathcal{M}$ is the unique minimizer of $\mathcal{H}$ just under the constraints of conservation laws. Indeed there is $\mathcal{M} = \mathcal{M}\times 1 = \mathcal{M}\phi^{*\prime}(\alpha_0)$ for some $\alpha_0 \in \mathbb{R}$ since $\phi^{*\prime}$ is a bijection from $\mathbb{R}^+$ to $\mathbb{R}^+$. Thus $\mathcal{M}$ has the form of the solution of the primal problem given by theorem 9 when $\mathbf{m}(\mathbf{v}) = \{\mathbf{1},\, \mathbf{v},\, \mathbf{v}^2\}$ and the constraint is $\int G\boldsymbol{m}d\mathbf{v} = (n,\, 0,\, 0)$. Adding more constraints (w.r.t. $\mathbb{A}(\mathbf{v}-\mathbf{u})$ and $\boldsymbol{b}(\mathbf{v}-\mathbf{u})$) prove that $\mathcal{H}(\mathcal{M}) = h(\boldsymbol{\rho}_{eq})$ since again $\mathcal{M}$ has the shape of the solution and satisfies the constraints. We have

$$\begin{aligned}(\forall\boldsymbol{\rho}\in\mathcal{R}^+_{\boldsymbol{a}}),\ \text{with}\ \boldsymbol{\rho} = (n,\,0,\,0,\,0,\,0),\ h_{\mathbf{a}}(\boldsymbol{\rho}_{eq}) \le h_{\mathbf{a}}(\boldsymbol{\rho})\\ \text{with equality}\ h_{\mathbf{a}}(\boldsymbol{\rho}_{eq}) = h_{\mathbf{a}}(\boldsymbol{\rho})\ \text{iff}\ \boldsymbol{\rho}_{eq} = \boldsymbol{\rho}\end{aligned} \tag{68}$$

thanks to the strict convexity of h in $\mathcal{R}^+_{\boldsymbol{a}}$ (Theorem 9). In other words, $\mathcal{M}$ is the unique minimizer of $\mathcal{H}$ of all functions in $\mathbb{L}^1(\boldsymbol{a})$ having the same mass, momentum and energy as f.
We define the function F which satisfies

$$\int F\mathbf{a}(\mathbf{v}-\mathbf{u})d\mathbf{v} = \boldsymbol{\rho}_f \quad \text{and} \quad \mathcal{H}(F) = h_{\boldsymbol{a}}(\boldsymbol{\rho}_f). \tag{69}$$

F is unique thanks to Theorem 9 and reads $F = \mathcal{M}(\phi^*)'(\boldsymbol{\alpha}_F\cdot\mathbf{a}(\mathbf{v}-\mathbf{u}))$. Consider as in the proof of Lemma 12 τf with $\tau = \tau_{-\mathbf{u}}\tau_{-I_d}\tau_{\mathbf{u}}$. Then

$$\Theta(\boldsymbol{\rho}_f) = \int_{\mathbb{R}^3} \tau f(\mathbf{v})\mathbf{a}(\mathbf{v}-\mathbf{u}))d\mathbf{v} = (n,\mathbf{0},0,n\Pi,-n\mathbf{b})\,, \quad h(\tilde{\boldsymbol{\rho}}) = h(\boldsymbol{\rho}).$$

There is $\tau\mathcal{M} = \mathcal{M}$ in such a way that $\mathcal{H}(\tau f) = \mathcal{H}(f)$ and $h_{\boldsymbol{a}}(\Theta\boldsymbol{\rho}) = h_{\boldsymbol{a}}(\boldsymbol{\rho})$, since τf is solution to the minimisation problem by changing $\boldsymbol{\rho}$ and $\Theta\boldsymbol{\rho}$.

From the strict convexity of $h_{\mathbf{a}}$, one finds

$$h_{\mathbf{a}}\left(n,\mathbf{0},0,n\Pi,\mathbf{0}\right) = h_{\mathbf{a}}\left(\frac{1}{2}\boldsymbol{\rho}+\frac{1}{2}\Theta(\boldsymbol{\rho})\right) \le \frac{1}{2}h_{\mathbf{a}}\left(\boldsymbol{\rho}\right)+\frac{1}{2}h_{\mathbf{a}}\left(\Theta(\boldsymbol{\rho})\right) = h_{\mathbf{a}}\left(\boldsymbol{\rho}\right). \tag{70}$$

with equality only if $\mathbf{b} = \mathbf{0}$.
Let then λ_a, $\lambda_b \in [0,1]$. Without a loss of generality, assume that $\lambda_b \le \lambda_a$ and take $\lambda\in[0,1]$ such that $\lambda_b = \lambda\lambda_a$ (if not, one sets $\lambda_a = \lambda\lambda_b$). We have

$$(n,\mathbf{0},0,n\Pi,\lambda\mathbf{b}) = \lambda\left(n,\mathbf{0},0,n\Pi,\mathbf{b}\right)+(1-\lambda)\left(n,\mathbf{0},0,n\Pi,\mathbf{0}\right),$$

and as a consequence $h_{\mathbf{a}}\left(n,\mathbf{0},0,n\Pi,\lambda\mathbf{b}\right)\leq h_{\mathbf{a}}\left(n,\mathbf{0},0,n\Pi,\mathbf{b}\right)$ where we have used (70). Again, the equality holds only if $\mathbf{b}=\mathbf{0}$. Likewise, we have

$$\left(\rho_1,\mathbf{0},\,0,\lambda_{\mathbb{A}}n\Pi,\lambda_b\mathbf{b}\right)=\left(1-\lambda_{\mathbb{A}}\right)\left(n,\mathbf{0},0,\mathbb{O},\mathbf{0}\right)+\lambda_{\mathbb{A}}\left(n,\mathbf{0},0,n\Pi,\lambda\mathbf{b}\right),$$

and finally (using former inequalities)

$$h_{\mathbf{a}}\left(n,\mathbf{0},0,\lambda_{\mathbb{A}}n\Pi,\lambda_{\mathbf{b}}\mathbf{b}\right)\leq h_{\mathbf{a}}\left(n,\mathbf{0},0,n\Pi,\mathbf{b}\right). \tag{71}$$

with equality only if $\Pi=\mathbb{O}$ and $\mathbf{b}=\mathbf{0}$ or $\lambda_{\mathbb{A}}=\lambda_{\mathbf{b}}=1$. There is

$$I:=\left\langle K\left(f\right)\partial_x\phi\left(\frac{f}{\mathcal{M}}\right)\right\rangle=\int\nu\left(G-f\right)\partial_x\phi\left(\frac{f}{\mathcal{M}}\right)d\mathbf{v}.$$

From the expression of G there holds

$$I=\int\left[\phi^{*\prime}\left(\boldsymbol{\alpha}\left(\mathtt{L}\left(\boldsymbol{\rho}_f\right)\right)\cdot\mathbf{a}\right)-\frac{f}{\mathcal{M}}\right]\partial_x\phi\left(\frac{f}{\mathcal{M}}\right)\mathcal{M}d\mathbf{v}.$$

Use Young inequality for ϕ: $\phi\left(y\right)-\phi\left(x\right)\geq\phi'\left(x\right)\left(y-x\right)$, with

$$x=\frac{f}{\mathcal{M}},\quad y=\phi^{*\prime}\left(\boldsymbol{\alpha}\left(\mathtt{L}\left(\boldsymbol{\rho}_f\right)\right)\cdot\mathbf{a}\right),$$

multiply by $\mathcal{M}$ and integrate over $\mathbb{R}^3$ gives

$$\begin{aligned}I\leq\int\nu\left(\phi\left(\phi^{*\prime}\left(\boldsymbol{\alpha}\left(\mathtt{L}\left(\boldsymbol{\rho}_f\right)\right)\cdot\mathbf{a}\right)\right)\mathcal{M}-\int\phi\left(\frac{f}{\mathcal{M}}\right)\mathcal{M}\right)&=\nu(\mathcal{H}(G)-\mathcal{H}(f))\\&=\nu(h_{\mathbf{a}}\left(\mathtt{L}\left(\boldsymbol{\rho}_f\right)\right)-\mathcal{H}(f)).\end{aligned}$$

Hence $I\leq\nu(h_{\mathbf{a}}\left(\mathtt{L}\left(\boldsymbol{\rho}_f\right)\right)-\mathcal{H}(f))$. This reads also

$$I\leq\nu\left(h_{\mathbf{a}}\left(\mathtt{L}\left(\boldsymbol{\rho}_f\right)\right)-h_{\mathbf{a}}\left(\boldsymbol{\rho}_f\right)+h_{\mathbf{a}}\left(\boldsymbol{\rho}_f\right)-\mathcal{H}(f)\right).$$

By definition of the entropy $h_{\mathbf{a}}\left(\boldsymbol{\rho}\right)$ there is $h_{\mathbf{a}}\left(\boldsymbol{\rho}_f\right)-\mathcal{H}(f)\leq 0$. But on the other hand we have by computing the moments on $\mathbf{a}\left(\mathbf{v}-\mathbf{u}\right)$

$$\boldsymbol{\rho}_f=\left(n,\mathbf{0},0,n\mathbb{A},n\mathbf{b}\right),\ \mathtt{L}\left(\boldsymbol{\rho}_f\right)=\left(n,\mathbf{0},0,\lambda_{\mathbb{A}}n\mathbb{A},\lambda_{\mathbf{b}}n\mathbf{b}\right),\ \lambda_{\mathbb{A}},\,\lambda_{\mathbf{b}}\in[0,1).$$

As $h_{\mathbf{a}}\left(\mathtt{L}\left(\boldsymbol{\rho}_f\right)\right)\leq h_{\mathbf{a}}\left(\boldsymbol{\rho}_f\right)$, the entropy theorem is proved.

From (71) and (69) $I=0$ iff $\mathtt{L}\left(\boldsymbol{\rho}_f\right)=\boldsymbol{\rho}_f$ and $f=F$. But $\mathtt{L}\left(\boldsymbol{\rho}_f\right)=\boldsymbol{\rho}_f$ iff $\boldsymbol{\rho}_f=\boldsymbol{\rho}_{eq}$ that is $f=\mathcal{M}$. □

Proof. (Proposition 14). Remark that thanks to Galilean invariance in the Grad space, the choice of the basis functions for defining the constraints does not change the result

of the minimization problem. This means that we may write

$$G[f](\mathbf{v}) = \mathcal{M}(\tilde{\boldsymbol{\mu}}(f)\cdot\mathbf{a}(\mathbf{v}))\,\phi^{*\prime}(\tilde{\boldsymbol{\alpha}}(f)\cdot\mathbf{a}(\mathbf{v})).$$

But it is more convenient to write those Lagrange multipliers when the basis is $\mathbf{a}(\mathbf{v}-\mathbf{u})$. That is

$$G[f](\mathbf{v}) = \exp(\boldsymbol{\mu}(f)\cdot\mathbf{a}(\mathbf{v}-\mathbf{u}))\,\phi^{*\prime}(\boldsymbol{\alpha}(f)\cdot\mathbf{a}(\mathbf{v}-\mathbf{u})).$$

Then there is

$$\exp(\boldsymbol{\mu}(\mathcal{M})\cdot\mathbf{a}(\mathbf{v}-\mathbf{u}))\,\phi^{*\prime}(\boldsymbol{\alpha}(\mathcal{M})\cdot\mathbf{a}(\mathbf{v}-\mathbf{u})) = \mathcal{M},$$

since $\mathcal{M}$ is the unique solution to the variational problem with the moment constraints

$$\int f\,\mathbf{m}(\mathbf{v})\,d\mathbf{v} = (n,\mathbf{0},0,0,0) = \boldsymbol{\rho}_{\mathcal{M}}.$$

In the above equation $\phi^{*\prime}(\boldsymbol{\alpha}(\mathcal{M})\cdot\mathbf{a}(\mathbf{v}-\mathbf{u})) = 1$ implies $\boldsymbol{\alpha}(\mathcal{M})\cdot\mathbf{a}(\mathbf{v}-\mathbf{u}) = \alpha$ for some constant α because $\phi^{*\prime}$ is a bijection from $[p_0,+\infty)$ into $\mathbb{R}^+$. So finally there holds also

$$\phi^{*\prime\prime}(\boldsymbol{\alpha}(\mathcal{M})\cdot\mathbf{a}(\mathbf{v}-\mathbf{u})) = c,$$

for some constant c if ϕ^* is twice differentiable.

Note that the properties of $G[f]$ implies that $G[\mathcal{M}] = \mathcal{M}$ so $K(\mathcal{M}) = 0$. Let us consider $f = \mathcal{M}(1+\epsilon g)$. Differentiating formally the function $\boldsymbol{\alpha}$ and $\boldsymbol{\mu}$ there holds :

$$\boldsymbol{\alpha}(f) = \boldsymbol{\alpha}(\mathcal{M}) + \epsilon d\boldsymbol{\alpha}_{\mathcal{M}}(g) + O(\epsilon^2), \quad \boldsymbol{\mu}(f) = \boldsymbol{\mu}(\mathcal{M}) + \epsilon d\boldsymbol{\mu}_{\mathcal{M}}(g) + O(\epsilon^2)$$

Then we compute for $f = \mathcal{M}(1+\epsilon g)$ the following approximation:

$$\begin{aligned}
\exp(\boldsymbol{\mu}(f)\cdot\mathbf{a}(\mathbf{v}-\mathbf{u})) &= \exp(\boldsymbol{\mu}(\mathcal{M})\cdot\mathbf{a}(\mathbf{v}-\mathbf{u}))\,(1+\epsilon d\boldsymbol{\mu}_{\mathcal{M}}(g)\cdot\mathbf{a}(\mathbf{v}-\mathbf{u})+o(\epsilon)) \\
&= \mathcal{M}(1+\epsilon d\boldsymbol{\mu}_{\mathcal{M}}(g)\cdot\mathbf{a}(\mathbf{v}-\mathbf{u})+o(\epsilon)) \\
\phi^{*\prime}(\boldsymbol{\alpha}(f)\cdot\mathbf{a}(\mathbf{v}-\mathbf{u})) &= \phi^{*\prime}(\boldsymbol{\alpha}(\mathcal{M})\cdot\mathbf{a}(\mathbf{v}-\mathbf{u})+o(\epsilon)) \\
&\quad +\epsilon d\boldsymbol{\alpha}_{\mathcal{M}}(g)\cdot\mathbf{a}(\mathbf{v}-\mathbf{u})\,\phi^{*\prime\prime}(\boldsymbol{\alpha}(\mathcal{M})\cdot\mathbf{a}(\mathbf{v}-\mathbf{u})+o(\epsilon)).
\end{aligned}$$

so that

$$\begin{aligned}
G[f](\mathbf{v}) &= \mathcal{M}(\mathbf{v})\,(1+\epsilon\,(d\boldsymbol{\mu}_{\mathcal{M}}(g) + c\,d\boldsymbol{\alpha}_{\mathcal{M}}(g))\cdot\mathbf{a}(\mathbf{v}-\mathbf{u})+o(\epsilon)) \\
&= \mathcal{M}(\mathbf{v})\,(1+\epsilon\Lambda_{\mathcal{M}}(g)\cdot\mathbf{a}(\mathbf{v}-\mathbf{u})+o(\epsilon))
\end{aligned}$$

Finally, there is

$$K(\mathcal{M}(1+\epsilon g)) = = \nu\,[\mathcal{M}+\epsilon\Lambda_{\mathcal{M}}(g)\cdot\mathbf{a}(\mathbf{v}-\mathbf{u})\,\mathcal{M}(\mathbf{v}) - \mathcal{M}(1+\epsilon g)+o(\epsilon)]$$

and by definition of $\mathcal{L}$ (6)

$$\mathcal{L}(g) = \nu\left(\Lambda_{\mathcal{M}}(g) \cdot \mathbf{a}(\mathbf{v}-\mathbf{u}) - g\right). \tag{72}$$

It is convenient to write this relaxation equation with a_i being a scalar function of $\mathbf{v}-\mathbf{u}$ rather than tensors. We need now to focus on $\Lambda_{\mathcal{M}}(g)$ in order to identify the expression of $\mathcal{L}_{\mathcal{M}}$. To do that we use the prescribe condition on moment:

$$\int G\left[\mathcal{M}(1+\epsilon g)\right]\mathbf{a}_i(\mathbf{v}-\mathbf{u}) = \left(1-\frac{\nu_i}{\nu}\right)\int \mathcal{M}(1+\epsilon g)\,\mathbf{a}_i(\mathbf{v}-\mathbf{u})$$

So with $K(f) = \nu(G[f]-f)$ we conclude that:

$$\int K\left(\mathcal{M}(1+\epsilon g)\right)\mathbf{a}_i(\mathbf{v}-\mathbf{u}) = -\epsilon\nu_i\int g\mathcal{M}(\mathbf{v})\,\mathbf{a}_i(\mathbf{v}-\mathbf{u})$$

using the linear approximation for K we have

$$\epsilon\int \nu\left(\Lambda_{\mathcal{M}}(g)\cdot\mathbf{a}(\mathbf{v}-\mathbf{u})-g\right)\mathbf{a}_i(\mathbf{v}-\mathbf{u})\,\mathcal{M} = -\epsilon\nu_i\int g\mathcal{M}(\mathbf{v})\,\mathbf{a}_i(\mathbf{v}-\mathbf{u}).$$

That is finally for any component i we have:

$$\int\left(\Lambda_{\mathcal{M}}(g)\cdot\mathbf{a}(\mathbf{v}-\mathbf{u})\right)\mathbf{a}_i(\mathbf{v}-\mathbf{u})\,\mathcal{M} = \left(1-\frac{\nu_i}{\nu}\right)\int g\mathcal{M}(\mathbf{v})\,\mathbf{a}_i(\mathbf{v}-\mathbf{u})$$

by expanding the dot product, we have formally:

$$\sum_j\int\left(\Lambda^j_{\mathcal{M}}(g)\cdot\mathbf{a}_j(\mathbf{v}-\mathbf{u})\right)\mathbf{a}_i(\mathbf{v}-\mathbf{u})\,\mathcal{M} = \left(1-\frac{\nu_i}{\nu}\right)\int g\mathcal{M}(\mathbf{v})\,\mathbf{a}_i(\mathbf{v}-\mathbf{u}).$$

So, by using orthogonality relations

$$\Lambda^i_{\mathcal{M}}\|\mathbf{a}_i\|^2 = (1-\frac{\nu_i}{\nu})\int_{\mathbb{R}^3} g\mathcal{M}\mathbf{a}_i(\mathbf{v}-\mathbf{u}).$$

Then, according to formula (72) we conclude that

$$\mathcal{L}g = \nu\left(\sum_i\left(1-\frac{\nu_i}{\nu}\right)\mathcal{P}_{\mathbf{a}_i}g - g\right)$$

where $\nu_i = 0$ if $a_i \in \mathbb{K}$. From this, it is easy to see that $\mathcal{L}$ is self adjoint, Fredholm with $Ker(\mathcal{L}) = \mathbb{K}$.
We end this proof by showing that

$$\forall f,\ \left[\int K(f)\,\phi = 0\right] \Leftrightarrow \phi \in \mathbb{K}$$

The right implication is just a consequence of (10). Assume now that

$$\forall f, \quad \int K\left(f\right)\phi(\mathbf{v})\,d\mathbf{v} = 0$$

Expanding $K(\mathcal{M}(1+\varepsilon g))$, we have for any $g \in L^2(\mathcal{M})$

$$\forall g, \quad \int \mathcal{M}\,\mathcal{L}\left(g\right)\phi(\mathbf{v})\,d\mathbf{v} = 0$$

Since $\mathcal{L}$ is self adjoint, we have

$$\forall g, \quad \int \mathcal{M}g\mathcal{L}\left(\phi\right)\,d\mathbf{v} = 0$$

which proves that $\phi \in \mathbb{K}$. □

7 Conclusion

In the present article, we have proposed a methodology to construct relaxation operators. The derivation is performed in three steps. We first consider the projection of the inverse linearized Boltzmann operator $\mathcal{L}_B^{-1}$ on a polynomial space of finite dimension. We then state relaxation equations on the moments of the probability distribution f basing on its diagonalization. The model must satisfy those equations together with the conservation laws. From this one derives linear relations between the moments of f and the target function G to be found. The later is then found by solving a variational problem. Different mathematical problems related to this construction have been addressed. We have firstly revisited a theorem by Junk [30] relating realizable moments (i.e moments of nonnegative integrable functions) to nonnegative polynomials. From this we have derived necessary conditions for the realizability of the moments of G and proved that it allows to specify the admissible relaxation equations on the Grad thirteen moments. The variational problem has been studied in detail by using different functional to be minimized under moment constraints. We have reestablished a theorem of Csiszar [20] on the existence of solution to such minimization problems by using convex analysis and exactly derived the shape of the solution by duality. In the last part of the article, we have proposed different models from this construction and analyzed their well-posedness. In particular, when relaxations occur on the Grad thirteen moments, the model satisfies almost all properties of the original Boltzmann equation: nonnegativity of the solution, conservation laws, H theorem, Galilean invariance and the right hydrodynamic limit up to Navier-Stokes level. However, the control of the entropy defined by the ϕ-divergence is only local. In the general case, those properties are also preserved but the control of the entropy is not yet proved. Finally, the present approach encompasses the derivation of many known models and for some of them their generalization.

There are many perspectives and questions related to this work. In principle the new model based only on Grad thirteen moments should not bring more than the

ESBGK or Shakhov models. It remains however to compare them. We also intend to study the generalization of those relaxation operators beyond the Grad case. In such cases, the present method does not require the effective computation of the relaxation operator if moment methods such as in [2, 33] are used. The computation of the approximate inverse linearized Boltzmann operator is of the same order of complexity than that of the transport coefficients for multicomponant fluids for which there exists plenty efficient methods. One may then in a first time compare from a numerical point of view the solution of this general model to that of the linearized Boltzmann equation and in a second time compare it to that of the known relaxation models and to the Boltzmann equation itself. Also, some study related to existence of solutions to the generalized Shakhov model can be addressed the framework of Bae and Yun [5].

Data Availability. No data was used for the research described in the article.

Conflict of Interest. The authors declare that thy hav no conflicts of interest.

Acknowledgements. The authors would like to thank J.J. Alibert from the laboratory IMATH (university of Toulon) for helpful discussions in the approach of the variational problem.

References

[1] Abdelmalik, M., Cai, M., Pichard, T.: On the Renormalization Maps for the phi-Divergence Moment Closures Applied in Radiative Transfer, J.Comput. Theor. Transport **52**, 6, 399-428 (2023)

[2] Abdelmalik, M. R. A., van Brummelen, H.: Moment closure approximations of the Boltzmann equation based on ϕ-divergences., J. Stat. Phys. **164**, 77-104 (2016)

[3] Akhiezer, N.I.: The classical moment problem, Amsterdam (1965)

[4] Andries, P., Le Tallec, P., Perlat, J.P., Perthame, B.: The Gaussian-BGK Model of Boltzmann Equation with Small Prandtl Number, Eur. J. Mech. B Fluids. **19**, 813-830 (2000)

[5] Bae, G., Yun, S.B.: The Shakhov model near global Maxwellian, Nonlinear analysis real world application **52**, 6, 399-428 (2023)

[6] Benoist, O.: Writing positive polynomials as sums of (few) squares, EMS. Newsletter, 8-13 (2017)

[7] Bhatnagar, P. L. and Gross, E. P. and Krook, M.: A Model for Collision Processes in Gases. I. Small Amplitude Processes in Charged and Neutral One-Component Systems, Physical Review **94**, 3, 511-525 (1954)

[8] Borwein, J.M., Lewis, S.: Duality relationships for entropy-like minimization problems, SIAM J. Control Optim. **29**, 2, 325-338 (1991)

[9] Bouchut, F., Perthame, B.: A BGK model for small Prandtl number in the Navier-Stokes approximation, J. Statist. Phys. **71**, 191-207 (1993)

[10] Bourgat, J.F., Desvillettes, L., Le Tallec, P., Perthame, B.: Microreversible collisions for polyatomic gases and Boltzmann's theorem , European J. Mech. B Fluids **13**, 2, 237-254 (1994)

[11] Brull, S., Pavan, V., Schneider, J.: Derivation of a BGK model for mixture, European Journal of Mechanics - B/Fluids **33**, 74-86 (2012)

[12] Brull, S., Schneider, J.: A new approach of the Ellipsoidal Statistical Model Cont. Mech. Thermodyn. **20**, 63-74 (2008)

[13] Brull, S., Schneider, J.: On the Ellipsoidal Statistical Model for polyatomic gases, Cont. Mech. Thermodyn. **20**, 489-508 (2009)

[14] Brull, S.: An Ellipsoidal Statistical Model for gas mixtures, Comm. Math. Sci. **13**, 1-13 (2015)

[15] Brull, S.: An Ellipsoidal Statistical Model for a monoatomic and polyatomic gas mixture, Comm. Math. Sci. **19**, 2177-2194 (2021)

[16] Cercignani C.: The Boltzmann Equation and its Application, Springer Verlag (1992)

[17] Chapman, S., Cowling, T.G.: The Mathematical Theory of non-uniform Gases, Third edition, Cambridge University Press (1990)

[18] Csiszar, I.: A class of measures of informativity of observation channels, Period. Math. Hung. **2**, 191-213 (1972)

[19] Csiszar, I.: I divergence geometry of probability distribution function and minimization problems Sanov property, Ann. Probab. **3**, 146-158 (1975)

[20] Csiszar, I.: Generalized projections for nonnegative functions, Acta. Math. Hung. **68**, 161-185 (1995)

[21] Curto, R., Fialkow, L: Recursiveness, positivity, and truncated moment problem, Houton J. Math. **17**, 4, 603-635 (1991)

[22] Curto, R., Fialkow, L: A duality proof of Tchakalff's theorem, J. Math. Anal. Appl. **262**, 519-532 (2002)

[23] Curto, R., Fialkow, L.: Truncated K-moment problems in several variables, J. Functional Analysis **54**, 1, 189-226 (2005)

[24] Curto, R., Fialkow, L.: An analogue of the Riesz-Haviland theorem for the truncated moment problem, J. Functional Analysis **255**, 2709-2731 (2008)

[25] Fialkow, L.:The truncated K-moment problem: a Survey., Theta Ser. Adv. Math. **18**, 25-51 (2016)

[26] Hauck, C., Levermore, C.D., Tits, A.L.: Convex duality and entropy-based moment closures: characterizing degenerate densities, SIAM J. Control Optimal **47**, 4, 1977-2015 (2008)

[27] Haviland, E.K.: On the momentum problem for distributions in more than one dimension ii, Amer. J. Math. **57**, 3, 562-568 (1935)

[28] Haviland, E.K.: On the momentum problem for distributions in more than one dimension, Amer. J. Math. **58**, 1, 164-168 (1936)

[29] Holway, L. H.: New statistical models for kinetic theory: methods of construction, Phys. Fluids. **9**, 1658-1673 (1966)

[30] Junk M.: Maximum entropy for reduced moment problems., Math. Models and Methods in Appl. Sci. **10**, 7, 1001-1025 (2000)

[31] Junk, M., Unterreiter, A.: A maximum entropy moment systems and Galilean invariance., Contin. Mech. Thermodyn. **14**, 6, 563-576 (2002)

[32] Lasserre, J.B.: Moments, Positive Polynomials and Their Applications, Imperial College Press, London (2009)

[33] Levermore, C. D.: Moment closure hierarchies for kinetic theories., J. Stat. Phys. **83**, 1021-1065 (1996)

[34] Mieussens L., Struchtrup H.: Numerical comparison of BGK-models with proper Prandtl number, Phys. Fluids **16**, 8, 2797-2813 (2004)

[35] Pavan, V.: General entropic approximations for canonical systems described by kinetic equations, J. Stat. Phys **142**, 4, 792-827 (2011)

[36] Pichard, T.: A moment closure based on a projection on the boundary of the realizability domain: 1D case, Kin. Rel. Models 13, 6, 1243-1280 (2020)

[37] Pichard, T.: A moment closure based on a projection on the boundary of the realizability domain: Extension and analysis, Kin. Rel. Models **15**, 5, 793-822 (2022)

[38] Rockafellar, R.T.: Convex analysis, Princeton University Press (1970)

[39] Schneider, J.: Entropic approximation in kinetic theory, Math. Model. Numer. Anal. **38**, 3, 541-561 (2004)

[40] Schneider, J.: A well-posed simulation model for multicomponent reacting gases, Comm. Math. Sci. **13**, 5, 1075-1103 (2015)

[41] Shakhov, E. M.: Generalization of the Krook kinetic relaxation equation, Fluid Dynamics. **3**, 95-96 (1968)

[42] Struchtrup, H.: The BGK model with velocity dependant collision frequency, Cont. Mech. Thermodyn. **9**, 23-31 (2004)

[43] Struchtrup, H., Zheng, Y.: Ellipsoidal Statistical Bhatnagar-Gross-Krook model with velocity-dependent collision frequency, Phys. Fluids. **12** (2005)