A CONSERVATIVE AND ENTROPIC DISCRETE-VELOCITY MODEL FOR RAREFIED POLYATOMIC GASES *

DUBROCA Bruno[1] and MIEUSSENS Luc[2]

Abstract. In this paper, a new discrete-velocity model is proposed for modeling rarefied polyatomic gases. This model is based on a polyatomic BGK equation that allows for the internal energy of molecules. For computational storage requirements, it is attractive to eliminate this new variable by using a reduced distribution technique. Then the BGK equation reduces to a system of two relaxation kinetic equations. Using the minimum entropy principle to define a discrete equilibrium function, a velocity discretization of this system is proposed. This discretization satisfies the conservation laws and the entropy dissipation. Owing to these properties, we believe that this discrete-velocity model can be used to design fast and robust numerical methods for computing rarefied polyatomic gas flows.

1. Introduction

As compared to a monoatomic molecule, a polyatomic one possesses additional degrees of freedom, such as the atom position in the molecules, the angular momentum, and the vibration energy. Therefore a complete description of a polyatomic gas is very complex but fortunately several simplifications are possible. For instance, the atom positions change more rapidly than the interesting time scale, permitting to consider the distribution as independent of these variables (see [8]). Moreover, it has been proved by experiments (see [9]) that the orientation of angular momentum can be neglected. Its magnitude only, or equivalently its rotational energy, has to be taken into account. Eventually, the vibration energy, which is quantified, can be neglected if the temperature of the gas is not too high.

With these simplifications, the description of a polyatomic gas requires only one additional variable - the rotational energy $\varepsilon > 0$, which will be supposed continuous - and a parameter δ equal to the number of rotational internal degrees of freedom. For instance, a diatomic molecule can rotate in three directions, but since the rotation about the line joining the two atoms has a negligible contribution to the rotational energy, this gives $\delta = 2$.

Modeling the energy exchange in polyatomic gases is a difficult problem, still investigated in both numerics and physics (see for instance [10] and the references mentioned). However, if one only allows for the rotational energy exchange, as described above, the problem is much simpler. The most frequently used model is the statistical model of Borgnakke-Larsen [5] that enables Direct Simulation Monte-Carlo (DSMC) computations for polyatomic gases. Many numerical results using this model have proved the previous simplifications to be relevant in numerous situations (see [4, 11]). Recently, a deterministic interpretation of the Borgnakke-Larsen model has been proposed by Bourgat et al. [6] in terms of Boltzmann collision operator.

* *This work was proposed and supported by the CEA-CESTA.*

[1] CEA-CESTA, DEV/SIS, 33114 Le Barp BP2, France

[2] Mathématiques pour l'Industrie et la Physique, UFR MIG, Université Paul Sabatier Toulouse 3, 118 route de Narbonne, 31062 TOULOUSE cedex 4, France

Concerning deterministic simulation, the work of [6] has been used by Buet [7]. This author adapts a former work on velocity discretization of monoatomic Boltzmann equation to the model of [6]. He obtains an entropic and conservative discrete-velocity model, and proposes some stochastic procedures to reduce the quadratic cost of the algorithms. However, the increase of the computational cost due to the additional variable is so large that deterministic numerical simulations are very expensive. As a matter of fact, to our knowledge, the only deterministic computations obtained are from Bahy [1] who used an extension of Rogier-Schneider's scheme [17] to polyatomic gases.

It is thus clear that there is a need for simpler deterministic methods for polyatomic gases. It has been proposed by L. Mieussens [13] a numerical method based on a conservative and entropic discretization of the BGK equation coupled with an implicit scheme, for monoatomic gases. Despite the simplicity of the BGK model, several results have been obtained which compare well to DSMC, both for accuracy and CPU time. This method is also able to easily describe flows that are hardly computed by DSMC (such as recirculation zones or near continuum flows). As a first step towards an extension of this method to polyatomic gases, we propose in this paper a robust velocity discretization of a polyatomic BGK model derived from the polyatomic Boltzmann equation of [6]. This discretization uses the minimum entropy principle to define a discrete equilibrium function. A mathematical proof for the existence of the discrete equilibrium is given. As in the monoatomic case, we obtain a discrete-velocity model which is conservative and entropic.

2. A polyatomic BGK model

As explained above, we consider a gas as described by the mass density $F(t,x,v,I)$ of molecules that at time t have position $x \in \mathbb{R}^D$, velocity $v \in \mathbb{R}^D$, and internal energy $\varepsilon(I) = I^{2/\delta}$, where δ is the number of internal degrees of freedom. Classical fluid quantities as mass density ρ, mean velocity $u = (u^{(1)}, ..., u^{(D)})$, and total energy E are defined as the first $D+2$ moments of F with respect to v and I

$$\rho = \langle\langle F \rangle\rangle, \quad \rho u = \langle\langle vF \rangle\rangle, \quad E = \langle\langle (\tfrac{1}{2}|v|^2 + I^{2/\delta})F \rangle\rangle,$$

where $\langle\langle G \rangle\rangle = \int_{\mathbb{R}^D_v} \int_{\mathbb{R}^+} G \, dvdI$ denotes the integral of any function $G(v,I)$. The temperature θ of the gas is defined through the total energy by the relation $E = \frac{1}{2}\rho|u|^2 + \frac{D+\delta}{2}\rho\theta$.

The simplest BGK model can be obtained from the Boltzmann equation of [6]

$$\partial_t F + v \cdot \nabla_x F = \frac{1}{\tau}(\mathcal{M}[F] - F), \tag{1}$$

where the collision operator is replaced by a relaxation term of F towards the local Maxwellian equilibrium defined by

$$\mathcal{M}[F] = \frac{\rho}{(2\pi\theta)^{D/2}} \exp\Big(-\frac{|v-u|^2}{2\theta}\Big) \frac{\Lambda_\delta}{\theta^{\delta/2}} \exp\Big(-\frac{I^{2/\delta}}{\theta}\Big).$$

The parameter $\Lambda_\delta = (\frac{\delta}{2}\Gamma(\frac{\delta}{2}))^{-1}$ ensures that $\mathcal{M}[F]$ has the same first $D+2$ moments as F. The relaxation time is defined by $\tau = \frac{\mu}{\rho\theta}$ so as to recover the correct viscosity μ in the hydrodynamic limit. The value of μ is generally $c\theta^\omega$ where ω depends on the molecular interaction potential ($\omega = 0$ for Maxwellian molecules, $\omega = \frac{1}{2}$ for Hard-Spheres molecules).

Despite its simplicity, this model possesses numerous properties of the Boltzmann equation. Namely, it satisfies the local conservation laws of $(\rho, \rho u, E)$

$$\partial_t \langle\langle \begin{pmatrix} 1 \\ v \\ \frac{1}{2}|v|^2 + I^{2/\delta} \end{pmatrix} F \rangle\rangle + \nabla_x \cdot \langle\langle v \begin{pmatrix} 1 \\ v \\ \frac{1}{2}|v|^2 + I^{2/\delta} \end{pmatrix} F \rangle\rangle = 0,$$

the local dissipation of entropy $\mathcal{H}(F) = \langle\langle F \log F \rangle\rangle$

$$\partial_t \langle\langle F \log F \rangle\rangle + \nabla_x \cdot \langle\langle vF \log F \rangle\rangle \leq 0, \tag{2}$$

and Euler and Navier-Stokes equations in the hydrodynamic limit $\tau \ll 1$.

It is clear that this model has some physical drawbacks due to its single relaxation time. For instance, the heat transfer coefficient κ in the Navier-Stokes limit is found to be incorrect, which leads to a Prandtl number of 1 instead of $\frac{5}{7}$ for a diatomic gas ($\delta = 2$). Moreover the time scale for elastic and inelastic collisions (that respectively involve translational and rotational energy exchanges) is the same, which for instance leads to incorrect values of the volume viscosity (see [8]).

However, we find it very attractive to use this model. The main reason is that the corresponding monoatomic BGK model (which has the same physical drawbacks) turned out to be relevant for many cases, and it permits us to obtain very fast deterministic algorithms (see [13]). Moreover, modified BGK models exist that give more physical transport coefficients, as the Ellipsoidal-Statistical BGK model (ES-BGK) of Holway extended to polyatomic gases by Andriès et al. [3].

3. Reduced distribution technique

3.1. **Reduced BGK system**

The method of reduced distribution functions has been frequently used to reduce the dependency of the distribution on the z-component of the velocity for plane flow simulations (see [18]). The same idea has been used by Perthame [15] to eliminate the dependency of the distribution function on the internal energy variable I, in the context of kinetic schemes for polyatomic Euler equations. In a recent paper [3], this technique is used for the polyatomic ES-BGK model.

Now we present this technique as applied to BGK equation (1). We define the mass distribution $f(t, x, v)$ and the internal energy distribution $g(t, x, v)$ by

$$f(t, x, v) = \int_0^{+\infty} F(t, x, v, I)\, dI, \qquad g(t, x, v) = \int_0^{+\infty} I^{2/\delta} F(t, x, v, I)\, dI.$$

The macroscopic quantities ρ, ρu, E can be deduced from f and g as follows:

$$\rho = \langle f \rangle, \quad \rho u = \langle v f \rangle, \quad E = \langle \tfrac{1}{2}|v|^2 f + g \rangle,$$

where we set $\langle . \rangle = \int_{\mathbb{R}^D_v} . \, dv$. In order to use a more compact definition of these macroscopic quantities, we introduce the vector $\boldsymbol{\rho} = (\rho, \rho u, E) \in \mathbb{R}^{D+2}$, and we denote by $\boldsymbol{m}(v) = (1, v, \frac{1}{2}|v|^2) \in \mathbb{R}^{D+2}$ the vector of mass (normalized to 1), momentum and translational energy of molecules. We also set $\mathbf{e}_{D+1} = (0, \dots, 0, 1)$ the last vector of the canonical basis of $\mathbb{R}^{D+2}$, then $\boldsymbol{\rho}$ is defined by

$$\boldsymbol{\rho} = \langle \boldsymbol{m} f + \mathbf{e}_{D+1} g \rangle.$$

Now integrating BGK equation (1) with measures dI and $I^{\delta/2} dI$ yields the system

$$\begin{aligned} \partial_t f + v \cdot \nabla_x f &= \frac{1}{\tau}(M^{tr}[\boldsymbol{\rho}] - f), \\ \partial_t g + v \cdot \nabla_x g &= \frac{1}{\tau}(M^{int}[\boldsymbol{\rho}] - g), \end{aligned} \tag{3}$$

where $(M^{tr}[\boldsymbol{\rho}], M^{int}[\boldsymbol{\rho}])$ are the translational and internal equilibrium defined by

$$M^{tr}[\boldsymbol{\rho}] = M[\boldsymbol{\rho}], \qquad M^{int}[\boldsymbol{\rho}] = \tfrac{\delta}{2}\theta M[\boldsymbol{\rho}], \tag{4}$$

θ is the temperature (see Sec. 2), and $M[\boldsymbol{\rho}]$ is the usual Maxwellian

$$M[\boldsymbol{\rho}] = \frac{\rho}{(2\pi\theta)^{D/2}} \exp\Big(-\frac{|v-u|^2}{2\theta}\Big).$$

The remarkable feature of this system is that it gives *exact* values – *i.e.* as defined by $F(t,x,v,I)$ – of the physical relevant macroscopic quantities ρ, ρu, E, stress tensor $\Sigma = \langle\langle (v-u)\otimes(v-u)F\rangle\rangle$ and heat flux $q = \langle\langle \frac{1}{2}|v-u|^2(v-u)F\rangle\rangle$. We point out that this is a consequence of the structure of the BGK collision operator. This cannot be done with the Boltzmann collision operator.

3.2. Reduced kinetic entropy

A method to obtain a kinetic entropy for system (3) is described hereafter. The result has been proposed by Perthame in [16] but we find it instructive to present its derivation.

The problem is to find a macroscopic functional $H(f,g)$ which is locally dissipated by (3), and such that the equilibrium $(M^{tr}[\boldsymbol{\rho}], M^{int}[\boldsymbol{\rho}])$ is its unique minimum, subject to the realization of macroscopic quantities $\boldsymbol{\rho}$. Obviously, $H(f,g)$ should be related to the entropy $\mathcal{H}(F)$ (see (2)). In fact, system (3) may be viewed as an entropic moment closure of (1) with respect to the internal energy variable I. Thus following [12], it is natural to set $H(f,g) = \langle S(f,g)\rangle$, where

$$S(f,g) = \min\Big\{\mathcal{H}(G), G \geq 0 \text{ s.t. } \int_0^{+\infty} \begin{pmatrix} 1 \\ I^{2/\delta} \end{pmatrix} G\, dI = \begin{pmatrix} f \\ g \end{pmatrix}\Big\}.$$

This problem can be solved analytically by Lagrange multiplier method, and we find

$$S(f,g) = f\Big((1+\tfrac{\delta}{2})\log\frac{f}{g^{\frac{\delta}{\delta+2}}} + \Lambda_\delta - 1 - \tfrac{\delta}{2} + \tfrac{\delta}{2}\log\tfrac{\delta}{2}\Big),$$

which gives $H(f,g)$. By dropping constant coefficients, we obtain the following definition

$$H(f,g) = \langle f\log\frac{f}{g^{\frac{\delta}{\delta+2}}} - f\rangle. \tag{5}$$

It is now easy to prove the following result (see [16])

Proposition 3.1. *The entropy $H(f,g)$ defined by (5) is strictly convex, and $(M^{tr}[\boldsymbol{\rho}], M^{int}[\boldsymbol{\rho}])$ is the unique solution of the entropy minimization problem*

$$(\mathcal{P}) \qquad H(M^{tr}[\boldsymbol{\rho}], M^{int}[\boldsymbol{\rho}]) = \min_{\mathcal{X}_{\boldsymbol{\rho}}}\Big\{H(\tilde{f},\tilde{g})\Big\},$$
$$\textit{with } \mathcal{X}_{\boldsymbol{\rho}} = \Big\{\tilde{f}\geq 0 \textit{ and } \tilde{g} > 0 \textit{ s.t. } \langle \boldsymbol{m}\tilde{f} + \mathbf{e}_{D+1}\tilde{g}\rangle = \boldsymbol{\rho}\Big\}. \tag{6}$$

Moreover, $H(f,g)$ is locally dissipated by (3)

$$\partial_t\langle f\log\frac{f}{g^{\frac{\delta}{\delta+2}}} - f\rangle + \nabla_x\cdot\langle v(f\log\frac{f}{g^{\frac{\delta}{\delta+2}}} - f)\rangle \leq 0. \tag{7}$$

Remark 3.2. It is worth mentioning another expression of the local equilibrium $(M^{tr}[\boldsymbol{\rho}], M^{int}[\boldsymbol{\rho}])$. If we write the local Maxwellian under the form used in [13]

$$M[\boldsymbol{\rho}] = \exp(\boldsymbol{\alpha}\cdot\boldsymbol{m}(v))$$

where $\boldsymbol{\alpha} = (\alpha^{(i)})_{i=0}^{D+1} = \left(\log\left(\frac{\rho}{(2\pi\theta)^{D/2}}\right) - \frac{|u|^2}{2\theta}, \frac{u}{\theta}, -\frac{1}{\theta}\right) \in \mathbb{R}^{D+2}$, then the translational and internal equilibrium given in (4) are defined by

$$M^{tr}[\boldsymbol{\rho}] = \exp(\boldsymbol{\alpha}\cdot \boldsymbol{m}(v)) \quad \text{and} \quad M^{int}[\boldsymbol{\rho}] = \tfrac{\delta}{2}[-\frac{1}{\alpha^{(D+1)}}]\exp(\boldsymbol{\alpha}\cdot \boldsymbol{m}(v)). \tag{8}$$

The vector $\boldsymbol{\alpha}$ is related to the Lagrange multipliers of the entropy minimization problem $(\mathcal{P})$. This will be used in the next section (for the discrete-velocity case).

4. Velocity discretization of the reduced BGK model

The velocity space $\mathbb{R}^D$ is replaced by a bounded Cartesian grid $\mathcal{V} = \{v_k = k\Delta v, k \in \mathcal{K}\}$ of N points, where $\mathcal{K}$ is a multi-index set of $\mathbb{Z}^D$, and $\Delta v \in \mathbb{R}$ is the step of the grid. The continuous-in-velocity distribution functions f and g are approximated on this grid by two vectors $f_{\mathcal{K}} = (f_k)_{k\in\mathcal{K}}$ and $g_{\mathcal{K}} = (g_k)_{k\in\mathcal{K}}$ with N components. These vectors are called discrete distribution functions. Each component $f_k(t,x)$ and $g_k(t,x)$ is assumed to be an approximation of $f(t,x,v_k)$ and $g(t,x,v_k)$.

The vector of macroscopic quantities $\boldsymbol{\rho}$ is approximated by the vector of moments $\boldsymbol{\rho}_{\mathcal{K}}$ of $(f_{\mathcal{K}}, g_{\mathcal{K}})$, which is defined by a discrete sum on the velocity grid. That is, setting $\langle \tilde{f}\rangle_{\mathcal{K}} = \sum_{k\in\mathcal{K}} \tilde{f}_k\,\Delta v^D$ for any vector $\tilde{f}\in\mathbb{R}^N$, we have

$$\boldsymbol{\rho}_{\mathcal{K}} = \langle \boldsymbol{m} f_{\mathcal{K}} + \mathbf{e}_{D+1} g_{\mathcal{K}}\rangle_{\mathcal{K}}.$$

We also define a discrete entropy in the same manner

$$H_{\mathcal{K}}(f_{\mathcal{K}}, g_{\mathcal{K}}) = \langle f_{\mathcal{K}} \log \frac{f_{\mathcal{K}}}{g_{\mathcal{K}}^{\frac{\delta}{\delta+2}}} - f_{\mathcal{K}}\rangle_{\mathcal{K}}.$$

For the approximation of system (3), the transport terms are just evaluated at each point v_k of the grid, and we define some discrete approximation $(M^{tr}_{\mathcal{K}}[\boldsymbol{\rho}_{\mathcal{K}}], M^{int}_{\mathcal{K}}[\boldsymbol{\rho}_{\mathcal{K}}])$ of the equilibrium function $(M^{tr}[\boldsymbol{\rho}], M^{int}[\boldsymbol{\rho}])$. Thus we obtain the discrete-velocity model of (3)

$$\begin{aligned} \partial_t f_k + v_k\cdot\nabla_x f_k &= \frac{1}{\tau}(M^{tr}_k[\boldsymbol{\rho}_{\mathcal{K}}] - f_k), \\ \partial_t g_k + v_k\cdot\nabla_x g_k &= \frac{1}{\tau}(M^{int}_k[\boldsymbol{\rho}_{\mathcal{K}}] - g_k). \end{aligned} \tag{9}$$

Now the main problem is to obtain an approximation $(M^{tr}_{\mathcal{K}}[\boldsymbol{\rho}_{\mathcal{K}}], M^{int}_{\mathcal{K}}[\boldsymbol{\rho}_{\mathcal{K}}])$ of $(M^{tr}[\boldsymbol{\rho}], M^{int}[\boldsymbol{\rho}])$ such that system (9) is conservative and entropic. Following the idea of [13], we use the discrete version of the minimum entropy principle (6)

$$\begin{aligned} (\mathcal{P}_{\mathcal{K}}) \qquad H_{\mathcal{K}}(M^{tr}_{\mathcal{K}}[\boldsymbol{\rho}_{\mathcal{K}}], M^{int}_{\mathcal{K}}[\boldsymbol{\rho}_{\mathcal{K}}]) &= \min_{\mathcal{X}_{\boldsymbol{\rho}_{\mathcal{K}}}} \Big\{ H_{\mathcal{K}}(\tilde{f}, \tilde{g})\Big\}, \\ \text{with } \mathcal{X}_{\boldsymbol{\rho}_{\mathcal{K}}} &= \Big\{ \tilde{f} \geq 0 \text{ and } \tilde{g} > 0 \in \mathbb{R}^N \text{ s.t. } \langle \boldsymbol{m}\tilde{f} + \mathbf{e}_{D+1}\tilde{g}\rangle_{\mathcal{K}} = \boldsymbol{\rho}_{\mathcal{K}}\Big\}. \end{aligned} \tag{10}$$

Therefore, this approximation will be called discrete equilibrium. From this definition, it immediately follows that the discrete-velocity reduced system (9) is conservative and entropic.

Obviously, it must be checked that this problem has a unique and easily solvable solution (solving directly $(\mathcal{P}_{\mathcal{K}})$ in $\mathbb{R}^N\times\mathbb{R}^N$ would be numerically expensive). In the continuous case, the condition $\rho, \theta > 0$ is sufficient to characterize the solution of (6) by the distribution $(M^{tr}[\boldsymbol{\rho}], M^{int}[\boldsymbol{\rho}]) = (M[\boldsymbol{\rho}], \frac{\delta}{2}\theta M[\boldsymbol{\rho}])$. In the following theorem, we prove that under a condition on $\boldsymbol{\rho}_{\mathcal{K}}$, more complex than $\rho_{\mathcal{K}}, \theta_{\mathcal{K}} > 0$, and under a natural assumption on $\mathcal{V}$, the discrete equilibrium $(M^{tr}_{\mathcal{K}}[\boldsymbol{\rho}_{\mathcal{K}}], M^{int}_{\mathcal{K}}[\boldsymbol{\rho}_{\mathcal{K}}])$ has a similar form. This result is similar to the result obtained for a monoatomic gas in [13], but the proof is rather more complicated.

Theorem 4.1. *Let $\boldsymbol{\rho}_{\mathcal{K}}$ be a vector in $\mathbb{R}^{D+2}$. Assume that the grid $\mathcal{V}$ is such that*

$$\{\boldsymbol{m}(v_k), k \in \mathcal{K}\} \text{ is of rank } D+2, \tag{11}$$

then the following assertions are equivalent

(i) *problem $(\mathcal{P}_{\mathcal{K}})$ has a unique solution $(M^{tr}_{\mathcal{K}}[\boldsymbol{\rho}_{\mathcal{K}}], M^{int}_{\mathcal{K}}[\boldsymbol{\rho}_{\mathcal{K}}])$, and there exists a unique vector $\boldsymbol{\alpha}_{\mathcal{K}} \in \mathbb{R}^{D+2}$ such that*

$$M^{tr}_k[\boldsymbol{\rho}_{\mathcal{K}}] = \exp(\boldsymbol{\alpha}_{\mathcal{K}} \cdot \boldsymbol{m}(v_k)) \quad \text{and} \quad M^{int}_k[\boldsymbol{\rho}_{\mathcal{K}}] = \tfrac{\delta}{2}[-\frac{1}{\boldsymbol{\alpha}_{\mathcal{K}}^{(D+1)}}]M^{tr}_k[\boldsymbol{\rho}_{\mathcal{K}}] \tag{12}$$

for any $k \in \mathcal{K}$.

(ii) *$\boldsymbol{\rho}_{\mathcal{K}}$ is strictly realizable on $\mathcal{V}$,* i.e.

$$\exists(\tilde{f}, \tilde{g}) \in \mathcal{X}_{\boldsymbol{\rho}_{\mathcal{K}}} \text{ such that } \tilde{f}, \tilde{g} > 0. \tag{13}$$

Remark 4.2. The exponential characterization (12) of the discrete equilibrium is to be compared to the similar expression of the continuous equilibrium (see (8)). The difference is that here, we do not have *a priori* any explicit relation between $\boldsymbol{\alpha}_{\mathcal{K}}$ and $\boldsymbol{\rho}_{\mathcal{K}}$. In particular $-\frac{1}{\boldsymbol{\alpha}_{\mathcal{K}}^{(D+1)}} \neq \theta_{\mathcal{K}}$.

Remark 4.3. Assumption (ii) of strict realizability is probably true if $\rho, \theta > 0$. Actually, it is possible to prove that in the most unfavorable case ($N = D+2$ with $D = 1$), (ii) and $\rho, \theta > 0$ are equivalent. But it is worth mentioning that this is not true in monoatomic case (see [13]). We believe that this difference is due to the additional degrees of freedom given by the internal energy distribution function g.

The model is now completely defined except in the case where $\boldsymbol{\rho}_{\mathcal{K}}$ is not strictly realizable. For that case, we can set $(M^{tr}_{\mathcal{K}}[\boldsymbol{\rho}_{\mathcal{K}}], M^{int}_{\mathcal{K}}[\boldsymbol{\rho}_{\mathcal{K}}]) = 0$. This has practically no interest because the model implicitly contains the fact that $\boldsymbol{\rho}_{\mathcal{K}}$ is realized by $(f_{\mathcal{K}}, g_{\mathcal{K}}) > 0$ if the initial condition is also strictly positive. We can now state that the model is well defined and has the expected properties.

Theorem 4.4. *Let (f^0, g^0) be a strictly positive vector of $\mathbb{R}^N \times \mathbb{R}^N$. Consider the initial value problem associated with model (3), where $(M^{tr}_{\mathcal{K}}[\boldsymbol{\rho}_{\mathcal{K}}], M^{int}_{\mathcal{K}}[\boldsymbol{\rho}_{\mathcal{K}}])$ is defined either by $(\mathcal{P}_{\mathcal{K}})$, or by 0 (if $\boldsymbol{\rho}_{\mathcal{K}}$ is not strictly realizable). If this problem has a solution $(f_{\mathcal{K}}, g_{\mathcal{K}})$, then we have (at least formally)*

$$f_k(t,x), g_k(t,x) > 0 \qquad \forall k, t, x, \tag{14}$$

$$(M^{tr}_{\mathcal{K}}[\boldsymbol{\rho}_{\mathcal{K}}], M^{int}_{\mathcal{K}}[\boldsymbol{\rho}_{\mathcal{K}}]) \text{ is defined by } (12) \tag{15}$$

$$\partial_t \langle \boldsymbol{m} f_{\mathcal{K}} + \mathbf{e}_{D+1} g_{\mathcal{K}} \rangle_{\mathcal{K}} + \nabla_x \langle v(\boldsymbol{m} f_{\mathcal{K}} + \mathbf{e}_{D+1} g_{\mathcal{K}}) \rangle_{\mathcal{K}} = 0, \tag{16}$$

$$\partial_t \langle f_{\mathcal{K}} \log \frac{f_{\mathcal{K}}}{g_{\mathcal{K}}^{\frac{\delta}{\delta+2}}} - f_{\mathcal{K}} \rangle_{\mathcal{K}} + \nabla_x \langle v(f_{\mathcal{K}} \log \frac{f_{\mathcal{K}}}{g_{\mathcal{K}}^{\frac{\delta}{\delta+2}}} - f_{\mathcal{K}}) \rangle_{\mathcal{K}} \leq 0. \tag{17}$$

where the last two relations are respectively local conservation law of $\boldsymbol{\rho}_{\mathcal{K}}$ and local dissipation of entropy $H_{\mathcal{K}}(f_{\mathcal{K}}, g_{\mathcal{K}})$.

Proof. The proof of assertions (14,16,17) is very similar to that of the continuous case, and is left to the reader. We only need to show that the discrete equilibrium is always an exponential. Noting $(f_{\mathcal{K}}, g_{\mathcal{K}})$ is always strictly positive proves that $\boldsymbol{\rho}_{\mathcal{K}}$ is obviously strictly realizable, and theorem 4.1 then allows us to conclude. $\square$

It remains to prove theorem 4.1, which is the subject of the next section.

5. Study of the discrete equilibrium

5.1. Formal calculation

Define the Lagrangian of the minimization problem (10)

$$\mathcal{L}(\tilde f,\tilde g,\boldsymbol{\gamma}) = \langle \tilde f \log\frac{\tilde f}{\tilde g^{\frac{\delta}{\delta+2}}} - \tilde f\rangle_{\mathcal{K}} - \boldsymbol{\gamma}\cdot\Big(\langle \boldsymbol{m}\tilde f + \mathbf{e}_{D+1}\tilde g\rangle_{\mathcal{K}} - \boldsymbol{\rho}_{\mathcal{K}}\Big),$$

where $(\tilde f,\tilde g)$ are some positive vectors of $\mathbb{R}^N$, and $\boldsymbol{\gamma}$ is a vector of $\mathbb{R}^{D+2}$. Then the solution of (10) can be formally computed by the following relations

$$\partial_{\tilde f}\mathcal{L}(M^{tr}_{\mathcal{K}}[\boldsymbol{\rho}_{\mathcal{K}}], M^{int}_{\mathcal{K}}[\boldsymbol{\rho}_{\mathcal{K}}],\boldsymbol{\beta})h = \langle\Big(\log\frac{M^{tr}_{\mathcal{K}}[\boldsymbol{\rho}_{\mathcal{K}}]}{M^{int}_{\mathcal{K}}[\boldsymbol{\rho}_{\mathcal{K}}]^{\frac{\delta}{\delta+2}}}\Big)h\rangle_{\mathcal{K}} - \boldsymbol{\beta}\cdot\langle \boldsymbol{m}h\rangle_{\mathcal{K}} = 0, \tag{18}$$

$$\partial_{\tilde g}\mathcal{L}(M^{tr}_{\mathcal{K}}[\boldsymbol{\rho}_{\mathcal{K}}], M^{int}_{\mathcal{K}}[\boldsymbol{\rho}_{\mathcal{K}}],\boldsymbol{\beta})h = \langle -\frac{\delta}{\delta+2}\frac{M^{tr}_{\mathcal{K}}[\boldsymbol{\rho}_{\mathcal{K}}]}{M^{int}_{\mathcal{K}}[\boldsymbol{\rho}_{\mathcal{K}}]}h\rangle_{\mathcal{K}} - \beta^{(D+1)}\langle h\rangle_{\mathcal{K}} = 0, \tag{19}$$

for any positive vector $h\in\mathbb{R}^N$, and where $\boldsymbol{\beta}$ is some vector of $\mathbb{R}^{D+2}$. Relation (19) yields

$$M^{int}_{\mathcal{K}}[\boldsymbol{\rho}_{\mathcal{K}}] = -\frac{\delta}{(2+\delta)\beta^{(D+1)}}M^{tr}_{\mathcal{K}}[\boldsymbol{\rho}_{\mathcal{K}}], \tag{20}$$

which implies that $\beta^{(D+1)}$ is necessarily strictly negative. Moreover we obtain with relation (18)

$$\log\frac{M^{tr}_{k}[\boldsymbol{\rho}_{\mathcal{K}}]}{M^{int}_{k}[\boldsymbol{\rho}_{\mathcal{K}}]^{\frac{\delta}{2+\delta}}} = \boldsymbol{\beta}\cdot\boldsymbol{m}(v_k), \qquad \forall k\in\mathcal{K}.$$

Combined with equality (20) and after some calculations, it comes

$$M^{tr}_{k}[\boldsymbol{\rho}_{\mathcal{K}}] = \exp\left(\frac{2+\delta}{2}\boldsymbol{\beta}\cdot\boldsymbol{m}(v_k) + \frac{\delta}{2}\log\Big(\frac{\delta}{2+\delta}[-\frac{1}{\beta^{(D+1)}}]\Big)\right), \qquad \forall k\in\mathcal{K}. \tag{21}$$

To obtain the nice expression (12), we use the nonlinear change of variables

$$\boldsymbol{\beta}\mapsto\boldsymbol{\alpha}_{\mathcal{K}} = \frac{2+\delta}{2}\boldsymbol{\beta} + \frac{\delta}{2}\log\Big(\frac{\delta}{2+\delta}[-\frac{1}{\beta^{(D+1)}}]\Big)\mathbf{e}_0, \tag{22}$$

where $\mathbf{e}_0 = (1,0,dotsc,0)$. Therefore, using (20–22) gives the expected result (12).

Remark 5.1. For the sequel, it is useful to invert (22), we find

$$\boldsymbol{\beta} = \frac{2}{2+\delta}\boldsymbol{\alpha}_{\mathcal{K}} - \frac{\delta}{2+\delta}\log\Big(\frac{\delta}{2}[-\frac{1}{\boldsymbol{\alpha}_{\mathcal{K}}^{(D+1)}}]\Big)\mathbf{e}_0 \tag{23}$$

5.2. Existence of $\boldsymbol{\alpha}$

Now we use the dual problem of the minimization problem (10), which is obtained as described below. Using (12) and (23), we rewrite the Lagrangian of the problem as a functional of $\boldsymbol{\alpha}_{\mathcal{K}}$

$$\begin{aligned}\mathcal{L}(M^{tr}_{\mathcal{K}}[\boldsymbol{\rho}_{\mathcal{K}}], M^{int}_{\mathcal{K}}[\boldsymbol{\rho}_{\mathcal{K}}],\boldsymbol{\alpha}_{\mathcal{K}}) = &-\frac{2}{2+\delta}\langle\exp(\boldsymbol{\alpha}_{\mathcal{K}}\cdot\boldsymbol{m})\rangle_{\mathcal{K}} + \frac{2}{2+\delta}\boldsymbol{\alpha}_{\mathcal{K}}\cdot\boldsymbol{\rho}_{\mathcal{K}}\\ &-\frac{\delta}{2+\delta}\log\Big(\frac{\delta}{2}[-\frac{1}{\alpha_{\mathcal{K}}^{D+1}}]\Big)\rho^{(0)}_{\mathcal{K}}\end{aligned}$$

The existence and uniqueness of $\boldsymbol{\alpha}_{\mathcal{K}}$ may then be obtained through the study of the maximum of this functional.

In fact we find it more convenient to simplify the last expression and to take the opposite. This leads to the following functional

$$J(\boldsymbol{\gamma}) = \langle \exp(\boldsymbol{\gamma}\cdot\boldsymbol{m})\rangle_{\mathcal{K}} - \boldsymbol{\gamma}\cdot\boldsymbol{\rho}_{\mathcal{K}} - \frac{\delta}{2}\log(-\frac{2}{\delta}\gamma^{(D+1)})\rho_{\mathcal{K}}^{(0)},$$

defined for any $\boldsymbol{\gamma}$ of $\mathcal{D} = \{\boldsymbol{\gamma}\in\mathbb{R}^{D+2}, \quad \gamma^{(D+1)} < 0\}$

Now, assuming $\boldsymbol{\rho}_{\mathcal{K}}$ is strictly realizable, we just have to prove that J admits a unique minimum $\boldsymbol{\alpha}_{\mathcal{K}}$ in $\mathcal{D}$. As a classical result of optimization, this is a direct consequence of the following lemma.

Lemma 5.2. *If $\boldsymbol{\rho}_{\mathcal{K}}$ is strictly realizable (in the sense of (13)), then J is strictly convex, and is coercive on $\mathcal{D}$,* i.e. *$\forall R>0$, $\exists B(R)$ and $\varepsilon(R)$ such that $\forall\boldsymbol{\gamma}\in\mathcal{D}$:*

$$\Big(|\gamma| \geq B(R) \quad or \quad 0 < -\gamma^{(D+1)} \leq \varepsilon(R)\Big) \Rightarrow J(\boldsymbol{\gamma}) > R.$$

Then the vector $\boldsymbol{\alpha}_{\mathcal{K}}$ allows us to define the discrete distributions $(M_{\mathcal{K}}^{tr}[\boldsymbol{\rho}_{\mathcal{K}}], M_{\mathcal{K}}^{int}[\boldsymbol{\rho}_{\mathcal{K}}])$ as in (12). Since $\boldsymbol{\alpha}_{\mathcal{K}}$ is the unique minimum of J in $\mathcal{D}$, then we have $J'(\boldsymbol{\alpha}_{\mathcal{K}}) = 0$, which reads

$$\langle \boldsymbol{m}\exp(\boldsymbol{\alpha}_{\mathcal{K}}\cdot\boldsymbol{m})\rangle_{\mathcal{K}} - \boldsymbol{\rho}_{\mathcal{K}} + \frac{\delta}{2}[-\frac{1}{\alpha_{\mathcal{K}}^{(D+1)}}]\rho_{\mathcal{K}}^{(0)}\mathbf{e}_{D+1} = 0.$$

This implies that $\langle \boldsymbol{m}M_{\mathcal{K}}^{tr}[\boldsymbol{\rho}_{\mathcal{K}}] + \mathbf{e}_{D+1}M_{\mathcal{K}}^{int}[\boldsymbol{\rho}_{\mathcal{K}}]\rangle_{\mathcal{K}} = \boldsymbol{\rho}_{\mathcal{K}}$, hence $(M_{\mathcal{K}}^{tr}[\boldsymbol{\rho}_{\mathcal{K}}], M_{\mathcal{K}}^{int}[\boldsymbol{\rho}_{\mathcal{K}}]) \in \mathcal{X}_{\boldsymbol{\rho}_{\mathcal{K}}}$. Finally, using this last property and the strict convexity of $H_{\mathcal{K}}$, we have for all $(\tilde f,\tilde g)\in\mathcal{X}_{\boldsymbol{\rho}_{\mathcal{K}}}$ different from $(M_{\mathcal{K}}^{tr}[\boldsymbol{\rho}_{\mathcal{K}}], M_{\mathcal{K}}^{int}[\boldsymbol{\rho}_{\mathcal{K}}])$

$$\begin{aligned}
H_{\mathcal{K}}(M_{\mathcal{K}}^{tr}[\boldsymbol{\rho}_{\mathcal{K}}], M_{\mathcal{K}}^{int}[\boldsymbol{\rho}_{\mathcal{K}}]) &< H_{\mathcal{K}}(\tilde f,\tilde g) \\
&\quad + H_{\mathcal{K}}'(M_{\mathcal{K}}^{tr}[\boldsymbol{\rho}_{\mathcal{K}}], M_{\mathcal{K}}^{int}[\boldsymbol{\rho}_{\mathcal{K}}])(M_{\mathcal{K}}^{tr}[\boldsymbol{\rho}_{\mathcal{K}}]-\tilde f, M_{\mathcal{K}}^{int}[\boldsymbol{\rho}_{\mathcal{K}}]-\tilde g) \\
&= H_{\mathcal{K}}(\tilde f,\tilde g) \\
&\quad + \frac{2}{2+\delta}\boldsymbol{\alpha}_{\mathcal{K}}\cdot\langle \boldsymbol{m}(M_{\mathcal{K}}^{tr}[\boldsymbol{\rho}_{\mathcal{K}}]-\tilde f) + \mathbf{e}_{D+1}(M_{\mathcal{K}}^{int}[\boldsymbol{\rho}_{\mathcal{K}}]-\tilde g)\rangle_{\mathcal{K}} \\
&\quad - \frac{\delta}{2+\delta}\log\Big(\frac{\delta}{2}[-\frac{1}{\alpha_{\mathcal{K}}^{(D+1)}}]\Big)\langle M^{tr}[\boldsymbol{\rho}]-\tilde f\rangle_{\mathcal{K}} \\
&= H_{\mathcal{K}}(\tilde f,\tilde g) + 0 + 0.
\end{aligned} \tag{24}$$

Thus $(M_{\mathcal{K}}^{tr}[\boldsymbol{\rho}_{\mathcal{K}}], M_{\mathcal{K}}^{int}[\boldsymbol{\rho}_{\mathcal{K}}])$ is the unique minimum of $H_{\mathcal{K}}$.

Remark 5.3. The definition of $(M_{\mathcal{K}}^{tr}[\boldsymbol{\rho}_{\mathcal{K}}], M_{\mathcal{K}}^{int}[\boldsymbol{\rho}_{\mathcal{K}}])$ by the minimum entropy principle is fundamental to obtain an entropic model. For instance, an alternate definition, very natural, is

$$M_{\mathcal{K}}^{tr}[\boldsymbol{\rho}_{\mathcal{K}}] = \exp(\boldsymbol{\alpha}_{\mathcal{K}}\cdot\boldsymbol{m}(v_k)), \qquad M_{\mathcal{K}}^{int}[\boldsymbol{\rho}_{\mathcal{K}}] = \frac{\delta}{2}\theta_{\mathcal{K}}M_{\mathcal{K}}^{tr}[\boldsymbol{\rho}_{\mathcal{K}}],$$

with $\boldsymbol{\alpha}_{\mathcal{K}}$ such that $\langle \boldsymbol{m}M_{\mathcal{K}}^{tr}[\boldsymbol{\rho}_{\mathcal{K}}] + \mathbf{e}_{D+1}M^{int}[\boldsymbol{\rho}]\rangle_{\mathcal{K}} = \boldsymbol{\rho}_{\mathcal{K}}$. This definition is very close to the continuous one, and also gives a conservative model. However, in the discrete case we do not have $\theta_{\mathcal{K}} = -\frac{1}{\alpha_{\mathcal{K}}^{(D+1)}}$, as opposed to the continuous case. Thus we do not have the relation $M_{\mathcal{K}}^{int}[\boldsymbol{\rho}_{\mathcal{K}}] = \frac{\delta}{2}[-\frac{1}{\alpha_{\mathcal{K}}^{(D+1)}}]M_{\mathcal{K}}^{tr}[\boldsymbol{\rho}_{\mathcal{K}}]$ which is necessary to prove the entropy property in (24).

For concluding the proof of theorem 4.1, we now prove lemma 5.2.

Proof of lemma 5.2. The proof of this lemma is more complicated than for the monoatomic case. This is due to the fact that $\mathcal{D}$ is bounded in one direction, thus the analysis of [13] cannot be completely applied.

First, note that J is clearly twice differentiable on $\mathcal{D}$. Its Hessian matrix is

$$J''(\gamma) = \langle \boldsymbol{m} \otimes \boldsymbol{m} \exp(\boldsymbol{\gamma} \cdot \boldsymbol{m}) \rangle_{\mathcal{K}} + \frac{\delta}{2(\gamma^{(D+1)})^2} \rho_{\mathcal{K}}^{(0)} \mathbf{e}_{D+1} \otimes \mathbf{e}_{D+1}.$$

This matrix is the sum of two matrices, the first one is positive definite (owing to assumption (11) on the grid $\mathcal{V}$, see a proof in [13]), the second one is positive. Consequently the sum is positive definite and thus J is strictly convex.

For the coercivity, we first prove that J is coercive in each direction of $\mathbb{R}^{D+2}$.

Lemma 5.4. *Let $\boldsymbol{\beta}$ be a vector of $\mathcal{D}$ and $R > 0$ a real number. For all $\boldsymbol{\omega}$ of the unit sphere S^{D+1} of $\mathbb{R}^{D+2}$, there exists a real $\bar{s}(R, \boldsymbol{\omega}, \boldsymbol{\beta}) > 0$ such that*

$$J(\boldsymbol{\beta} + s\boldsymbol{\omega}) \geq R, \qquad \forall s \geq \bar{s}.$$

Proof of lemma 5.4. First, we recall that assumption (11) on $\mathcal{V}$ in theorem 4.1 is equivalent to

$$\boldsymbol{\gamma} \cdot \boldsymbol{m}(v_k) = 0 \quad \forall k \in \mathcal{K} \iff \boldsymbol{\gamma} = 0. \tag{25}$$

Now, note that

$$\begin{aligned} J(\boldsymbol{\beta} + s\boldsymbol{\omega}) = &\sum_{k \in \mathcal{K}} \exp(\boldsymbol{\beta} \cdot \boldsymbol{m}(v_k)) \exp(s\boldsymbol{\omega} \cdot \boldsymbol{m}(v_k)) \Delta v^D - s\omega \cdot \boldsymbol{\rho}_{\mathcal{K}} \\ &- \frac{\delta}{2} \log(-\frac{2}{\delta}(\beta^{(D+1)} + s\omega^{(D+1)})) \rho_{\mathcal{K}}^{(0)} - \boldsymbol{\beta} \cdot \boldsymbol{\rho}_{\mathcal{K}}. \end{aligned}$$

As s increases, the vector $\boldsymbol{\beta} + s\boldsymbol{\omega}$ grows to infinity or intersects the boundary $\{\gamma^{(D+1)} = 0\}$ of $\mathcal{D}$, depending on the sign of $\omega^{(D+1)}$. Thus two cases are to be considered (see figure 1) :

$$(a) \quad \omega^{(D+1)} \leq 0 \quad \text{and} \quad (b) \quad \omega^{(D+1)} > 0.$$

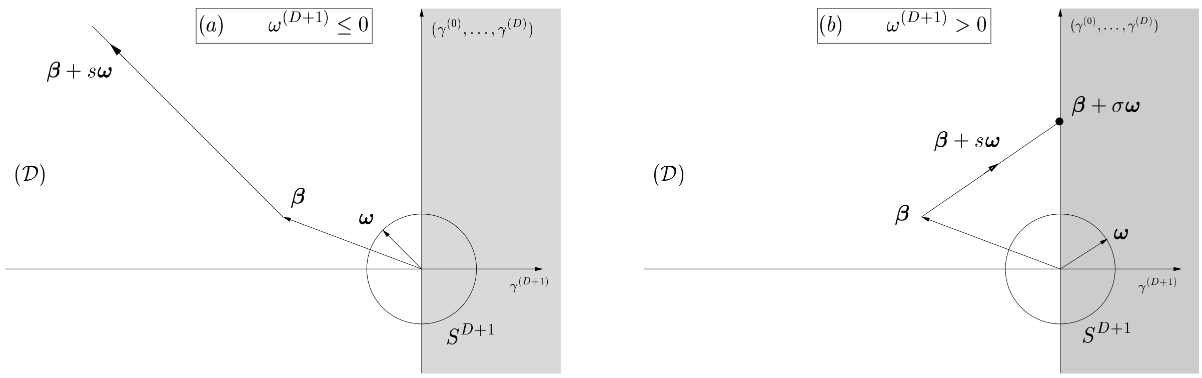

FIGURE 1. Behavior of $\boldsymbol{\beta} + s\boldsymbol{\omega}$ depending on the sign of $\omega^{(D+1)}$.

For case (a), we see that $J(\boldsymbol{\beta} + s\boldsymbol{\omega})$ is defined for any $s > 0$. First we assume that there exists k_0 such that $\boldsymbol{\omega} \cdot \boldsymbol{m}(v_{k_0}) > 0$. Therefore, denoting by $c_1(\boldsymbol{\beta})$ and $c_2(\boldsymbol{\beta})$ the terms that do not depend on s, we obtain

$$\begin{aligned} J(\boldsymbol{\beta} + s\boldsymbol{\omega}) \geq &c_1(\boldsymbol{\beta}) \exp(s\boldsymbol{\omega} \cdot \boldsymbol{m}(v_{k_0})) - s\boldsymbol{\omega} \cdot \boldsymbol{\rho}_{\mathcal{K}} \\ &\frac{\delta}{2} \log(-\frac{2}{\delta}(\beta^{(D+1)} + s\omega^{(D+1)})) \rho_{\mathcal{K}}^{(0)} - c_2(\boldsymbol{\beta}), \end{aligned}$$

which tends to $+\infty$. Thus we can find $\bar{s}(R,\boldsymbol{\omega},\boldsymbol{\beta})$ as it is stated in the lemma. If such a k_0 does not exist, then $\boldsymbol{\omega}\cdot\boldsymbol{m}(v_k)\leq 0$ for all k, and property (25) implies that there exists k_0 such that $\boldsymbol{\omega}\cdot\boldsymbol{m}(v_{k_0})<0$. Therefore from property (13) we have $\boldsymbol{\omega}\cdot\boldsymbol{\rho}_{\mathcal{K}}<0$, and thus

$$J(\boldsymbol{\beta}+s\boldsymbol{\omega})\geq -s\boldsymbol{\omega}\cdot\boldsymbol{\rho}_{\mathcal{K}}-\frac{\delta}{2}\log(-\frac{2}{\delta}(\beta^{(D+1)}+s\omega^{(D+1)}))\rho_{\mathcal{K}}^{(0)}-c_2(\boldsymbol{\beta}),$$

which tends to $+\infty$. Thus we can also find $\bar{s}(R,\boldsymbol{\omega},\boldsymbol{\beta})$ as stated in the lemma.

For case (b), we see that $J(\boldsymbol{\beta}+s\boldsymbol{\omega})$ is defined for $s<\sigma=-\frac{\beta^{(D+1)}}{\omega^{(D+1)}}$. Therefore the exponentials and the linear term in $J(\boldsymbol{\beta}+s\boldsymbol{\omega})$ are bounded. We obtain

$$J(\boldsymbol{\beta}+s\boldsymbol{\omega})\geq c(\boldsymbol{\beta})-\frac{\delta}{2}\log(-\frac{2}{\delta}(\beta^{(D+1)}+s\omega^{(D+1)}))\rho_{\mathcal{K}}^{(0)}$$

for all $s<\sigma$. Then it is clear that $J(\boldsymbol{\beta}+s\boldsymbol{\omega})$ tends to $+\infty$ as s goes to σ. Thus we can find $\bar{s}(R,\boldsymbol{\omega},\boldsymbol{\beta})$ as required, which completes the proof of lemma 5.4. □

Now, we use the following lemma, proved in the appendix, to obtain the coercivity of J.

Lemma 5.5. *Any C^1 function J defined on $\mathcal{D}=\{\boldsymbol{\gamma}\in\mathbb{R}^{D+2},\quad \gamma^{(D+1)}<0\}$ which is convex, and coercive in each direction (in the sense of lemma 5.4) is necessarily coercive in $\mathcal{D}$.*

This is in fact a general result which does not depend on the expression of J. The proof of lemma 5.2 is then complete. □

Consequently we have proved that under assumption (13) of a strict realizable vector $\boldsymbol{\rho}_{\mathcal{K}}$, the problem $(\mathcal{P}_{\mathcal{K}})$ has a unique solution (this is the part (ii) $\Rightarrow$ (i) of Theorem 4.1). The part (i) $\Rightarrow$ (ii) is obvious: by definition $(M_{\mathcal{K}}^{tr}[\boldsymbol{\rho}_{\mathcal{K}}],M_{\mathcal{K}}^{int}[\boldsymbol{\rho}_{\mathcal{K}}])$ realizes $\boldsymbol{\rho}_{\mathcal{K}}$ and from (i), we have $(M_{\mathcal{K}}^{tr}[\boldsymbol{\rho}_{\mathcal{K}}],M_{\mathcal{K}}^{int}[\boldsymbol{\rho}_{\mathcal{K}}])>0$, which proves that $\boldsymbol{\rho}_{\mathcal{K}}$ is strictly realizable. The proof of theorem 4.1 is now complete.

6. Conclusion

A new discrete-velocity model for rarefied polyatomic gases has been proposed. This model is based on a reduced BGK system of two equations: one equation for the mass distribution function and one equation for the internal energy distribution function. In term of computational storage, this model is economic, since there is no more variables than in the monoatomic model.

The velocity discretization of this system is based on a rigorous approximation of the Maxwellian equilibrium, by using a discrete version of the minimum entropy principle. The existence of this approximation (called discrete equilibrium) has been proved.

Owing to this definition, our discrete-velocity model is conservative and entropic, as the continuous one. This is essential to a robust and economic numerical scheme. By the way, an extension of the numerical method proposed in [13, 14] to this model will be presented in a forthcoming article.

References

[1] Y. Bahy. PhD thesis, ONERA, 1997

[2] P. Andriès, J.-F. Bourgat, P. Le Tallec, and B. Perthame. Numerical comparison between the Boltzmann and ES-BGK models for rarefied gases. Technical Report 3872, INRIA, 2000.

[3] P. Andriès, P. Le Tallec, J.-F. Perlat, and B. Perthame. The Gaussian-BGK model of Boltzmann equation with small Prandtl number. *Eur. J. Mech. B/Fluids*, 2000. to appear.

[4] G.A. Bird. *Molecular Gas Dynamics and the Direct Simulation of Gas Flows.* Oxford Science Publications, 1994.

[5] C. Borgnakke and P.S. Larsen. Statistical collision model for Monte-Carlo simulation of polyatomic gas mixture. *J. Comput. Phys.*, 18:405–420, 1975.

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

[7] C. Buet. Conservative and entropy schemes for Boltzmann collision operator of polyatomic gases. *M3AS*, 7(2):165–192, 1997.
[8] S. Chapman and T.G. Cowling. *The Mathematical Theory of Non-Uniform Gases.* Cambridge University Press, 1970.
[9] I. Kuscer. Models of energy exchange in polyatomic gases. *Operator Theory: Advances and Applications*, 51, 1992.
[10] E. V. Kustova and F. Mallinger. Level kinetic approach in the case of strong vibrational nonequilibrium for a pure diatomic gas. Technical Report 3577, INRIA, 1998.
[11] J.-C. Lengrand. Mise en oeuvre de la méthode de Monte Carlo pour la simulation numérique d'un écoulement de gaz raréfié. Technical report, CNRS, 1986. RC 86-4.
[12] C.D. Levermore. Moment Closure Hierarchies for Kinetic Theories. *J. Stat. Phys.*, 83:1021–1065, 1996.
[13] L. Mieussens. Discrete Velocity Model and Implicit Scheme for the BGK Equation of Rarefied Gas Dynamics. *Math. Models and Meth. in Applied Sciences*, 2000. in press.
[14] L. Mieussens. Discrete-velocity models and numerical schemes for the Boltzmann-BGK equation in plane and axisymmetric geometries. *J. Comput. Phys.*, 162:429–466, 2000.
[15] B. Perthame. Boltzmann type schemes for gas dynamics and the entropy property. *SIAM J. Numer. Anal.*, 27(6):1405–1421, 1990.
[16] B. Perthame. An introduction to kinetic schemes for gas dynamics. In *An introduction to recent developments in theory and numerics for conservation laws (Freiburg/Littenweiler, 1997)*, volume 5 of *Lect. Notes Comput. Sci. Eng.*, pages 1–27. Springer, Berlin, 1999.
[17] F. Rogier and J. Schneider. A Direct Method For Solving the Boltzmann Equation. *Transp. Th. Stat. Phys.*, 23(1-3):313–338, 1994.
[18] J.Y. Yang and J.C. Huang. Rarefied Flow Computations Using Nonlinear Model Boltzmann Equations. *Journal of Computational Physics*, 120:323–339, 1995.

7. Appendix: proof of lemma 5.5

This proof is divided into four steps.

step 1. Parametrization of the set $\{\gamma \in \mathcal{D};\ J(\gamma) = R\}$.

Let $R > 0$ be a real and $\beta \in \mathcal{D}$ such that $J(\beta) < R$. Let ω be a vector of the unit sphere S^{D+1}. From the convexity of $\mathcal{D}$, we know that the vector $\beta + s\omega$ is in $\mathcal{D}$ for any $0 \le s \le \bar{s}(\omega)$ (where we denote by $\bar{s}(\omega)$ the real $\bar{s}(R,\beta,\omega)$ of lemma 5.4). Therefore, the function $s \mapsto \chi(s) = J(\beta + s\omega)$ is defined for any s such that $0 \le s \le \bar{s}(\omega)$. From the properties of J, this function is convex and continuous in $[0, \bar{s}(\omega)]$ and it satisfies $\chi(0) < R$ and $\chi(\bar{s}(\omega)) \ge R$. Consequently there exists a unique real in $]0, \bar{s}(\omega)]$ denoted by $s^\star(\omega) > 0$ such that $\chi(s^\star(\omega)) = J(\beta + s^\star(\omega)\omega) = R$.

step 2. We prove that the mapping $\omega \in S^{D+1} \mapsto s^\star(\omega)$ is bounded. This is a direct consequence of the following lemma:

Lemma 7.1. *the mapping $\omega \mapsto s^\star(\omega)$ is continuous from S^{D+1} onto $\mathbb{R}$.*

Since S^{D+1} is compact, then $s^\star(\omega)$ is bounded, *i.e.* there exists $\sigma(\beta, R) > 0$ such that $s^\star(\omega) \le \sigma$, for all $\omega \in S^{D+1}$.

Now we turn to the proof of lemma 7.1.

Proof of lemma 7.1. Let ψ the function defined by $\psi(s,\omega) = J(\beta + s\omega) - R$, which is defined on the manifold of $\mathbb{R} \times \mathbb{R}^{D+2}$

$$\begin{aligned}\Sigma_\psi =& \{(s,\omega) \in \mathbb{R} \times S^{D+1} \text{ s.t. } s > 0 \text{ and } \omega^{(D+1)} \le 0\} \\ & \cup \{(s,\omega) \in \mathbb{R} \times S^{D+1} \text{ s.t. } 0 < s < -\frac{\beta^{(D+1)}}{\omega^{(D+1)}} \text{ and } \omega^{(D+1)} > 0\}.\end{aligned}$$

Let $\omega \in S^{D+1}$, then by definition, the real $s^\star(\omega)$ satisfies $\psi(s^\star(\omega), \omega) = 0$. The idea is then to use the implicit function theorem for ψ in Σ_ψ. Due to the fact that J is a C^1 function in $\mathcal{D}$, then ψ is differentiable at $(s^\star(\omega), \omega)$. Moreover, we have

$$\partial_s \psi(s^\star(\omega), \omega) = J'(\beta + s^\star(\omega)\omega) \cdot \omega = \frac{1}{s^\star(\omega)} J'(\gamma) \cdot (\gamma - \beta),$$

where we have set $\boldsymbol{\gamma} = \boldsymbol{\beta} + s^\star(\boldsymbol{\omega})\boldsymbol{\omega}$. Therefore, the convexity of J in $\mathcal{D}$ implies

$$\partial_s \psi(s^\star(\boldsymbol{\omega}), \boldsymbol{\omega}) \geq \frac{1}{s^\star(\boldsymbol{\omega})}(J(\boldsymbol{\gamma}) - J(\boldsymbol{\beta})).$$

Now we note that by assumption $J(\boldsymbol{\gamma}) = J(\boldsymbol{\beta} + s^\star(\boldsymbol{\omega})\boldsymbol{\omega}) = R$ and $J(\boldsymbol{\beta}) < R$, which gives $\partial_s \psi(s^\star(\boldsymbol{\omega}), \boldsymbol{\omega}) > 0$. This holds for any $\boldsymbol{\omega}$, consequently the implicit function theorem can be applied to ψ and gives that the mapping $\boldsymbol{\omega} \mapsto s^\star(\boldsymbol{\omega})$ is continuously differentiable in S^{D+1}. Then *a fortiori* , it is continuous in S^{D+1}, and this completes the proof of lemma 7.1. $\square$

step 3. We prove that the R-level set of J is compact in $\mathcal{D}$. Set $\Omega = \{\boldsymbol{\gamma} \in \mathcal{D}; J(\boldsymbol{\gamma}) \leq R\}$ and $\boldsymbol{\gamma} \neq \boldsymbol{\beta} \in \Omega$. This vector can be written as $\boldsymbol{\gamma} = \boldsymbol{\beta} + s\boldsymbol{\omega}$, with $s = |\gamma - \beta|$ and $\boldsymbol{\omega} = \frac{1}{s}(\gamma - \beta) \in S^{D+1}$. As in step 1, the properties of $\chi(s) = J(\boldsymbol{\beta} + s\boldsymbol{\omega})$ imply that $s \leq s^\star(\boldsymbol{\omega})$. Then step 2 implies that $s \leq \sigma(\boldsymbol{\beta}, R)$, for any $\boldsymbol{\gamma} \in \Omega$. Therefore $|\gamma| \leq \boldsymbol{\beta} + \sigma(\boldsymbol{\beta}, R)$, for any $\boldsymbol{\gamma} \in \Omega$, and thus Ω is bounded. Moreover, $\Omega = J^{-1}(]-\infty, R])$ implies that this set is closed in $\mathcal{D}$. Consequently, Ω is a compact subset of $\mathcal{D}$.

step 4. Conclusion.

This implies that there exist two reals $B(R) > 0$ and $\varepsilon(R) > 0$ such that

$$\boldsymbol{\gamma} \in \Omega \Rightarrow \big(|\gamma| \leq B(R) \text{ and } \varepsilon(R) \leq -\gamma^{(D+1)}\big).$$

Therefore, if $\boldsymbol{\gamma} \in \mathcal{D}$ satisfies $|\gamma| > B(R)$ or $0 < -\gamma^{(D+1)} < \varepsilon(R)$, then $\boldsymbol{\gamma} \notin \Omega$, and hence $J(\boldsymbol{\gamma}) > R$.

Since $B(R)$ and $\varepsilon(R)$ can be obtained for any $R > 0$, this proves that J is coercive in $\mathcal{D}$. The proof of lemma 5.5 is now complete.