Convergence rate for the method of moments with linear closure relations

Y. Bourgault[1], D. Broizat[2], P.-E. Jabin[3]

Abstract

We study linear closure relations for the moments' method applied to simple kinetic equations. The equations are linear collisional models (velocity jump processes) which are well suited to this type of approximation. In this simplified, 1 dimensional setting, we are able to prove stability estimates for the method (with a kinetic interpretation by a BGK model). Moreover we are also able to obtain convergence rates which automatically increase with the smoothness of the initial data.

1 Introduction

1.1 Quick presentation of the moments' methods for kinetic equations

We study here an unusual but very simple choice of closure for the moments' method where we can completely characterize the stability and convergence rates of the approximation. As far as we know this very simple situation was never considered before. Before presenting this choice though, let us very briefly give the main ideas behind the method of moments.

Moments' methods have been introduced in [10] in the context of the Boltzmann equation. This well known equation is posed on the density

[1]Department of Mathematics and Statistics, University of Ottawa, 585 King Edward Ave, Ottawa, Ontario, Canada. E-mail: ybourg@uottawa.ca

[2]Laboratoire J.-A. Dieudonné, Université de Nice – Sophia Antipolis, Parc Valrose, 06108 Nice Cedex 02, France, E-mail: broizat@unice.fr

[3]CSCAMM and Department of Mathematics, University of Maryland, College Park, MD 20742-4015, USA, E-mail: pjabin@cscamm.umd.edu

$f(t,x,v)$ of particles in the phase space and reads

$$\partial_t f + v\cdot\nabla_x f = Q(f), \tag{1.1}$$

where Q is a non linear operator (in the velocity variable v) which expresses how the velocity of a particle may change when it has a random collision.

Solving numerically an equation like (1.1) is in general very costly. The structure of the right hand side Q is non local in v and moreover the equation is posed in phase space which means that one has to work in dimension $2d+1$ if $x, v \in \mathbb{R}^d$ (7 for instance if $x \in \mathbb{R}^3$).

The moments' method is one possible answer to this problem and it consists in solving equations on the polynomial moments of f. Let $m(v)$ be polynomial in v then

$$\partial_t < m, f > +\nabla_x\cdot < v\, m(v), f >=< m, Q(f) >, \tag{1.2}$$

where we denote

$$< m, f >= \int m(v)\, f(t,x,v)\, dv. \tag{1.3}$$

Now instead of solving one equation in dimension $2d+1$, one has to solve several equations but in dimension $d+1$. Moreover in general one can expect that $< m,\, Q(f) >$ is not too complicated to compute.

However the system given by (1.2) is not closed as vm is always one degree higher than m. Therefore no matter how many moments m_i one chooses, it is never possible to express all the $v\, m_j$ in terms of the m_i. This is the closure problem and it means that (1.1) is never equivalent to (1.2) for any finite number of moments.

Instead one typically chooses a closure equation, *i.e.* a relation between $< vm_i,\, f >$ and the $< m_j, f >$ for those i where vm_i cannot be expressed in terms of the m_j.

The first big difficulty for this type of method is how to choose the closure in order to ensure that the corresponding moments' system has good properties and gives a good approximation of Eq. (1.1). This problem was of course recognized early on, see for instance [3], as well as the role of entropy in that respect, see [15] among many others.

One of the first systematic ways of finding a closure was introduced in [12] and [13]. It is still not easy to actually compute the relation which means that it is often computed numerically instead (see [18], [6], [7]). Different closures can of course be used (see [21] for example).

Theoretically even checking that the corresponding method leads to a hyperbolic system is not easy (we refer for instance to [4], [5]). Proving convergence rates seems to be out of reach for the time being although in practice it seems to be working well (see [11] for a numerical study).

Let us also mention that the methods of moments has also been used for theoretical purposes (as in [9]) and not only numerical computations.

We conclude this very brief overview by referring to [2], [17] or [19] for more on numerical simulations for kinetic equations in this context.

1.2 Linear closure relations

The guiding question in this article is whether it can make sense to consider a linear closure relation. This is certainly delicate in the nonlinear case of Boltzmann eq. (1.1). Instead we choose a simplified 1d setting where it is possible to fully analyze the method.

Instead of (1.1), we consider the linear model

$$\begin{cases} \partial_t f + v\partial_x f = L(f), \qquad (x,v) \in \mathbb{R}\times\mathbb{R} \\ \\ L(f) = \displaystyle\int_{\mathbb{R}} Q(v,v^*) f(t,x,v^*)dv^* - \lambda f \end{cases}, \tag{1.4}$$

with $\lambda > 0$ and where the operator Q corresponds to a velocity jump process. It models a dynamic where the particles' velocities are piecewise constant but may jump instantaneously, following a Poisson like process, according for instance to collisions with a random medium. The kernel Q captures the probability of such collisions starting with velocity v^* and ending with velocity v.

While much simplified with respect to (1.1), it is not uninteresting in itself, with applications to physics (see [8], [23]) or biology (see [14] for example). The equation has to be supplemented with some initial data, which for simplicity we assume to be compactly supported in velocity

$$f(t=0,x,v) = f^0(x,v) \in L^2(\mathbb{R}^2), \quad \text{supp}\, f^0 \subset \mathbb{R}\times I. \tag{1.5}$$

In general Q could even be assumed to depend on t and x. Here we make the additional approximation

$$Q(v,v^*) = \left(q(v) \sum_{j=0}^{d} \alpha_j v^{*j} \right) \mathbb{1}_{\{(v,v^*)\in I^2\}}, \tag{1.6}$$

with q smooth and compactly supported in some interval I, $d \in \mathbb{N}^*$ and $(\alpha_j)_{0\le j\le d} \in \mathbb{R}^{d+1}$. With this special form, one of course expects to be in a very favorable situation for the method of moments. Hence this should be seen as a simple toy model where the method can easily be tested.

Denoting the moments of the solution f by

$$\mu_i^f(t,x) := \int_I v^i f(t,x,v)dv, \qquad i \in \mathbb{N}, \tag{1.7}$$

Eq. (1.4) simply becomes

$$\partial_t f + v\partial_x f = L(f) = q(v)\sum_{j=0}^{d} \alpha_j \mu_j^f - \lambda f. \tag{1.8}$$

As we work in dimension 1, the structure of the hierarchy of equations on the moments is also very simple

$$\partial_t \mu_i^f + \partial_x \mu_{i+1}^f = \gamma_i \left(\sum_{j=0}^{d} \alpha_j \mu_j^f\right) - \lambda \mu_i^f, \qquad\qquad i \in \mathbb{N}, \tag{1.9}$$

where we truncate at order N and we define the moments of q

$$\gamma_i = \mu_i^q = \int_I v^i q(v)dv, \qquad i \in \mathbb{N}. \tag{1.10}$$

In order to close the system, it would be necessary to be able to express μ_{N+1}^f in terms of the μ_i^f for $i \le N$. The linear closure relation that we study here consists in assuming that μ_{N+1}^f is a linear combination of the lower moments.

That means that instead of (1.8) or (1.9), we solve

$$\begin{cases} \partial_t \mu_i + \partial_x \mu_{i+1} = \gamma_i \left(\displaystyle\sum_{j=0}^{d} \alpha_j \mu_j\right) - \lambda \mu_i, \qquad i = 0,\dots,N \\ \mu_{N+1} = \displaystyle\sum_{i=0}^{N} a_i \mu_i. \end{cases}, \tag{1.11}$$

with $(a_0,\dots,a_N)$ given real coefficients that one should choose in the "best" possible way.

One can rewrite (1.11) in matrix form as

$$\partial_t M_N + \partial_x (A M_N) = B M_N, \tag{1.12}$$

where $M_N = M_N(t,x) = (\mu_0(t,x), \dots, \mu_N(t,x))^T \in \mathbb{R}^{N+1}$, and $A, B \in \mathcal{M}_{N+1}(\mathbb{R})$ are defined by

$$A = \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ a_0 & \dots & a_{N-1} & a_N \end{pmatrix}, \quad B = \left(\begin{array}{c|c} \vdots & 0 \\ (\gamma_i \alpha_j)_{\substack{0 \le i \le N \\ 0 \le j \le d}} & \vdots \\ \vdots & 0 \end{array} \right) - \lambda I_{N+1}. \tag{1.13}$$

1.3 Basic properties of the method and the main result

Given the simple form of (1.11) or (1.12), some properties are obvious. Given its linear nature, the system is hyperbolic if the characteristic polynomial

$$\chi_A(X) = det(X I_{N+1} - A) = X^{N+1} - \sum_{i=0}^{N} a_i X^i \tag{1.14}$$

has $N+1$ real roots, denoted by

$$spec(A) = \{\lambda_0, \dots, \lambda_N\}. \tag{1.15}$$

This is enough to guarantee the well posedness of the numerical system (1.12) but not necessarily good stability properties. Norms of the numerical approximation could for instance grows fast as N increases. If the initial data is compactly supported in x then the solution is as well and the support propagates with speed $\max_k |\lambda_k|$.

On the other hand, the major inconvenients of such a method are also pretty clear. For instance positivity of the even moments will likely not be preserved. Even the positivity of the macroscopic density μ_0^f has no reason to be propagated.

However a careful analysis can in fact reveal that it is possible to choose appropriately the coefficients a_i in order to have not only stability but also very fast convergence.

For simplicity, assume that $I=[-1,\ 1]$ (just rescale and translate otherwise) and choose the Tchebychev polynomial for χ_A or

$$\forall 0\leq k\leq N,\qquad \lambda_k=\cos\left(\left(\frac{2k+1}{2N+2}\right)\pi\right), \tag{1.16}$$

or

$$\chi_A^{(N+1)}(X)=\prod_{k=0}^{N}\left(X-\cos\left(\left(\frac{2k+1}{2N+2}\right)\pi\right)\right). \tag{1.17}$$

Then it is possible to show:

Theorem 1.1 *Assume that f^0 satisfies* (1.5)*, that Q satisfies* (1.6)*. Then the solution to the truncated moments' hierarchy* (1.11) *where the coefficients a_i are chosen according to* (1.16) *or* (1.17) *satisfies*

$$\sup_{0\leq i\leq N}\ \sup_{t\in[0,T]}\|\mu_i(t)\|_{L^2(\mathbb{R})}\leq e^{TC_{d,\alpha}\|q\|_{L^2(I)}}\|f^0\|_{L^2(\mathbb{R}^2)}, \tag{1.18}$$

where $C_{d,\alpha}=\sqrt{\pi(d+1)}\left(\sum_{j=0}^{d}{\alpha_j}^2\right)^{1/2}$ is independent of N.

In addition if $f^0\in H^k(\mathbb{R}^2)$, $d=0$, denoting by f the corresponding solution to (1.4) *and defining its moments by* (1.7)*, one has*

$$\|\mu_0-\mu_0^f\|_{L^\infty([0,T],L^2(\mathbb{R}))}\leq\frac{C}{N^{k-3/4}}\times\|f^0\|_{H^k(\mathbb{R}^2)}. \tag{1.19}$$

where $C\geq 0$ depends on T,λ,q and k.

Remark 1.2 *The convergence result is given for the first moment and $d=0$. Higher moments would lose small powers of N so the same proof would give*

$$\|\mu_i-\mu_i^f\|_{L^\infty([0,T],L^2(\mathbb{R}))}\leq\frac{C\,N^{k\,(i+d)/N}}{N^{k-3/4}}\times\|f^0\|_{H^k(\mathbb{R}^2)}.$$

We do not know whether those results are optimal or not and in particular whether the numerical value $3/4$ in (1.19) *is optimal (though actual numerical simulations may suggest it is).*

How to choose the λ_k or the polynomial $\chi_A^{(N+1)}$ is of course the key point in this method. There might be other possibilities than the Tchebychev polynomial. As it will appear later in the analysis, a good choice needs to satisfy several constraints. First of all it has to be connected with an exact numerical integration formula, that is there should exist a nice function ρ_N s.t.

$$\int_I \frac{R(v)}{\rho_N(v)}\,dv = \sum_k R(\lambda_k),$$

for any polynomial R of degree at most $2N+1$. Moreover one wants to bound uniformly in N for every i the constants

$$\sum_k \lambda_k^{2i}.$$

This naturally leads to the Tchebychev polynomial.

In this setting, the conclusion of this analysis is that the linear method of moments should be seen in the same light as spectral methods (see [16] for instance). It is stable, converges automatically at order $k-3/4$ if the initial data is in H^k but does not propagate any additional property (positivity being probably the most important).

The next section is a more detailed presentation of the stability and convergence analysis in the general case (without necessarily choosing the Tchebychev points). The corresponding technical justifications and calculations are presented in two separate sections. Section 3 is devoted to the proof of stability results, while section 4 deals with the error estimate.

Theorem 1.1 is proved in section 5. The last section is an appendix and recalls some well known results.

2 Stability and convergence results

We present here in more details the kind of stability and convergence results that can be proved for the method (1.11) without assuming any particular choice of the eigenvalues (like (1.16)). We give here the main ideas in the approach and leave the technical proofs to further sections.

2.1 Eigenvectors for System (1.12)

As they enter in some of the estimates, we start by a short parenthesis about the eigenvectors for the matrix A. Define $P\in\mathcal{M}_{N+1}(\mathbb{R})$ the matrix of the eigenvectors; then

$$P_{i,j} = \lambda_j^i, \qquad 0\le i,j\le N. \tag{2.1}$$

Its inverse P^{-1} can be computed as easily and

$$P^{-1}_{i,j} = \frac{\tilde{p}_{i,j}}{\pi_i}, \qquad 0 \le i, j \le N, \tag{2.2}$$

where

$$\pi_i = \prod_{j \neq i} (\lambda_i - \lambda_j), \qquad \tilde{p}_{i,j} = (-1)^{N-j} \sum_{\substack{k_1 < \ldots < k_{N-j} \\ k_l \neq i \ \forall l}} \prod_{l=1}^{N-j} \lambda_{k_l} = \sum_{l=0}^{j} \frac{a_l}{\lambda_i^{j-l+1}}, \tag{2.3}$$

with the convention that $\tilde{p}_{i,N} = 1$.

Moreover, an easy computation shows that

$$\tilde{p}_{i,j} = \lambda_i^{N-j} - a_N \lambda_i^{N-j-1} - \cdots - a_{j+2}\lambda_i - a_{j+1}, \qquad 0 \le i, j \le N. \tag{2.4}$$

We can notice that $\tilde{p}_{i,j}$ is an homogeneous polynomial of degree $N - j$ in the eigenvalues $(\lambda_0, \ldots, \lambda_N)$.

2.2 Stability estimate and kinetic interpretation of the method

Let us first state our main stability estimate.

Theorem 2.1 *Assume* (1.5) *and that* $\{\lambda_0, \ldots, \lambda_N\} \subset I$. *Moreover assume that there exists a function* $\rho_N(v)$, *positive on* I, *such that*

$$\int_I \frac{R(v)}{\rho_N(v)} dv = \sum_{k=0}^{N} R(\lambda_k), \qquad \forall R \in \mathbb{R}_{2N+1}[X]. \tag{2.5}$$

Then, the hyperbolic system (1.12) *is stable and*

$$\sup_{t \in [0,T]} \|\mu_i(t)\|_{L^2(\mathbb{R})} \le \left(\sum_{k=0}^{N} \lambda_k^{2i} \right)^{1/2} \times e^{TC_N(q)} \times C_N(f^0), \qquad i = 0, \ldots, N \tag{2.6}$$

where

$$C_N(f^0) = \left(\iint_{\mathbb{R} \times I} |f^0(x,v)|^2 \rho_N(v) dx dv \right)^{1/2}, \tag{2.7}$$

$$C_N(q) = \Lambda_{N,d} \left(\int_I |q(v)|^2 \rho_N(v) dv \right)^{1/2} - \lambda, \tag{2.8}$$

$$\Lambda_{N,d} = \sqrt{\sum_{j=0}^{d} \alpha_j^2} \sqrt{\sum_{j=0}^{d} \sum_{k=0}^{N} \lambda_k^{2j}} = \sqrt{\sum_{j=0}^{d} \alpha_j^2} \sqrt{\sum_{k=0}^{N} \frac{1-\lambda_k^{2d+2}}{1-\lambda_k^2}}. \tag{2.9}$$

This result does not assume any particular distribution on the eigenvalues but of course it is by no means guaranteed in general that one could find ρ_N satisfying (2.5). Notice that the corresponding relation is really a quadrature formula for computing integrals on I which we ask to be exact for polynomials of degree $2N+1$.

We do not prove this result directly on the system (1.12). Instead we show a corresponding result on a linear BGK problem (see [1] for the simplification of collision kernels into BGK models).

For any $\varepsilon > 0$ and $N \geq d$, consider the equation

$$\begin{aligned} &\partial_t f_N^\varepsilon + v \partial_x f_N^\varepsilon = \frac{M^{(N)} f_N^\varepsilon - f_N^\varepsilon}{\varepsilon} + L(f_N^\varepsilon) \\ &f_N^\varepsilon(0,x,v) = f_0(x,v), \end{aligned} \tag{2.10}$$

where the linear operator L is defined by (1.8) and the Maxwellian $M^{(N)} : f \mapsto M^{(N)} f$ satisfies the moment conditions

$$\left\{ \begin{aligned} \int_I v^i M^{(N)} f dv &= \int_I v^i f dv \qquad i = 0, \ldots, N, \\ \int_I v^{N+1} M^{(N)} f dv &= \sum_{i=0}^{N} a_i \int_I v^i f dv \end{aligned} \right. . \tag{2.11}$$

This problem is a kinetic approximation of the macroscopic problem (1.11) as formally when $\varepsilon \to 0$ then one recovers (1.11) from (2.10).

Indeed, multiplying (2.10) by v^i and integrating over $v \in I$, we obtain

$$\partial_t \mu_i^\varepsilon + \partial_x \mu_{i+1}^\varepsilon = \gamma_i \left(\sum_{j=0}^{d} \alpha_j \mu_j^\varepsilon \right) - \lambda \mu_i^\varepsilon \quad \text{for } i = 0, \ldots, N,$$

where $\mu_i^\varepsilon := \int_I v^i f_N^\varepsilon dv$ for $i \in \mathbb{N}$. Moreover, when $\varepsilon \to 0$, we have formally

$$\mu_{N+1}^\varepsilon = \int_I v^{N+1} f_N^\varepsilon dv \sim \int_I v^{N+1} M^{(N)} f_N^\varepsilon dv = \sum_{i=0}^N a_i \int_I v^i f_N^\varepsilon dv = \sum_{i=0}^N a_i \mu_i^\varepsilon,$$

which is our closure relation.

The interest of (2.10) is to make some computations more transparent and easier to follow and in addition at the kinetic level which is more natural given the original equation. If we can prove stability estimates for (2.10) that are uniform in ε then simply by passing to the limit, we will obtain estimates for (1.12)-(1.11).

The most obvious choice for the Maxwellian is simply

$$M^{(N)} f = \sum_{i=0}^N \left(\int_I v^i f dv \right) m_i, \tag{2.12}$$

where the $v \mapsto m_i(v)$, $i = 0, \dots, N$, are any functions satisfying

$$\int_I v^j m_i(v) dv = \delta_{i,j}, \quad 0 \le j \le N, \qquad\qquad \int_I v^{N+1} m_i(v) dv = a_i. \tag{2.13}$$

The conditions (2.13) ensure that (2.11) holds.

There are obviously many ways to choose the m_i s.t. (2.13) is satisfied. What we are looking for is a choice compatible with an inner product Φ_N such that the application $M^{(N)}$ is an orthogonal projection for Φ_N. Formally this implies that for any f

$$\Phi_N(f, M^{(N)} f) \le \Phi_N(f, f).$$

In addition this inner product should have good symmetry properties s.t. for f and g

$$\Phi_N(f, vg) = \Phi_N(vf, g).$$

The simplest way to ensure this is to look for a weight ρ_N s.t.

$$\Phi_N(f, g) = \int_I f(v)\, g(v)\, \rho_N(v)\, dv.$$

In that case formally

$$\Phi_N(f, v\partial_x f) = \frac{1}{2} \partial_x \, \Phi_N(f, vf).$$

Therefore if f_N^ε solves (2.10) then one expects that

$$\frac{d}{dt}\int_{\mathbb{R}} \Phi_N(f_N^\varepsilon, f_N^\varepsilon)\,dx \le 2\int_{\mathbb{R}} \Phi_N(f_N^\varepsilon,\, L(f_N^\varepsilon))\,dx.$$

This is the strategy that we implement. Find appropriate conditions on ρ_N and the m_i to obtain the correct structure and then simply bound $\Phi_N(f, L(f))$ in terms of $\Phi_N(f,f)$.

For the first part of this strategy we actually prove

Theorem 2.2 *Let $\rho_N(v)$ a positive function on I such that*

$$\int_I \frac{R(v)}{\rho_N(v)}dv = \sum_{k=0}^{N} R(\lambda_k), \qquad \forall R \in \mathbb{R}_{2N+1}[X]. \tag{2.14}$$

We set

$$E_N = L^2(I, \rho_N(v)dv) = \left\{ f \text{ measurable}, \quad \int_I |f(v)|^2 \rho_N(v)dv < \infty \right\}.$$

Then, the map ϕ_N defined by

$$\phi_N(f,g) = \int_I f(v)g(v)\rho_N(v)dv, \qquad (f,g) \in E_N^2 \tag{2.15}$$

is an inner product on E_N, and the Maxwellian $M^{(N)} : E_N \to E_N$, defined by

$$M^{(N)}f = \sum_{i=0}^{N} \left(\int_I v^i f(v)dv \right) \frac{\tilde{T}_i(v)}{\rho_N(v)}, \qquad f \in E_N,$$

is an orthogonal projection and satisfies (2.11), where $(\tilde{T}_i)_{0\le i\le N}$ is the basis of $\mathbb{R}_N[X]$ defined by:

$$\tilde{T}_i(X) = \sum_{k=0}^{N} Q_{k,i}X^k, \qquad 0 \le i \le N.$$

with $Q = (P^T)^{-1}P^{-1}$.

Furthermore, with that choice of $M^{(N)}$, any solution f_N^ε of the problem (2.10) formally satisfies:

$$\frac{d}{dt}\int_{\mathbb{R}} \phi_N(f_N^\varepsilon, f_N^\varepsilon)dx \le 2\int_{\mathbb{R}} \phi_N(f_N^\varepsilon, L(f_N^\varepsilon))dx. \tag{2.16}$$

2.3 Error estimate

We now turn to the convergence of the method of moments to the solution. For simplicity we state the results here for the case $d = 0$ in (1.6), namely

$$Q(v, v^*) = q(v)\mathbb{1}_{\{(v,v^*)\in I^2\}}.$$

We also assume that $I \subset [-1,\ 1]$, still for simplicity.

The convergence results of course require a stability estimate. However in themselves, they do not use the specific form of the closure.

Therefore here we do not assume any specific closure relation. Instead we assume that we have a well defined methods of moments, *i.e.*, some way of computing μ_i which satisfy

$$\begin{array}{ll} \partial_t\mu_i + \partial_x\mu_{i+1} = \gamma_i\mu_0 - \lambda\mu_i, \qquad i = 0, \dots, N \\ \mu_{N+1} = F(\mu_1, ..., \mu_N). \end{array}, \tag{2.17}$$

where

$$\gamma_i = \int_I v^i q(v)dv.$$

Moreover we assume that the corresponding method has good stability estimates in the sense that

$$\exists C,\ \gamma \geq 0,\ \text{for any } (\mu_i)_{i=0..N} \text{ and } (\tilde\mu_i)_{i=0..N} \text{ solutions to (1.11) for } f^0,\ \tilde f^0,$$

$$\text{then} \qquad \sum_{i=0}^N \|\mu_i - \tilde\mu_i\|_{L^\infty([0,T],\ H^k(\mathbb{R}))} \leq CN^\gamma \|f^0 - \tilde f^0\|_{L^2(I,\ H^k(\mathbb{R}))}, \tag{2.18}$$

where C, γ may depend on T, q, λ but not on N or f^0.

Note that Th. 2.1 can indeed be expected to imply such a result. The exponent γ will depend on the choice of the coefficients in our method. Of course Th. 2.1 controls only the L^2 norm of the μ_i. However as the method (1.11) is linear, the $\partial_x^k\mu_i$ are also a solution to the same system and an estimate like (2.18) can be derived. For a more detailed analysis of how to obtain (2.18) from Th. 2.1, we refer to Section 5 where it is performed when the λ_k are the Tchebychev points.

For any method that satisfies (2.18), then one has the following convergence result

Theorem 2.3 *Assume* (1.5) *with* $I \subset [-1,\ 1]$, *that the method* (2.17) *satisfies* (2.18) *for some* $\gamma \geq 0$ *and that* $f^0 \in H^k(\mathbb{R}^2)$, *with* $k \in \mathbb{N}^*$. *Consider* f *the solution to* (1.4) *with* $d = 0$ *and the corresponding solution* μ_i *to* (2.17). *Then, for all* $T \geq 0$ *and for all* $N \geq 1$, *we have the estimate*

$$\|\mu_0 - \mu_0^f\|_{L^\infty([0,T],L^2(\mathbb{R}))} \leq \frac{C}{N^{k-\gamma}} \times \|f^0\|_{H^k(\mathbb{R}^2)}. \tag{2.19}$$

where $C \geq 0$ *depends on* T, λ, q *and* k *but not on* N.

This error estimate is a sort of interpolation between the stability bounds and the following result for C^∞ solution to (2.17)

Proposition 2.4 *Assume that the* $(\mu_i)_{i=0...N+1}$ *solve*

$$\partial_t \mu_i + \partial_x \mu_{i+1} = \gamma_i\, \mu_0 - \lambda \mu_i, \qquad i = 0, ..., N, \tag{2.20}$$

with $\mu_i(t = 0) = 0$ *for* $i = 0...N$. *Then, for all* $T \geq 0$ *and for all* $N \geq 1$, *we have the estimate*

$$\|\mu_0 - \mu_0^f\|_{L^\infty([0,T],L^2(\mathbb{R}))} \leq C \times \frac{T^N}{N!} \sum_{i=0}^{N} \|\partial_x^N \mu_i\|_{L^\infty([0,T],\ L^2(\mathbb{R}))}, \tag{2.21}$$

where $C \geq 0$ *depends on* T, λ, q.

3 Proof of Theorems 2.1 and 2.2

To simplify the presentation, we omit here the subscript N in E_N, ϕ_N, ρ_N, and the superscript N in $M^{(N)}$.

3.1 Elementary space decomposition

The difficulty is to combine the fact that M has to be an orthogonal projection for Φ with the symmetry property on Φ. We take here a slightly more general approach by not assuming directly that Φ satisfies (2.15).

Consider a Maxwellian which has the form (2.12) and satisfies (2.13). First, such an application M is a projection because $M \circ M = M$, which is a straightforward consequence of

$$M(m_i) = \sum_{j=0}^{N} \left(\int_I v^j m_i dv \right) m_j = m_i, \qquad i = 0, \ldots, N.$$

Moreover, one has that

$$KerM = \left\{ f \in E, \quad \int_I v^i f(v) dv = 0, \quad i = 0, \dots, N \right\} := K,$$

$$Ker(M - I) = Im(M) = Span(m_0, m_1, \dots, m_N) := V,$$

and we have the space decomposition

$$E = V \oplus K,$$

with

$$dim(V) = codim(K) = N + 1.$$

We start by a more detailed explanation of the sufficient conditions to obtain Th. 2.2

Lemma 3.1 *Assume that the inner product $\phi : E \times E \to \mathbb{R}$ satisfies*

$$\begin{cases} i.\ K = V^{\perp,\phi}, \text{(i.e. the decomposition } V \oplus K \text{ becomes orthogonal).} \\ ii.\ \forall (f,g) \in E^2, \qquad \phi(vf, g) = \phi(vg, f). \end{cases} \tag{3.1}$$

Then, for any solution $f_N^\varepsilon = f_N^\varepsilon(t,x,v)$ to problem (2.10)*, inequality* (2.16) *formally holds:*

$$\frac{d}{dt} \int_{\mathbb{R}} \|f_N^\varepsilon(t,x,.)\|_\phi^2 dx \leq 2 \int_{\mathbb{R}} \phi(f_N^\varepsilon, L(f_N^\varepsilon)) dx.$$

<u>Proof</u> :

Take a smooth $f = f(t,x,v)$ such that $\partial_t f + v\partial_x f = \frac{1}{\varepsilon}(Mf - f) + L(f)$. We have

$$\frac{d}{dt} \int_{\mathbb{R}} \|f(t,x,.)\|_\phi^2 dx = 2 \int_{\mathbb{R}} \phi(f, \partial_t f) dx$$
$$= 2 \left(\frac{1}{\varepsilon} \int_{\mathbb{R}} \phi(f, Mf - f) dx - \int_{\mathbb{R}} \phi(f, v\partial_x f) dx + \int_{\mathbb{R}} \phi(f, L(f)) dx \right).$$

Since $Mf - f \in K$ and $Mf \in V$, we have $\phi(Mf, Mf - f) = 0$, thus

$$\phi(f, Mf - f) = \phi(f - Mf, Mf - f) = -\|f - Mf\|_\phi \leq 0.$$

We deduce

$$\frac{d}{dt}\int_{\mathbb{R}} \|f(t,x)\|_\phi^2 dx \leq -2\int_{\mathbb{R}} \phi(f, v\partial_x f)dx + 2\int_{\mathbb{R}} \phi(f, L(f))dx.$$

Thus, having $\int_{\mathbb{R}} \phi(f, v\partial_x f)dx = 0$ is sufficient to obtain (2.16).

Since $\phi(f, v\partial_x f) = \partial_x \phi(f, vf) - \phi(\partial_x f, vf)$, we can write

$$\int_{\mathbb{R}} \phi(f, v\partial_x f)dx = \frac{1}{2}\int_{\mathbb{R}} \left(\phi(f, v\partial_x f) - \phi(vf, \partial_x f)\right) dx.$$

Of course this requires that $\phi(f, vf)$ vanish at infinity in x. This is ensured if f is L^2 in x for instance as $\phi(.,.)$ is a local operator in x (it acts only on the v variable). For instance later on, we will choose

$$\phi(f,g) = \int_I f(t,x,v)\, g(t,x,v)\, \rho_N(v)\, dv.$$

Therefore the symmetry condition (3.1) on ϕ is enough to conclude. If f is not smooth enough to follow the previous steps, one simply regularizes it by convolution in x. As the equation is linear and x is only a parameter in M and L then the regularized function solves the same equation. Therefore it satisfies (2.16) and letting the regularizing parameter vanish, one recovers the same inequality for f. $\square$

Now, take a inner product ϕ on E such that $K = V^{\perp,\phi}$, *i.e.* K and V are orthogonal for this inner product. The symmetry part *ii* of condition (3.1) is obviously equivalent to

$$\forall (f,g) \in V^2, \qquad \phi(vf, g) = \phi(vg, f), \tag{3.2}$$

$$\forall (f,g) \in K \times V, \qquad \phi(vf, g) = \phi(vg, f), \tag{3.3}$$

$$\forall (f,g) \in K^2, \qquad \phi(vf, g) = \phi(vg, f). \tag{3.4}$$

We study each of those conditions in the following subsections.

3.2 Study of the condition (3.2)

Proposition 3.2 *The condition* (3.2) *is equivalent to*

$$A^T Q = QA,$$

where A is the matrix defined by (1.13) *and Q is the symmetric definite positive matrix defined by*

$$Q_{i,j} = \phi(m_i, m_j), \qquad 0 \leq i, j \leq N.$$

Proof: The condition (3.2) is equivalent to

$$\phi(vm_i, m_j) = \phi(m_i, vm_j), \qquad \forall\, 0 \leq i, j \leq N.$$

Let $(i, j) \in \{0, \ldots, N\}^2$. Since $m_j \in V$, he have

$$\phi(vm_i, m_j) = \phi(M(vm_i), m_j).$$

Moreover, (2.13) implies

$$M(vm_i) = m_{i-1} + a_i m_N, \qquad 0 \leq i \leq N, \tag{3.5}$$

with the convention $m_{-1} = 0$. Thus

$$\phi(vm_i, m_j) = \phi(m_{i-1}, m_j) + a_i \phi(m_N, m_j) = Q_{i-1,j} + a_i Q_{N,j}.$$

But

$$(A^T Q)_{i,j} = \sum_{k=0}^{N} A_{k,i} Q_{k,j} = Q_{i-1,j} + a_i Q_{N,j},$$

with the convention $Q_{-1,j} = 0$. Therefore, we have

$$\phi(vm_i, m_j) = (A^T Q)_{i,j}.$$

Thus, (3.2) amounts to the matrix $A^T Q$ be (real and) symmetric, i.e. $A^T Q = QA$, since Q is symmetric. □

This result suggests a particular way of defining the inner product on V:

Corollary 3.3 *Choose ϕ s.t.*

$$Q = (P^{-1})^T P^{-1} = (PP^T)^{-1},$$

where P is defined by (2.1)*. This choice implies*

$$\phi(m_i, m_j) = Q_{i,j} = \sum_{k=0}^{N} P^{-1}_{k,i} P^{-1}_{k,j}. \tag{3.6}$$

Proof. We have, denoting by $D = diag(\lambda_0, \dots, \lambda_N)$:

$$\begin{aligned} A^T Q =& A^T (P^{-1})^T P^{-1} = (P^{-1}A)^T P^{-1} = (DP^{-1})^T P^{-1} = (P^{-1})^T (DP^{-1}) \\ &= (P^{-1})^T P^{-1} A = QA, \end{aligned}$$

thus Q satisfies $A^T Q = QA$. Moreover, it is easy to check that Q is symmetric, definite, and positive. □

In the rest of the proof, we always choose Φ according to (3.6).

3.3 Study of the condition (3.3)

Now, we assume (3.6) to be satisfied and analyze (3.3).

Proposition 3.4 *Assume* (3.6)*, and consider the following polynomials*

$$T_i(X) = \frac{1}{Q_{0,N}} \sum_{k=0}^{N} Q_{k,i} X^k, \qquad 0 \le i \le N, \tag{3.7}$$

where the matrix $Q \in \mathcal{M}_{N+1}(\mathbb{R})$ is defined by (3.3)*.*

Then, the condition (3.3) *is equivalent to*

$$T_N(v) m_i(v) = T_i(v) m_N(v), \qquad v \in I, \qquad 0 \le i \le N-1, \tag{3.8}$$

$$\forall f \in K, \qquad \phi_{|K}\left(\frac{1}{Q_{0,N}} \times \frac{\chi_A(v) m_N(v)}{T_N(v)}, f \right) = \int_I v^{N+1} f(v) dv. \tag{3.9}$$

The proof of this proposition is split in several lemmas. First we find two equivalent conditions to (3.3).

Lemma 3.5 *The condition* (3.3) *is equivalent to*

$$\forall i \in \{1, \dots, N\}, \quad vm_i - m_{i-1} - a_i m_N = \frac{Q_{i,N}}{Q_{0,N}} (vm_0 - a_0 m_N) \tag{3.10}$$

$$\forall f \in K, \quad \phi_{|K}\left(\frac{1}{Q_{0,N}} (vm_0 - a_0 m_N), f \right) = \int_I w^{N+1} f(w) dw, \tag{3.11}$$

We then have to study conditions (3.10) and (3.11).
Proof of Lemma 3.5: The condition (3.3) is obviously equivalent to

$$\phi(vm_i, f) = \phi(m_i, vf), \qquad 0 \leq i \leq N, \quad f \in K.$$

Let $i \in \{0, \dots, N\}$ and $f \in K$. Since $f \in K$, he have

$$\phi(vm_i, f) = \phi_{|K}(vm_i - M(vm_i), f) = \phi_{|K}(vm_i - m_{i-1} - a_i m_N, f),$$

using (3.5). Similarly, since $m_i \in V$,

$$\phi(m_i, vf) = \phi_{|V}(m_i, M(vf)) = Q_{i,N} \int_I w^{N+1} f(w) dw,$$

where the last equality is deduced from

$$\forall f \in K, \qquad M(vf) = \left(\int_I w^{N+1} f(w) dw \right) m_N,$$

recalling that $\int w^n f \, dv = 0$, $\forall n \leq N$.

Thus, the condition (3.3) amounts to for any $0 \leq i \leq N$ and $f \in K$

$$\phi_{|K}(vm_i - m_{i-1} - a_i m_N, f) = Q_{i,N} \int_I w^{N+1} f(w) dw.$$

from which we deduce (3.11), and

$$\phi_{|K} \left(\frac{1}{Q_{i,N}} (vm_i - m_{i-1} - a_i m_N) - \frac{1}{Q_{0,N}} (vm_0 - a_0 m_N), f \right) = 0.$$

Therefore, for all $i \in \{1, \dots, N\}$, we have

$$\frac{1}{Q_{i,N}} (vm_i - m_{i-1} - a_i m_N) - \frac{1}{Q_{0,N}} (vm_0 - a_0 m_N) \in K \cap K^{\perp,\phi} = \{0\},$$

which shows the relation (3.10).
Conversely, the conditions (3.10)+(3.11) imply (3.3) just by following the previous steps in reverse order. □

We start with condition (3.10)

Lemma 3.6 *Consider the polynomial*

$$D(X) = \sum_{i=0}^{N} \beta_i X^i,$$

where

$$\beta_i = \frac{Q_{i,N}}{Q_{0,N}} = \frac{\phi(m_i, m_N)}{\phi(m_0, m_N)}, \qquad 0 \le i \le N.$$

The condition (3.10) *is equivalent to*

$$D(v)m_i(v) = T_i(v)m_N(v), \qquad v \in I, \qquad 0 \le i \le N,$$

where $(T_i)_{0\le i\le N}$ *are the following polynomials:*

$$\begin{aligned} &T_N(X) = D(X), \\ &T_{i-1}(X) = XT_i(X) - a_iD(X) - \beta_i\chi_A(X), \qquad 1 \le i \le N. \end{aligned} \tag{3.12}$$

Proof. Setting $\beta_i = \frac{Q_{i,N}}{Q_{0,N}}$ for $i \in \{0, \ldots, N\}$, we deduce from (3.10) the recursive formula

$$m_{i-1} = vm_i - (a_i - \beta_i a_0)m_N - \beta_i(vm_0), \qquad i = 1 \ldots, N,$$

which leads to, for any $0 \le i \le N$

$$m_i = \left(v^{N-i} - \sum_{k=0}^{N-i-1} v^k (a_{i+k+1} - \beta_{i+k+1}a_0) \right) m_N - \left(\sum_{k=1}^{N-i} \beta_{i+k}v^k \right) m_0.$$

Thus

$$D(v)m_0(v) = \left(v^N - \sum_{k=0}^{N-1} (a_{k+1} - \beta_{k+1}a_0)v^k \right) m_N(v),$$

where

$$D(v) := \beta_0 + \beta_1 v + \cdots + \beta_N v^N.$$

We deduce

$$vD(v)m_0(v) = (\chi_A(v) + a_0 D(v))\, m_N(v).$$

Thus, it comes from the recursive formula on the m_i that

$$D(v)m_{i-1}(v) = vD(v)m_i(v) - (a_iD(v) + \beta_i\chi_A(v))m_N(v), \qquad i = 1 \ldots, N,$$

which is exactly

$$D(v)m_i(v) = T_i(v)m_N(v), \qquad i = 0, \ldots, N, \tag{3.13}$$

with

$$T_N(X) = D(X),$$

$$T_{i-1}(X) = XT_i(X) - a_iD(X) - \beta_i\chi_A(X), \quad i = 1, \ldots N.$$

Conversely, if the functions $(m_i)_{0\leq i\leq N}$ satisfy these relations, then (3.10) is obvious, almost everywhere $v \in I$ (except at the roots of the polynomial D). $\square$

Remark 3.7 *We notice that*

$$XT_0(X) - a_0D(X) = \chi_A(X). \tag{3.14}$$

Furthermore, the formula (3.12) *implies:*

$$1 + deg(T_i) \leq max(deg(T_{i-1}), N + 1), \qquad 1 \leq i \leq N.$$

But $T_0 \in \mathbb{R}_N[X]$*, thus* $T_i \in \mathbb{R}_N[X]$ *for all* $i \in \{0, \ldots, N\}$*.*

We can now prove

Lemma 3.8 *We have the explicit formula* (3.7) *for the* $(T_i)_{0\leq i\leq N}$ *which is recalled here*

$$T_i(X) = \frac{1}{Q_{0,N}}\sum_{k=0}^{N} Q_{k,i}X^k, \qquad 0 \leq i \leq N.$$

Proof: First, according to the definition of D in Lemma 3.6, we have, for all $k \in \{0, \ldots, N\}$,

$$\begin{aligned} D(\lambda_k) =& \frac{1}{Q_{0,N}}\sum_{i=0}^{N} Q_{i,N}\lambda_k^i = \frac{1}{Q_{0,N}}\sum_{i=0}^{N}\sum_{j=0}^{N} P^{-1}_{j,i}P^{-1}_{j,N}P_{i,k} = \frac{1}{Q_{0,N}}P^{-1}_{k,N} \\ =& \frac{1}{Q_{0,N}\ \pi_k}. \end{aligned}$$

Then, the recursive formula (3.12) easily implies:

$$T_j(\lambda_k) = (\lambda_k^{N-j} - a_N \lambda_k^{N-j-1} - \cdots - a_{j+2}\lambda_k - a_{j+1})D(\lambda_k), \qquad 0 \leq j \leq N$$

which gives (according to (2.2)-(2.4))

$$T_j(\lambda_k) = \tilde{p}_{k,j} D(\lambda_k) = \frac{\tilde{p}_{k,j}}{Q_{0,N}\ \pi_k} = \frac{P^{-1}_{k,j}}{Q_{0,N}}, \qquad 0 \leq j, k \leq N. \tag{3.15}$$

Since each polynomial T_i have a degree equal to N, it is enough to check equality (3.7) on the set $\{\lambda_0, \ldots, \lambda_N\}$, which allows to conclude. □

Remark 3.9 *The formula* (3.7) *shows that the polynomials* $(T_i)_{0 \leq i \leq N}$ *are a basis of* $\mathbb{R}_N[X]$ *since the matrix* Q *is invertible.*

We may finally characterize (3.11)

Lemma 3.10 *We assume that* (3.10) *holds. Then, the condition* (3.11) *is equivalent to*

$$\forall f \in K, \qquad \phi_{|K}\left(\frac{1}{Q_{0,N}} \times \frac{\chi_A(v) m_N(v)}{T_N(v)}, f\right) = \int_I v^{N+1} f(v) dv.$$

Proof: It is straightforward using (3.13) and the formula (3.14). □

We now have all what is needed to prove Prop. 3.4

Proof of Prop. 3.4: By Lemma 3.5, condition (3.3) is equivalent to (3.10)-(3.11). By Lemma 3.6, condition (3.10) is equivalent to (3.8) provided that the T_i are defined by (3.12). Lemma 3.8 shows that the recursive formula (3.12) actually gives the explicit formula (3.7). Finally by Lemma 3.10, we know that condition (3.11) is equivalent to (3.9) thus concluding the proof. □

3.4 Study of the condition (3.4)

We prove

Proposition 3.11 *Assume* (3.6)+(3.8)+(3.9). *Then, setting*

$$\rho(v) := \frac{Q_{0,N} T_N(v)}{m_N(v)}, \tag{3.16}$$

and

$$\phi_{|K}(f, g) = \int_I f(v) g(v) \rho(v) dv, \qquad \forall (f, g) \in K^2,$$

the condition (3.4) *is satisfied.*

Proof: This choice is compatible with (3.11), since (3.11) can also be written as

$$\forall f \in K, \qquad \phi_{|K}\left(\frac{\chi_A}{\rho}, f\right) = \int_I v^{N+1} f(v)dv,$$

and since we have $\int_I \chi_A(v)f(v)dv = \int_I v^{N+1}f(v)dv$ since $f \in K$.

Moreover, (3.4) is obviously satisfied with this choice. □

3.5 Choice of ϕ on the subspace V

For the moment Φ is defined as a weighted L^2 type inner product on K by Prop. 3.11 and on V by Corollary 3.3.

We wish to define Φ as a weighted inner product on the whole $K \oplus V$. The following lemma shows that provided ρ satisfies the right relations then Φ as defined by Prop. 3.11 and Corollary 3.3 is automatically of the right form.

Lemma 3.12 *Assume* (3.6), (3.8), (3.9), (3.16), *and assume that the weight ρ satisfies the moment conditions:*

$$\int_I \frac{R(v)}{\rho(v)} dv = \sum_{k=0}^{N} R(\lambda_k), \qquad R \in \mathbb{R}_{2N}[X].$$

Then, we have:

$$\phi_{|V}(f,g) = \int_I f(v)g(v)\rho(v)dv. \tag{3.17}$$

Proof: It is sufficient to show that the formula (3.17) holds for $(f,g) = (m_i, m_j)$. We have, for $(i,j) \in \{0,\dots,N\}^2$,

$$\int_I m_i(v)m_j(v)\rho(v)dv = \int_I \left(\tilde{T}_i\tilde{T}_j\right)(v) \times \frac{dv}{\rho(v)},$$

setting

$$\tilde{T}_i = Q_{0,N}T_i, \qquad 0 \le i \le N.$$

Moreover, we have

$$\phi(m_i, m_j) = Q_{i,j} = \sum_{k=0}^{N} P^{-1}_{k,i}P^{-1}_{k,j} = \sum_{k=0}^{N} \left(\tilde{T}_i\tilde{T}_j\right)(\lambda_k),$$

according to (3.6) and (3.15).

Since the $\frac{(N+1)(N+2)}{2}$ polynomials $(\tilde{T}_i\tilde{T}_j)_{0\leq i\leq j\leq N}$ are in $\mathbb{R}_{2N}[v]$, we see that the assumption on ρ guarantees that (3.17) is satisfied. $\square$

3.6 Proof of the theorem (2.2): Synthesis

We summarize here all the definitions and check rigorously that they are compatible.

So, assume there exists a function $\rho = \rho(v)$, positive on I and satisfying the moment conditions:

$$\int_I \frac{R(v)}{\rho(v)}dv = \sum_{k=0}^{N} R(\lambda_k), \qquad \forall R \in \mathbb{R}_{2N+1}[X]. \tag{3.18}$$

Note that this in particular implies that $1/\rho$ is integrable on I.

The space $E = L^2(I, \rho(v)dv)$ is a Hilbert space, for the inner product

$$\phi(f,g) = \int_I f(v)g(v)\rho(v)dv, \qquad (f,g) \in E^2.$$

We define the map $M : E \to E$ by

$$Mf = \sum_{i=0}^{N} \left(\int_I v^i f(v)dv\right) \frac{\tilde{T}_i(v)}{\rho(v)}, \qquad f \in E,$$

where

$$\tilde{T}_i(X) = \sum_{k=0}^{N} Q_{k,i}X^k, \qquad 0 \leq i \leq N,$$

and $Q = (P^T)^{-1}P^{-1}$ (the matrix P is defined by (2.1)).

- First, the map M is well defined as $\forall f \in E$, $\forall i \in \{0,\ldots,N\}$

$$\left|\int_I v^i f(v)dv\right| \leq \left(\int_I \frac{|v|^{2i}}{\rho(v)}dv\right)^{1/2} \left(\int_I |f(v)|^2\rho(v)dv\right)^{1/2},$$

  and thus Mf makes sense. Moreover, for $f \in E$, we have $Mf \in E$

because

$$\int_I |Mf(v)|^2\rho(v)dv = \int_I \left| \sum_{i=0}^N \left(\int_I w^i f(w)dw \right) \frac{\tilde{T}_i(v)}{\rho(v)} \right|^2 \rho(v)dv$$

$$\leq \int_I \left(\sum_{i=0}^N \left| \int_I w^i f(w)dw \right|^2 \right) \left(\sum_{i=0}^N \left| \frac{\tilde{T}_i(v)}{\rho(v)} \right|^2 \right) \rho(v)dv$$

$$\leq \left(\sum_{i=0}^N \left| \int_I w^i f(w)dw \right|^2 \right) \times \sum_{i=0}^N \int_I \frac{|\tilde{T}_i(v)|^2}{\rho(v)} dv \quad < \ \infty.$$

- It is obvious that M is linear. We check that M is a projection: let $f \in E$, we have, for $0 \leq j \leq N$,

$$\int_I v^j Mf(v)dv \ = \ \sum_{i=0}^N \left(\int_I w^i f(w)dw \right) \left(\int_I \frac{v^j \tilde{T}_i(v)}{\rho(v)} dv \right)$$

$$= \ \sum_{i=0}^N \left(\int_I w^i f(w)dw \right) \left(\sum_{k=0}^N \lambda_k^j \tilde{T}_i(\lambda_k) \right).$$

  Moreover, we have

$$\lambda_k^j = P_{j,k}, \qquad \tilde{T}_i(\lambda_k) = P^{-1}_{k,i}, \qquad 0 \leq i,j,k \leq N,$$

  thus

$$\sum_{k=0}^N \lambda_k^j \tilde{T}_i(\lambda_k) = \delta_{i,j}, \qquad 0 \leq i,j \leq N.$$

  We deduce

$$\int_I v^j Mf(v)dv \ = \ \int_I w^j f(w)dw, \qquad 0 \leq j \leq N,$$

  which implies $M \circ M = M$.

- M is an orthogonal projection (for the inner product ϕ), because it is a self-adjoint projector:

$$\begin{aligned}\phi(Mf,g) &= \int_I Mf(v)g(v)\rho(v)dv \\ &= \sum_{i=0}^{N}\left(\int_I w^i f(w)dw\right)\left(\int_I \tilde{T}_i(v)g(v)dv\right) \\ &= \sum_{i=0}^{N}\sum_{k=0}^{N} Q_{k,i}\left(\int_I w^i f(w)dw\right)\left(\int_I v^k g(v)dv\right) \;=\; \phi(f,Mg),\end{aligned}$$

as the matrix Q is symmetric.

- The Maxwellian M satisfies moment conditions:

$$\left\{\begin{array}{lcll} \int_I v^i Mf(v)dv & = & \int_I v^i f(v)dv & i=0,\dots,N, \\ \int_I v^{N+1}Mf(v)dv & = & \sum_{i=0}^{N} a_i \int_I v^i f(v)dv & \end{array}\right. .$$

  In fact, the first conditions have been already established, and the second ones result from the following formula, using (3.18) and since $\chi_A(v)\tilde{T}_i(v)$ is a polynomial of degree $2N+1$,

$$\begin{aligned}\int_I \chi_A(v)Mf(v) &= \sum_{i=0}^{N}\left(\int_I w^i f(w)dw\right)\int_I \frac{\chi_A(v)\tilde{T}_i(v)}{\rho(v)}dv \\ &= \sum_{i=0}^{N}\left(\int_I w^i f(w)dw\right)\sum_{k=0}^{N}\chi_A(\lambda_k)\tilde{T}_i(\lambda_k) = 0.\end{aligned}$$

- By Corollary 3.3, Prop. 3.11 and Lemma 3.12, the inner product Φ satisfies the assumptions of Lemma 3.1. This concludes the proof of Theorem 2.2. $\square$

3.7 From Th. 2.2 to Th. 2.1: Uniform stability estimate on the BGK model

The first point is to control the collision term which is done by

Lemma 3.13 *Define ρ and Φ as per Th. 2.2 then*

$$\int_{\mathbb{R}}\int_I f_N^\varepsilon L(f_N^\varepsilon)\rho(v)dvdx \le \left(\Lambda_{N,d}\left(\int_I |q(v)|^2\rho(v)dv\right)^{1/2} - \lambda\right) \int_{\mathbb{R}}\int_I |f_N^\varepsilon|^2\rho(v)dvdx, \tag{3.19}$$

with

$$\Lambda_{N,d} = \sqrt{\sum_{j=0}^d \alpha_j^2}\sqrt{\sum_{j=0}^d\sum_{k=0}^N \lambda_k^{2j}} = \sqrt{\sum_{j=0}^d \alpha_j^2}\sqrt{\sum_{k=0}^N \frac{1-\lambda_k^{2d+2}}{1-\lambda_k^2}}.$$

<u>Proof</u>: We have

$$\begin{aligned}
&\int_{\mathbb{R}}\int_I f_N^\varepsilon(t,x,v)L(f_N^\varepsilon)(t,x,v)\rho(v)dvdx\\
&= \int_{\mathbb{R}}\int_I f_N^\varepsilon(t,x,v)q(v)\sum_{j=0}^d \alpha_j\left(\int_I w^j f_N^\varepsilon(t,x,w)dw\right)\rho(v)dvdx\\
&\quad - \lambda\int_{\mathbb{R}}\int_I |f_N^\varepsilon(t,x,v)|^2\rho(v)dvdx.
\end{aligned}$$

So

$$\begin{aligned}
&\int_{\mathbb{R}}\int_I f_N^\varepsilon(t,x,v)L(f_N^\varepsilon)(t,x,v)\rho(v)dvdx\\
&= \int_{\mathbb{R}} dx\left(\sum_{j=0}^d \alpha_j\int_I w^j f_N^\varepsilon(t,x,w)dw\right)\left(\int_I f_N^\varepsilon(t,x,v)q(v)\rho(v)dv\right)\\
&\quad - \lambda\int_{\mathbb{R}}\int_I |f_N^\varepsilon(t,x,v)|^2\rho(v)dvdx.
\end{aligned}$$

We can control the moments of f_N^ε in the following way:

$$\left|\int_I w^j f_N^\varepsilon(t,x,w)dw\right|^2 \le \left(\int_I \frac{w^{2j}}{\rho(w)}dw\right)\left(\int_I |f_N^\varepsilon(t,x,w)|^2\rho(w)dw\right),$$

and the assumption (2.14) implies

$$\left|\int_I w^j f_N^\varepsilon(t,x,w)dw\right|^2 \le \left(\sum_{k=0}^N \lambda_k^{2j}\right)\phi(f_N^\varepsilon, f_N^\varepsilon).$$

We deduce that

$$\left|\sum_{j=0}^{d}\alpha_j\int_I w^j f_N^\varepsilon(t,x,w)dw\right|^2 \le \left(\sum_{j=0}^{d}\alpha_j^2\right)\left(\sum_{j=0}^{d}\sum_{k=0}^{N}\lambda_k^{2j}\right)\phi(f_N^\varepsilon, f_N^\varepsilon).$$

Moreover, we have

$$\left|\int_I f_N^\varepsilon(t,x,v)q(v)\rho(v)dv\right|^2 = \phi(f_N^\varepsilon, q)^2 \le \phi_N(q,q)\,\phi(f_N^\varepsilon, f_N^\varepsilon).$$

Thus, combining the last two inequalities and using Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}\int_{\mathbb{R}}\int_I f_N^\varepsilon L(f_N^\varepsilon)\rho(v)dvdx \le &\sqrt{\sum_{j=0}^{d}\alpha_j^2}\sqrt{\sum_{j=0}^{d}\sum_{k=0}^{N}\lambda_k^{2j}}\sqrt{\phi(q,q)}\int_{\mathbb{R}}\phi(f_N^\varepsilon, f_N^\varepsilon)dx\\ &-\lambda\int_{\mathbb{R}}\phi(f_N^\varepsilon, f_N^\varepsilon)dx,\end{aligned}$$

which is the desired estimate. □

Combining Th. 2.2 and Lemma 3.13, we can obtain stability estimates for (2.10) uniform in ε:

Proposition 3.14 *Let f_N^ε a solution of* (2.10) *with* (1.5). *Then, the following estimate holds for any $t > 0$*

$$\sup_{\varepsilon>0}\int_{\mathbb{R}}\int_I |f_N^\varepsilon(t,x,v)|^2\rho(v)dvdx \le e^{tC_{N,q}}\int_{\mathbb{R}}\int_I |f^0(x,v)|^2\rho(v)dvdx, \quad (3.20)$$

where

$$C_{N,q} = 2\left(\Lambda_{N,d}\left(\int_I |q(v)|^2\rho(v)dv\right)^{1/2} - \lambda\right),$$

and we recall

$$\Lambda_{N,d} = \sqrt{\sum_{j=0}^{d}\alpha_j^2}\sqrt{\sum_{j=0}^{d}\sum_{k=0}^{N}\lambda_k^{2j}} = \sqrt{\sum_{j=0}^{d}\alpha_j^2}\sqrt{\sum_{k=0}^{N}\frac{1-\lambda_k^{2d+2}}{1-\lambda_k^2}}.$$

Proof: It is straightforward using Gronwall lemma, as according to (2.16) and (3.19), we have

$$\frac{d}{dt}\int_{\mathbb{R}}\int_I |f_N^\varepsilon(t,x,v)|^2\rho(v)dvdx \leq 2\left(\Lambda_{N,d}\left(\int_I |q(v)|^2\rho(v)dv\right)^{1/2} - \lambda\right)$$
$$\int_{\mathbb{R}}\int_I |f_N^\varepsilon(t,x,v)|^2\rho(v)dvdx.$$

□

We are now ready to conclude the proof of Th. 2.1 Now, we show that we can pass to the limit $\varepsilon \to 0$ in the BGK model (2.10) to obtain a stability estimate on the hyperbolic system (1.12).

We fix $N \geq d$. Using (3.20), we can see that the family $(f_N^\varepsilon)_{\varepsilon>0}$ is bounded in the space $L^2(]0,+\infty[_{loc}\times\mathbb{R}_x\times I_v, \rho(v)dtdxdv)$. Thus there exists a sequence $\varepsilon_k \underset{k\to\infty}{\longrightarrow} 0$ such that

$$f_N^{\varepsilon_k} \underset{k\to\infty}{\rightharpoonup} f_N \in L^2\left(]0,+\infty[_{loc}\times\mathbb{R}_x \times I_v, \rho_N(v)dtdxdv\right).$$

Therefore

$$\partial_t f_N^{\varepsilon_k} + v\partial_x f_N^{\varepsilon_k} \underset{k\to\infty}{\rightharpoonup} \partial_t f_N + v\partial_x f_N, \qquad L(f_N^{\varepsilon_k}) \underset{k\to\infty}{\rightharpoonup} L(f_N) \quad \text{in} \quad \mathcal{D}'(]0,\infty[\times\mathbb{R}\times I),$$

from which we deduce

$$M(f_N^{\varepsilon_k}) - f_N^{\varepsilon_k} = \varepsilon_k(\partial_t f_N^{\varepsilon_k} + v\partial_x f_N^{\varepsilon_k} - L(f_N^{\varepsilon_k})) \underset{k\to\infty}{\rightharpoonup} 0 \quad \text{in} \quad \mathcal{D}'(]0,\infty[\times\mathbb{R}\times I),$$

and thus

$$M(f_N^{\varepsilon_k}) \underset{k\to\infty}{\rightharpoonup} f_N \quad \text{in} \quad \mathcal{D}'(]0,\infty[\times\mathbb{R}\times I).$$

Hence, passing to the limit in (2.11), we show that the function f_N is a kinetic interpretation of the system (1.12), in the sense that $\left(\mu_i = \int_I v^i f_N dv\right)_{0\leq i\leq N}$ satisfies (1.12). As this system is linear, hyperbolic, it has a unique solution for a given initial data. That means that any solution of (1.12) can consequently be obtained as the moments of a limit f_N of f_N^ε.

Moreover, f_N also satisfies the bound (3.20) :

$$\forall t>0, \qquad \int_{\mathbb{R}}\int_I |f_N(t,x,v)|^2\rho(v)dvdx \leq e^{tC_{N,q}}\int_{\mathbb{R}}\int_I |f^0(x,v)|^2\rho(v)dvdx.$$

Thus we obtain the estimate (2.6), since

$$|\mu_i(t,x)|^2 = \left| \int_I v^i f_N(t,x,v)dv \right|^2 \le \left(\sum_{k=0}^{N} \lambda_k^{2i} \right) \int_I |f_N(t,x,v)|^2 \rho(v)dv,$$

according to the assumptions on ρ_N. The proof of Th. 2.1 is now complete. □

4 Error estimate: Proof of Th. 2.3 and Prop. 2.4

4.1 Estimates provided by the model

In order to prove Th. 2.3, we will need good smoothness properties on the solution to the exact equation (1.4). Fortunately this model is very simple to manipulate and the estimates we need easy to obtain in that case.

Let us first start with the support in velocity

Proposition 4.1 *Assume* (1.5)*,* (1.6) *with* $I \subset [-1,\ 1]$*. Then the solution to* (1.4) *satisfies*

$$\forall t \ge 0, \qquad a.e.\ x \in \mathbb{R}, \qquad \mathrm{supp_v}\, f(t,x,.) \subset I \subset [-1,\ 1].$$

Proof : If f is a solution of (1.4) with $d = 0$, then, using the Stokes formula, we have, formally

$$\begin{aligned}\frac{d}{dt}\int_{\mathbb{R}} |f(t,x,v)|^2 dx = {} & 2\int_{\mathbb{R}} f(t,x,v)q(v)\left(\int_I f(t,x,v^*)dv^*\right)dx \\ & - 2\lambda \int_{\mathbb{R}} |f(t,x,v)|^2 dx \\ \le {} & 2q(v)\int_{\mathbb{R}} f(t,x,v)\left(\int_I f(t,x,v^*)dv^*\right)dx.\end{aligned}$$

Therefore, integrating in time, we obtain

$$\begin{aligned}\int_{\mathbb{R}} |f(t,x,v)|^2 dx \le {} & \int_{\mathbb{R}} |f^0(x,v)|^2 dx \\ & + 2q(v)\int_0^t \int_{\mathbb{R}} f(s,x,v)\left(\int_I f(s,x,v^*)dv^*\right)dxds,\end{aligned}$$

and since $\mathrm{supp}\, q \subset I$, we get the result. □

The model (1.4) also propagates the H^k-norm of the solution

Proposition 4.2 *Assume* (1.5), (1.6) *with* $I \subset [-1,\ 1]$. *Then the solution to* (1.4) *satisfies*

$$\forall t \geq 0, \qquad \|f(t)\|_{H^k(\mathbb{R},\ L^2(I))} \leq e^{Ct}\|f^0\|_{H^k(\mathbb{R},\ L^2(I))},$$

where $C = \|q\|_{L^2(I)} - \lambda$.

Proof: First note that for any k, $\partial_x^k f$ is also a solution to (1.4) with $\partial_x^k f^0$ as initial data. Then

$$\begin{aligned}\frac{d}{dt}\|f(t)\|^2_{H^k(\mathbb{R},\ L^2(I))} =& 2\sum_{p=0}^{k}\int_{\mathbb{R}}\left(\int_{\mathbb{R}}\partial_x^p f(t,x,v)q(v)dv\right)\left(\int_I \partial_x^p f(t,x,v^*)dv^*\right)\\ &- 2\lambda\|f(t)\|^2_{H^k(\mathbb{R},\ L^2(I))}.\end{aligned}$$

Using Prop. (4.1) and Hölder inequality, we easily obtain

$$\frac{d}{dt}\|f(t)\|^2_{H^k(\mathbb{R},\ L^2(I))} \leq 2\left(\|q\|_{L^2(I)} - \lambda\right)\|f(t)\|^2_{H^k(\mathbb{R},\ L^2(I))},$$

and a simple Gronwall lemma gives the result. □

We may finally conclude from Props. 4.1-4.2 that

Corollary 4.3 *Assume* (1.5), (1.6) *with* $I \subset [-1,\ 1]$. *Then the moments of solution to* (1.4) *satisfy*

$$\forall i \in \mathbb{N}, \quad \forall t \geq 0, \qquad \|\mu_i^f(t)\|_{L^2(\mathbb{R})} \leq e^{Ct}\|f^0\|_{L^2(\mathbb{R}^2)}, \tag{4.1}$$

and

$$\sum_{i=0}^{N}\|\mu_i^f(t)\|_{L^2(\mathbb{R})} \leq (2N+1)^{1/2}\, e^{Ct}\,\|f^0\|_{L^2(\mathbb{R}^2)}.$$

where $C = \|q\|_{L^2(I)} - \lambda$.

Proof. Notice that

$$|\mu_i^f| \leq \int_{-1}^{1}|v|^i f(t,x,v)\,dv \leq \frac{1}{\sqrt{2i+1}}\left(\int_{-1}^{1}|f(t,x,v)|^2\,dv\right)^{1/2},$$

by Cauchy-Schwarz. Hence

$$\|\mu_i^f(t)\|_{L^2(\mathbb{R})} \leq \frac{1}{\sqrt{2i+1}}\|f(t)\|_{L^2(\mathbb{R}^2)},$$

and one concludes by using Prop. 4.2. □

4.2 Proof of Prop. 2.4: Error estimate in the smooth case

Since the functions (μ_i) are smooth in the space variable, we have by (2.17), for all $0 \leq i \leq N$,

$$\begin{aligned}\frac{d}{dt}\|\mu_i(t)\|^2_{L^2(\mathbb{R})} &= 2\int_{\mathbb{R}} \mu_i(t,x)\partial_t\mu_i(t,x)dx \\ &= -2\int_{\mathbb{R}} \mu_i(t,x)\partial_x\mu_{i+1}(t,x)dx \qquad , \\ &+ 2\gamma_i \int_{\mathbb{R}} \mu_i(t,x)\mu_0(t,x)dx - 2\lambda\int_{\mathbb{R}} |\mu_i(t,x)|^2 dx,\end{aligned}$$

or

$$\begin{aligned}\frac{d}{dt}\|\mu_i(t)\|^2_{L^2(\mathbb{R})} &\leq 2\|\mu_i(t)\|_{L^2(\mathbb{R})}\|\partial_x\mu_{i+1}(t)\|_{L^2(\mathbb{R})} \\ &+ 2|\gamma_i|\ \|\mu_i(t)\|_{L^2(\mathbb{R})}\|\mu_0(t)\|_{L^2(\mathbb{R})} - 2\lambda\|\mu_i(t)\|^2_{L^2(\mathbb{R})}.\end{aligned}$$

Thus, we obtain

$$\frac{d}{dt}\|\mu_i(t)\|_{L^2(\mathbb{R})} \quad \leq \quad \|\partial_x\mu_{i+1}(t)\|_{L^2(\mathbb{R})} + |\gamma_i|\ \|\mu_0(t)\|_{L^2(\mathbb{R})} - \lambda\|\mu_i(t)\|_{L^2(\mathbb{R})}.$$

Of course the same computation can be performed on the $\partial_x^k\mu_i$, obtaining

$$\begin{aligned}&\forall k \in \mathbb{N}, \quad \forall i \in \{0,\dots,N\}, \\ &\frac{d}{dt}\|\partial_x^k\mu_i(t)\|_{L^2(\mathbb{R})} \leq \|\partial_x^{k+1}\mu_{i+1}(t)\|_{L^2(\mathbb{R})} + |\gamma_i|\ \|\partial_x^k\mu_0(t)\|_{L^2(\mathbb{R})} \\ &\qquad\qquad - \lambda\|\partial_x^k\mu_i(t)\|_{L^2(\mathbb{R})}.\end{aligned} \tag{4.2}$$

This sequence of inequalities lets us control $\|\mu_0(t)\|_{L^2(\mathbb{R})}$ in term of the "last derivatives", namely $\left(\|\partial_x^N\mu_i\|_{L^2(\mathbb{R})}\right)_{0\leq i\leq N}$. To do that, we set

$$\forall t \geq 0, \quad \forall k \in \{0,\dots,N\}, \qquad H_k(t) := \sum_{i=0}^{k} \|\partial_x^k\mu_i\|_{L^2(\mathbb{R})}.$$

The coefficients γ_i are easily bounded by $\|q\|_{L^2}$. Hence the inequalities (4.2) imply

$$\forall t \geq 0, \quad \forall k \in \{0,\dots,N\}, \quad \frac{d}{dt}H_k(t) \leq H_{k+1}(t) + \left(2\|q\|_{L^2(I)} - \lambda\right) H_k(t). \tag{4.3}$$

We now use

Lemma 4.4 *Let $(H_k(t))_{k\geq 1}$ be a sequence of nonnegative C^1 functions such that:*

$$\forall k \in \mathbb{N}, \quad \forall t \geq 0, \qquad \begin{cases} H_k'(t) \leq CH_k(t) + H_{k+1}(t) \\ H_k(0) = 0 \end{cases},$$

where $C > 0$ is a numerical constant independent of k.
Then, we have:

$$\forall p \in \mathbb{N}^*, \quad \forall t \geq 0, \qquad H_0(t) \leq \frac{1}{(p-1)!}\int_0^t (t-s)^{p-1} e^{C(t-s)} H_p(s)ds.$$

Proof: First, the assumption may be rewritten as

$$\forall k \in \mathbb{N}, \quad \forall t \geq 0, \qquad \frac{d}{dt}\left(H_k(t)e^{-Ct}\right) \leq H_{k+1}(t)e^{-Ct},$$

thus, integrating in t, we obtain:

$$\forall k \in \mathbb{N}, \quad \forall t \geq 0, \qquad H_k(t) \leq \int_0^t e^{C(t-s)} H_{k+1}(s)ds.$$

A simple recursion allows to conclude. □

End of the proof of prop (2.4): Applying the previous lemma, we obtain

$$\forall t \geq 0, \qquad H_0(t) \leq \frac{1}{(N-1)!}\int_0^t (t-s)^{N-1} e^{(2\|q\|_{L^2(I)}-\lambda)(t-s)} H_N(s)ds,$$

and thus, for all $t \in [0,T]$,

$$\begin{aligned} \|\mu_0\|_{L^2(\mathbb{R})} &\leq \frac{1}{(N-1)!}\int_0^t (t-s)^{N-1} e^{(2\|q\|_{L^2(I)}-\lambda)(t-s)} \sum_{i=0}^N \|\partial_x^N \mu_i(s)\|_{L^2(\mathbb{R})} ds \\ &\leq \frac{e^{(2\|q\|_{L^2(I)}-\lambda)T}}{(N-1)!}\int_0^t (t-s)^{N-1} \sum_{i=0}^N \|\partial_x^N \mu_i(s)\|_{L^2(\mathbb{R})} ds \end{aligned}$$

which gives

$$\sup_{0\leq t\leq T} \|\mu_0(t,.)\|_{L^2(\mathbb{R})} \leq C\frac{T^N}{N!} \times \sum_{i=0}^N \|\partial_x^N \mu_i\|_{L_t^\infty L_x^2}, \tag{4.4}$$

where the constant C depends on T, λ and q. This concludes the proof of Prop. 2.4. □

In the case $d > 0$, the proof is very similar: The estimates in the previous subsection are identical. Here some minor changes are required, for instance (4.3) is still true but only for $k \geq d$.

4.3 Proof of Th. 2.3: Error estimate in the general case

The general idea for the proof is to regularize the initial data. Then we have to bound 3 terms. First the error between the exact solution and the truncated hierarchy for this regularized initial data. This term is controlled by Prop. 2.4. The next term is the difference between the solution for the non regularized initial data and the solution for the regularized one, both solutions to the exact equation (1.4). This is bounded using the estimates for (1.4). The final term is the difference between the solution for the non regularized initial data and the solution for the regularized one, but both solutions to the truncated hierarchy (2.17). We bound this term thanks to assumption (2.18).

- *Step* 1*: Regularization of the initial data*

  We fix $\varepsilon > 0$ and we choose $f_\varepsilon^0 \in H^k(\mathbb{R}^2)$, with $\text{supp}_\text{v}\, f_\varepsilon^0 \subset I$, and such that (see the appendix for more details)

$$a.e.\ v \in I, \quad f_\varepsilon^0(.,v) \in C^\infty(\mathbb{R}),$$

$$\|f_\varepsilon^0(v) - f^0(v)\|_{L^2(\mathbb{R})} \le C_k \varepsilon^k \|f^0(v)\|_{H^k(\mathbb{R})}, \tag{4.5}$$

$$\|f_\varepsilon^0(v)\|_{H^m(\mathbb{R})} \le C_k \varepsilon^{k-m} \|f^0(v)\|_{H^k(\mathbb{R})} \quad \forall m > k.$$

  Let $f_\varepsilon = f_\varepsilon(t,x,v)$ be the unique solution to Eq. (1.4) with $f_\varepsilon(t=0) = f_\varepsilon^0$. We define the moments as usual

$$\mu_i^{f_\varepsilon}(t,x) = \int_I v^i\, f_\varepsilon(t,x,v)\, dv, \qquad i \in \mathbb{N}.$$

  Of course, those moments still satisfy the hierarchy

$$\partial_t \mu_i^{f_\varepsilon} + \partial_x \mu_{i+1}^{f_\varepsilon} = \gamma_i \mu_0^{f_\varepsilon} - \lambda \mu_i^{f_\varepsilon} \qquad i \in \mathbb{N}.$$

  We denote by μ_i^ε the solution to the truncated hierarchy (2.17) for the initial data

$$\mu_i^\varepsilon(t=0,x) = \mu_i^{f_\varepsilon^0}(x) = \int_I v^i f_\varepsilon^0(x,v)\, dv, \qquad 0 \le i \le N. \tag{4.6}$$

- *Step 2: Error estimates in term of* ε. First note that $\mu_i^\varepsilon - \mu_i^{f_\varepsilon}$ satisfies the assumptions of Prop. 2.4. On the one hand

$$\sum_{i\le N} \|\partial_x^N(\mu_i^\varepsilon - \mu_i^{f_\varepsilon})\|_{L^2} \le \sum_{i\le N}\left(\|\partial_x^N \mu_i^\varepsilon\|_{L^2} + \|\partial_x^N \mu_i^{f_\varepsilon}\|_{L^2}\right) \le \sum_{i\le N} \|\partial_x^N \mu_i^\varepsilon\|_{L^2} + e^{Ct}\,(2N+1)^{1/2}\|\partial_x^N\, f_\varepsilon^0\|_{L^2},$$

by Corollary 4.3. On the other hand by applying (2.18) to μ_i^ε and 0, one gets

$$\sum_{i\le N} \|\partial_x^N \mu_i^\varepsilon\|_{L_t^\infty L_x^2} \le C\,N^\gamma\,\|\partial_x^N f_\varepsilon^0\|_{L^2_{x,v}}.$$

So applying Prop. 2.4, we have for $\gamma' = \max(\gamma, 1/2)$

$$\|\mu_0^\varepsilon - \mu_0^{f_\varepsilon}\|_{L^\infty([0,T],L^2(\mathbb{R}))} \le C\,\frac{N^{\gamma'}T^N}{N!} \times \|\partial_x^N f_\varepsilon^0\|_{L^2(\mathbb{R}^2)} \le C_k\,\varepsilon^{k-N}\,\frac{N^{\gamma'}T^N}{N!} \times \|f^0\|_{H_x^k\,L_v^2}, \tag{4.7}$$

by (4.5).

We can use the stability estimate (2.18) to control

$$\|\mu_0^\varepsilon - \mu_0\|_{L^\infty([0,T],L^2(\mathbb{R}))} \le CN^\gamma\|f^0 - f_\varepsilon^0\|_{L^2} \le C_k\,N^\gamma\,\varepsilon^k\|f^0\|_{H_x^k\,L_v^2}, \tag{4.8}$$

again by (4.5).

At last, we can control $\|\mu_0^f - \mu_0^{f_\varepsilon}\|_{L^\infty([0,T],L^2(\mathbb{R}))}$ according to Corollary 4.3

$$\|\mu_0^f - \mu_0^{f_\varepsilon}\|_{L^\infty([0,T],L^2(\mathbb{R}))} = \|\mu_0^{f-f_\varepsilon}\|_{L^\infty([0,T],L^2(\mathbb{R}))} \le C\|f^0 - f_\varepsilon^0\|_{L^2(\mathbb{R}^2)} \le C_k\,\varepsilon^k\,\|f^0\|_{H_x^k\,L_v^2}. \tag{4.9}$$

- *Step 3: Choice of the parameter* ε

We deduce from (4.7), (4.8) and (4.9) the complete error estimate

$$\|\mu_0 - \mu_0^f\|_{L^\infty([0,T],L^2(\mathbb{R}))} \le C\left(N^{\gamma'}\,\frac{T^N}{N!}\,\varepsilon^{k-N} + N^\gamma\,\varepsilon^k\right)\,\|f^0\|_{H_x^k\,L_v^2}, \tag{4.10}$$

where the numerical constant $C \geq 0$ depends on k, T, q, and λ.

We of course choose the "best" value of ε, which minimizes the error. When $N > k$, we get for $\gamma \geq 1/2$

$$\varepsilon = \varepsilon^* = \left(\frac{T^N}{N!}\right)^{1/N} \underset{N\to\infty}{\sim} \frac{eT}{N}.$$

If $\gamma < 1/2$ then one takes instead

$$\varepsilon = \varepsilon^* = \left(\frac{T^N\, N^{1/2-\gamma}}{N!}\right)^{1/N} \underset{N\to\infty}{\sim} \frac{eT}{N},$$

with the same asymptotic behavior.

In both cases

$$\begin{aligned} \|\mu_0 - \mu_0^f\|_{L^\infty([0,T],L^2(\mathbb{R}))} &\leq C_{T,q,\lambda,k}\, N^\gamma\, (\varepsilon^*)^k \\ &\underset{N\to\infty}{\sim} \frac{C_{T,q,\lambda,k}}{N^{k-\gamma}}, \end{aligned} \tag{4.11}$$

which concludes the proof of Theorem 2.3. □

5 Proof of Th. 1.1: Example of the Tchebychev points

We now make the assumptions in Th. 1.1 and in particular that $d = 0$ and that the λ_k satisfy (1.16).

The constant $\Lambda_{N,0}$ defined by (2.9) satisfies

$$\Lambda_{N,0} = (N+1)^{1/2}. \tag{5.1}$$

Moreover we define

$$\rho_N(v) = \frac{\pi}{N+1}\sqrt{1-v^2}.$$

In fact, the function $\frac{1}{\rho_N}$ is a normalization of the Tchebychev weight.

As the λ_k gives the usual method of integration we have

Proposition 5.1 *We have, for all $N \geq 1$ and for all $R \in \mathbb{R}_{2N+1}[X]$:*

$$\int_I \frac{R(v)}{\rho_N(v)} dv = \frac{N+1}{\pi} \int_{]-1,1[} \frac{R(v)}{\sqrt{1-v^2}} dv = \sum_{k=0}^{N} R(\lambda_k), \qquad \forall R \in \mathbb{R}_{2N+1}[X]. \tag{5.2}$$

We may therefore apply Th. 2.1 to this choice of λ_k and for this choice of ρ_N.

Compute

$$C_N(q) = \sqrt{N+1} \left(\int_{-1}^{1} |q(v)|^2 \rho_N(v) \, dv \right)^{1/2} \leq C \, \|q\|_{L^2}.$$

Similarly

$$C_N(f^0) \leq C \, N^{-1/2} \, \|f^0\|_{L^2}.$$

So by Th. 2.1

$$\sup_{i \leq N} \|\mu_i(t)\|_{L^2} \leq C \, e^{C\,T\,\|q\|_{L^2}} \, \|f^0\|_{L^2} \, N^{-1/2} \left(\sum_{k \leq N} \lambda_k^{2i} \right)^{1/2} \leq C \, e^{C\,T\,\|q\|_{L^2}} \, \|f^0\|_{L^2},$$

which is exactly (1.18). Moreover

$$\sum_{i \leq N} \|\mu_i(t)\|_{L^2} \leq C \, \|f^0\|_{L^2} \, N^{-1/2} \sum_{i \leq N} \left(\sum_{k \leq N} \lambda_k^{2i} \right)^{1/2}.$$

Of course by Cauchy-Schwarz, we have, for all $0 \leq L \leq N$,

$$N^{-1/2} \sum_{i \leq N} \left(\sum_{k=L\ldots N-L} \lambda_k^{2i} \right)^{1/2} \leq \left(\sum_{i \leq N} \sum_{k=L\ldots N-L} \lambda_k^{2i} \right)^{1/2}$$
$$\leq \left(\sum_{L \leq k \leq N-L} \frac{1}{1-|\lambda_k|^2} \right)^{1/2}$$

Note that

$$1 - |\lambda_k|^2 \geq \frac{(k+1)^2}{C\,N^2} \quad \text{if } k \leq N/2, \quad 1 - |\lambda_k|^2 \geq \frac{(N-k+1)^2}{C\,N^2} \quad \text{if } k \geq N/2.$$

Hence, if $L \leq N/2$,

$$N^{-1/2} \sum_{i \leq N} \left(\sum_{k=L \dots N-L} \lambda_k^{2i} \right)^{1/2} \leq C \left(\sum_{L \leq k \leq N/2} \frac{N^2}{(k+1)^2} \right)^{1/2} \leq C \, \frac{N}{L^{1/2}}.$$

Therefore

$$\sum_{i \leq N} \|\mu_i(t)\|_{L^2} \leq C \, \|f^0\|_{L^2} \, (N^{1/2} \, L^{1/2} + N \, L^{-1/2}),$$

and choosing $L = \sqrt{N}$ we obtain that this method satisfies the estimate (2.18) with $\gamma = 3/4$. It only remains to apply Th. 2.3 to conclude. $\square$

6 Appendix

The natural way to regularize is by convolution. However to obtain high order approximation, it is necessary to choose correctly the mollifier. In the L^2 framework though, things are quite simple by truncating in Fourier.

Proposition 6.1 *Let k a positive integer, $f \in H^k(\mathbb{R}^d)$ and $\varepsilon > 0$. It exists $f_\varepsilon \in H^\infty(\mathbb{R}^d)$ such that*

$$\|f - f_\varepsilon\|_{L^2(\mathbb{R}^d)} \leq \varepsilon^k \|D^k f\|_{L^2(\mathbb{R}^d)}, \tag{6.1}$$

$$\|D^m f_\varepsilon\|_{L^2(\mathbb{R}^d)} \leq \varepsilon^{k-m} \|D^k f\|_{L^2(\mathbb{R}^d)}, \quad \forall m \in \mathbb{N}. \tag{6.2}$$

Proof : We use Fourier's analysis. We consider $f_\varepsilon \in L^2(\mathbb{R}^d)$ defined by

$$\widehat{f_\varepsilon}(\xi) = \widehat{f}(\xi) \mathbb{1}_{\{|\xi| \leq 1/\varepsilon\}}.$$

First, we have

$$\begin{aligned} \|\widehat{f} - \widehat{f_\varepsilon}\|_{L^2(\mathbb{R}^d)} &= \left(\int_{|\xi| > 1/\varepsilon} |\widehat{f}(\xi)|^2 d\xi \right)^{1/2} \leq \left(\int_{\mathbb{R}^d} \varepsilon^{2k} |\xi|^{2k} |\widehat{f}(\xi)|^2 d\xi \right)^{1/2} \\ &= \varepsilon^k \|\widehat{D^k f}\|_{L^2(\mathbb{R}^d)}, \end{aligned}$$

which proves (6.1).

On the other hand, for all $m \in \mathbb{N}$,

$$\||\xi|^m \widehat{f}_\varepsilon\|_{L^2(\mathbb{R}^d)} = \left(\int_{|\xi|\le 1/\varepsilon} |\xi|^{2m} |\widehat{f}(\xi)|^2 d\xi\right)^{1/2} \le \left(\int_{\mathbb{R}^d} \varepsilon^{2k-2m} |\xi|^{2k} |\widehat{f}(\xi)|^2 d\xi\right)^{1/2}$$
$$= \varepsilon^{k-m} \|\widehat{D^k f}\|_{L^2(\mathbb{R}^d)},$$

thus, $f_\varepsilon \in H^m(\mathbb{R}^d)$ and the estimate (6.2) holds. □

Acknowledgements.
Y. Bourgault was supported by a NSERC Grant. P-E Jabin was partially supported by NSF Grant 1312142 and by the KI-Net research network, NSF Grant 1107444.

References

[1] P. L. Bhatnagar, E. P. Gross, and M. Krook. A model for collision processes in gases. I. small amplitude processes in charged and neutral one-component systems. *Phys. Rev.*, **94**(3), 511–525, 1954.

[2] G. A. Bird, Molecular Gas Dynamics and the Direct Simulation of Gas Flow. *Oxford Engineering Science Series, Vol. 42* (Oxford, 1994).

[3] A. V. Bobylev. The Chapman-Enskog and Grad methods for solving the Boltzmann equation. *Sov. Phys. Dokl.*, **27**(1), 29–31, 1982.

[4] F. Brini. Hyperbolicity region in extended thermodynamics with 14 moments. *Continuum Mech. Thermodyn.*, **13**(1), 1–8, 2001.

[5] Z. Cai, Y. Fan, and R. Li. Globally hyperbolic regularization of Grads moment system in one dimensional space. *arXiv: 1111.3409*, 2012.

[6] Z. Cai and R. Li. Numerical regularized moment method of arbitrary order for Boltzmann-BGK equation. *SIAM J. Sci. Comput.*, **32**(5), 2875–2907, 2010.

[7] Z. Cai, R. Li, and Y. Wang. An efficient NRxx method for Boltzmann-BGK equation. *J. Sci. Comput.*, **50**(1), 103–119, 2012.

[8] S. Chandrasekhar, Stochastic problems in physics and astronomy. *Rev. Mod. Phys.* **15**, 2–89, 1943.

[9] L. Desvillettes, Some applications of the method of moments for the homogeneous Boltzmann and Kac equations. *Archive Rat. Mech. Anal.* **123**, 387–404, 1993.

[10] H. Grad. On the kinetic theory of rarefied gases. *Comm. Pure Appl. Math.*, **2**(4), 331–407, 1949.

[11] P. Le Tallec and J. P. Perlat. Numerical analysis of Levermores moment system. Rapport de recherche 3124, INRIA Rocquencourt, March 1997.

[12] C.D. Levermore, Moment Closure Hierarchy for Kinetic Theories. *J. Statist. Phys.* **83**, no. 5-6, (1996), 1021–1065.

[13] C.D. Levermore, W.J. Morokoff, The Gaussian moment closure for gas dynamics. *SIAM J. Appl. Math.* **59** (1999), no. 1, 72–96.

[14] H.G. Othmer, S.R. Dunbar, W. Alt, Models of dispersal in biological systems. *J. Math. Biol.*, **26** (3), 263–298, 1988.

[15] B. Perthame, Boltzmann type schemes for gas dynamics and the entropy property. *SIAM J. Numer. Anal.* **27**, 1405–1421, 1990.

[16] J. Shen and T. Tang. Spectral and High-Order Methods with Applications, volume 3 of Mathematics Monograph Series. Science Press, Beijing, P. R. China, 2006.

[17] H. Struchtrup. Macroscopic Transport Equations for Rarefied Gas Flows: Approximation Methods in Kinetic Theory. Springer, 2005.

[18] H. Struchtrup and M. Torrilhon. Regularization of Grads 13 moment equations: Derivation and linear analysis. *Phys. Fluids*, **15**(9), 2668–2680, 2003.

[19] E. F. Toro. Riemann solvers and numerical methods for fluid dynamics - A practical introduction - 3nd edition. Springer, 2009.

[20] M. Torrilhon. Two dimensional bulk microflow simulations based on regularized Grads 13-moment equations. *SIAM Multiscale Model. Simul.*, **5**(3), 695–728, 2006.

[21] M. Torrilhon. Hyperbolic moment equations in kinetic gas theory based on multivariate Pearson-IV-distributions. *Commun. Comput. Phys.*, **7**(4), 639–673, 2010

[22] D. Vernon Widder, The Laplace transform. Princeton University Press 1946.

[23] G.M. Wing, An introduction to transport theory. New-York, Wiley, 1962.