J Stat Phys (2009) 136: 1105–1130
DOI 10.1007/s10955-009-9820-3

Some Ideas About Quantitative Convergence of Collision Models to Their Mean Field Limit

Rémi Peyre

Received: 8 May 2009 / Accepted: 3 September 2009 / Published online: 18 September 2009
© Springer Science+Business Media, LLC 2009

Abstract We consider a stochastic N-particle model for the spatially homogeneous Boltzmann evolution and prove its convergence to the associated Boltzmann equation when $N \longrightarrow \infty$, with non-asymptotic estimates: for any time $T > 0$, we bound the distance between the empirical measure of the particle system and the measure given by the Boltzmann evolution in a relevant Hilbert space. The control got is Gaussian, i.e. we prove that the distance is bigger than $xN^{-1/2}$ with a probability of type $O(e^{-x^2})$. The two main ingredients are a control of fluctuations due to the discrete nature of collisions and a kind of Lipschitz continuity for the Boltzmann collision kernel. We study more extensively the case where our Hilbert space is the homogeneous negative Sobolev space $\dot{H}^{-s}$. Then we are only able to give bounds for Maxwellian models; however, numerical computations tend to show that our results are useful in practice.

Keywords Kinetic theory · Spatially homogeneous Boltzmann equation · Mean field limit · Nonasymptotic bounds

Introduction

The Boltzmann equation was written down by L. Boltzmann [3] in 1872, five years after Maxwell's seminal paper [14], to describe the behaviour of a large number of gas molecules interacting by pairwise collisions. Proving rigorously the heuristic arguments of Boltzmann to get some convergence of the N-particle model to the continuous Boltzmann equation when $N \longrightarrow \infty$ is an extremely difficult challenge that mathematicians are still dealing with.

Here we are only going to handle the *spatially homogeneous* Boltzmann equation (also called *mean field Boltzmann equation*), in which one forgets the positions of the gas particles to concentrate only on the collision phenomenon. Then proving the convergence of the

R. Peyre (✉)
École Normale Supérieure de Lyon, Lyon, France
e-mail: rpeyre@phare.normalesup.org

N-particle system to the continuous equation is a typical mean field limit problem—a particle model is said to be *mean field* when each particle interacts with comparable strength with *all* the other ones. Such a problem, which was first proposed by Kac [11], is far more tractable than the original one, and convergence results, mostly qualitative, have already been obtained for it (see Sect. 6.4).

Here however we are interested in a *quantitative* and *non-asymptotic* version of these results. We would also like to set our results in an *infinite-dimensional* setting, that is, to say that not only any reasonable functional of the particle model converges to the corresponding functional of the limit system, but moreover that all these functionals converge *uniformly*. The quantitative convergence we are going to prove will even have an $N^{-1/2}$ speed, typical of the uniform central limit theory (see [7] about it).

Concerning concrete Boltzmann models, in the actual state of my work I am only able to use my results for Maxwellian systems, and moreover constants in convergence bounds deteriorate rapidly with time. However that does not seem to be a fundamental feature of my approach, and further improvements might overcome these issues.

Important Remark 0.1 There are *two* sides in this work. The first one, whose climax is Theorem 3.3, is abstract: it consists in showing how Hilbert spaces can be used to prove a new powerful type of convergence results for collision models like Boltzmann's. That work is *a priori* likely to be applied to a wide range of situations, but for each of them checking the hypotheses of the abstract theorem is a different challenge. The second side, which is more physical, consists in studying *one* particular case of application of our formulas, namely the Boltzmann model looked at in the $\dot{H}^{-s}$ space, for which we obtain precise numerical bounds (cf. Sect. 6.3). Though the results got for that particular choice can be proved to be definitely limited in some way (cf. Sect. 4.1), that may not be true any longer for a smarter choice of Hilbert space—which would however be more complicated to handle. So this article highlights a way of studying collision models, but remains at a simple level in the applications, hence the title "some ideas".

Here is some notation which will be used throughout this paper:

- The space $\mathbb{R}^d$ is equipped with its Euclidean structure, whose norm is denoted by $|\cdot|$.
- $f : E \longrightarrow F$ being a measurable function and μ a measure on E, the image measure of μ by f on F will be denoted $f\#\mu$.
- δ_x denotes a Dirac mass at x.
- $\mathcal{S}(\mathbb{R}^d)$ is the Schwartz space on $\mathbb{R}^d$, i.e. the set of (complex-valued) $\mathcal{C}^\infty$ functions on $\mathbb{R}^d$ which tend to 0 at infinity faster than any $|x|^{-k}$, as well as all their derivatives.
- The Fourier transform of a function $f \in \mathcal{S}(\mathbb{R}^d)$ is denoted by $\widehat{f}$, with the unitary convention $\widehat{f}(\xi) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} f(x) e^{-i\xi\cdot x} \mathrm{d}x$.
- The notation $\|\cdot\|$ will be used to denote Hilbert norms in functional spaces. If Q is a linear operator between two Hilbert spaces, its operator norm $\sup_{\|x\|\leqslant 1} \|Qx\|$ will be denoted $|||Q|||$.
- x, y and z being three points of an affine Hilbert space with $y, z \neq x$, $\widehat{yxz}$ denotes the angle between $\overrightarrow{xy}$ and $\overrightarrow{xz}$, which is an element of $[0, \pi]$.
- The identity matrix of size d is denoted $\mathbf{I}_d$.

1 The Model

1.1 The Microscopic Model

Let us describe the particle model for the spatially homogeneous Boltzmann evolution. Such models have been first proposed by Kac [11] and later thoroughly studied by Sznitman [21], Spohn [19] and others. There are N identical particles indexed by $0, \dots, N-1$, each particle i being characterized by its velocity $v_i \in \mathbb{R}^d$. One imposes random collision times, so that the microscopic evolution is a Markov process. The way two particles with respective velocities v and w hit each other is described by some positive measure $\gamma_{v,w}$ on $(\mathbb{R}^d)^2$, $N^{-1}\mathrm{d}\gamma_{v,w}(v', w')$ being the collision rate from state (v, w) to state (v', w'). In other words, the generator $\mathcal{L}$ of the Markov process is

$$\begin{aligned}&\mathcal{L}f(v_0, \dots, v_{N-1})\\&\quad = \frac{1}{2N} \sum_{0 \leqslant i,j < N} \int_{(\mathbb{R}^d)^2} \Big(- f(v_0, \dots, v_{N-1}) + f(\dots, v_i', \dots, v_j', \dots)\Big) \mathrm{d}\gamma_{v_i, v_j}(v_i', v_j').\end{aligned} \tag{1.1}$$

We may add to this model some extra physical conditions. First, we will always suppose that the momentum and energy are conserved by collisions, and that the model is invariant by velocity translation or rotation, i.e. that for all $v, w \in \mathbb{R}^d$, for any (positive) isometry J of $\mathbb{R}^d$;

$$\gamma_{v,w}\text{-a.e.} \quad v' + w' = v + w, \tag{1.2}$$

$$\gamma_{v,w}\text{-a.e.} \quad |w' - v'| = |w - v|, \tag{1.3}$$

$$\gamma_{Jv,Jw} = (J, J)^{\#}\gamma_{v,w}. \tag{1.4}$$

When conditions (1.2) to (1.4) are satisfied, the model is completely described by the family of measures $(\overline{\gamma}_u)_{u \in (0,\infty)}$ on $(0, \pi]$, where $\mathrm{d}\overline{\gamma}_u(\theta)$ is the proportion, by unit of time, of particles with relative speed u which undergo a collision making them deviate by an angle θ in the collision referential.

Moreover, it is often assumed that the $\overline{\gamma}_u$ have a scale invariance property, in the sense that there exists a real parameter g such that for any $\lambda \in (0, +\infty)$,

$$\overline{\gamma}_{\lambda u} = \lambda^g \overline{\gamma}_u. \tag{1.5}$$

For instance, the hard sphere model is scale-invariant with $g = 1$. Another very interesting particular case is when $g = 0$—then one says that the model is *Maxwellian*. In this article the concrete results obtained will actually concern Maxwellian models.

Before turning to the macroscopic model, let us make some remarks on the microscopic one:

Remark 1.1

1. The N^{-1} factor in (1.1) is essential to get the mean field limit: it morally says that the global collision rate of one particle is independent of the total number of particles.

2. Strictly speaking, generator (1.1) allows a particle to collide with itself, which is physically absurd. Yet because of the conservation law (1.3), the auto-collision term is actually zero, so there is no problem.
3. The $\gamma_{v,w}$ have to satisfy some integrability conditions for the Markov process to be well-defined. For instance, if conditions (1.2) to (1.5) are satisfied, then it suffices that for an arbitrarily chosen $u \in (0, \infty)$, $\int_0^\pi \theta^{d-1} \mathrm{d}\overline{\gamma}_u(\theta)$ is finite, cf. [20].

1.2 The Macroscopic Model

The macroscopic space-homogeneous Boltzmann equation [4] is obtained informally by letting N tend to infinity in the microscopic evolution. Then the particles' velocities are described by they empirical measure, which is a (non-atomic in general) probability measure μ_t on $\mathbb{R}^d$. The evolution of that measure is deterministic and is governed by the equation:

$$D_t \mu = Q(\mu_t, \mu_t), \tag{1.6}$$

where Q is the *Boltzmann collision kernel* of the system, formally defined by:

$$Q(\mu, \nu) = \frac{1}{2} \int \left(\int \left(-\delta_v - \delta_w + \delta_{v'} + \delta_{w'} \right) \mathrm{d}\gamma_{v,w}(v', w') \right) \mathrm{d}\mu(v) \mathrm{d}\nu(w). \tag{1.7}$$

Equation (1.6) is an ordinary differential equation in an infinite-dimensional space; that equation is non-linear because of the quadratic term $Q(\mu, \mu)$. Unique existence of a solution to it has been thoroughly studied over the last decades [6, 22]. For our theory to work, we will need to consider a setting where that unique existence is achieved in some convenient space—which is quite logical altogether. Later we will see concrete examples where (1.6) behaves well for our purpose.

1.3 Conservation Laws, Convergence to Equilibrium

Because of the conservation laws (1.2) and (1.3), we get $d + 1$ invariant functions for the microscopic system: the first d are synthetised in the momentum $P = \sum_{i=0}^{N-1} v_i$, and the last one is the energy $K = \frac{1}{2} \sum_{i=0}^{N-1} |v_i|^2$. In the macroscopic model, these invariants become $p = \int v \mathrm{d}\mu(v)$ and $k = \frac{1}{2} \int |v|^2 \mathrm{d}\mu(v)$. Moreover the fact that the macroscopic model derives from the description of an evolution of particles implies two extra properties for it: first *positivity* of (1.6), which means that if μ_0 is a positive measure, then so are the μ_t for t positive; secondly *conservation of mass* which gives the $(d + 2)$-nd invariant $m = \int \mathrm{d}\mu(v)$ for the macroscopic equation.

Concerning equilibrium, if we impose some minimal non-degeneracy condition (see [22]), then it is a well-known beautiful result due to Boltzmann [3] that (1.6) is dissipative for positive measures and converges to an equilibrium measure μ_{eq} depending only on p, k and m: for $m = 1$ and $p = 0$, it is

$$\mathrm{d}\mu_{\mathrm{eq}}(v) = \left(\frac{d}{4\pi k} \right)^{d/2} e^{-d|v|^2/4k} \mathrm{d}v, \tag{1.8}$$

and it has the invariance properties $\mu_{\mathrm{eq}}(p, k, 1) = \tau_p{}^{\#} \mu_{\mathrm{eq}}(0, k - p^2/2, 1)$, τ_p being the translation by vector p, and $\mu_{\mathrm{eq}}(\lambda p, \lambda k, \lambda m) = \lambda \mu_{\mathrm{eq}}(p, k, m)$. More recently a beautiful quantitative version of that convergence result has been proved by Carlen, Gabetta and Toscani [5].

For the microscopic model, there is also a unique ergodic equilibrium measure for each value of P and K (N being fixed), which is merely the uniform measure on the $(dN-d-1)$-dimensional sphere[1] of $(\mathbb{R}^d)^N$ made of the N-uples of vectors having these P and K. Note that for N particles with momentum Np and energy Nk, the marginals of that measure tend to the continuous equilibrium measure $\mu_{\mathrm{eq}}(p,k,1)$ when $N \longrightarrow \infty$.

Finally it is worth recalling that the microscopic process is reversible under its equilibrium measure, while on the contrary the macroscopic equation (1.6) exhibits a dissipative behaviour—a phenomenon which caused much trouble at Boltzmann's time, but has been well understood today.

2 Homogeneous Sobolev Spaces

2.1 Why Homogeneous Sobolev Spaces?

To be able to speak of quantitative convergence, we will work in some Banach space. Which one will we take? As we want to compare the empirical measure of our particle system to its limit evolution, a natural choice is to take some coupling distance between measures—say, the W_1 Wasserstein distance [23, Sect. 7], defined for μ, ν two positive measures with the same mass by:

$$W_1(\mu,\nu) = \sup_{f \text{ 1-Lip.}} \left| \int f \mathrm{d}(\nu-\mu) \right|, \tag{2.1}$$

where "f 1-Lip." means that the supremum is taken over all 1-Lipschitz functions on $\mathbb{R}^d$. However it turns out that it is hopeless to get an $N^{-1/2}$ rate of convergence in such a space, because testing $\nu-\mu$ against so much test functions makes the uniform central limit theory fail: see [7, Sect. 6.4] for more details. We also give a more intuitive, completely different explanation of that fact in Appendix.

Thus the idea is to test $\nu-\mu$ against a smaller space made of more regular functions. Sobolev spaces $W^{s,p}$, $s>0$, are such natural test spaces; then $\nu-\mu$ will be seen as an element of the dual space $W^{-s,p/(p-1)}$. For our theory we will have to work in a Hilbert space, so we take $p=2$ and work in $W^{-s,2}=H^{-s}$; then we can take s fractional, which will turn out to be useful indeed. Yet since defining a norm for H^{-s} spaces requires to choose some arbitrary length, which is physically annoying, we will rather consider *homogeneous* $\dot{H}^{-s}$ spaces, which do have a canonical norm (plus other advantages). Note however, cf. Remark 0.1, that this choice is only one possibility—certainly particularly reasonable—among other ones, and that trickier choices might also be relevant.

2.2 Definition and Useful Properties

Let us define properly the $\dot{H}^{-s}$ spaces.

Definition 2.1 Let $s \in \mathbb{R}$, and for $f \in \mathcal{S}(\mathbb{R}^d)$, set

$$\|f\|_{\dot{H}^{-s}} = \left(\int_{\mathbb{R}^d} |\widehat{f}(\xi)|^2 |\xi|^{-2s} \mathrm{d}\xi \right)^{1/2}. \tag{2.2}$$

[1] Possibly of radius 0.

Then those of the $f \in \mathcal{S}(\mathbb{R}^d)$ for which $\|f\|_{\dot{H}^{-s}} < \infty$, equipped with the norm $\|\cdot\|_{\dot{H}^{-s}}$, constitute a pre-Hilbert space with scalar product

$$\langle f, g\rangle_{\dot{H}^{-s}} = \int_{\mathbb{R}^d} \widehat{f}(\xi)\overline{\widehat{g}}(\xi)|\xi|^{-2s}\mathrm{d}\xi. \tag{2.3}$$

The Hilbert space obtained by completing it is denoted $\dot{H}^{-s}$.

Remark 2.2 For a physicist, $f : \mathbb{R}^d \longrightarrow \mathbb{C}$ has some homogeneity: say, the elements in $\mathbb{R}^d$ are measured in x (generally x is a unit of length, say meters) and the elements in $\mathbb{C}$ are measured in y (which will often be a density, say kg·m^{-d}). Then $\|f\|_{\dot{H}^{-s}}$ is measured in y·x$^{s+d/2}$ (in our example, $\|f\|_{\dot{H}^{-s}}$ would be measured in kg·m$^{s-d/2}$). Equivalently, if μ is a measure on $\mathbb{R}^d$, the physical dimension of $\|\mu\|_{\dot{H}^{-s}}$ is z·x$^{s-d/2}$, x being the physical dimension of the elements of $\mathbb{R}^d$ and z the physical dimension of μ (which in our example would be kg).

As we told in Sect. 2.1, bounding a function or a measure in $\dot{H}^{-s}$ means bounding uniformly its integral against some class of regular functions:

Proposition 2.3 *Define $\dot{H}^s$ in the same way as $\dot{H}^{-s}$. Then, for any f for which it makes sense*:

$$\|f\|_{\dot{H}^{-s}} = \sup_{\|g\|_{\dot{H}^s} \leqslant 1} \left| \int_{\mathbb{R}^d} f(x)\overline{g}(x)\mathrm{d}x \right|. \tag{2.4}$$

Lemma 2.4 *For $s \in]0, d[$, let ϕ_s be the locally integrable function*

$$\phi_s(x) = |x|^{-(d-s)}, \tag{2.5}$$

then one has for all $f, g \in \mathcal{S}(\mathbb{R}^d)$:

$$\langle f, g\rangle = c(s,d)^2 \big\langle f * \phi_s, g * \phi_s \big\rangle_{L^2(\mathbb{R}^d)}, \tag{2.6}$$

with

$$c(s,d) = \frac{\Gamma((d-s)/2)}{(2\pi)^{d/2}\Gamma(s/2)}, \tag{2.7}$$

$\Gamma(\cdot)$ *being Euler's Gamma function.*

Proof Use that the Fourier transform of $|\xi|^{-s}$ is $(2\pi)^{d/2}c(s,d)\phi_s(x)$, cf. [18, exercise V-10]. □

Immediate proposition 2.5 *Let J_λ be a similarity of $\mathbb{R}^d$ with dilation factor λ, then for any map $f \in \dot{H}^{-s}$,*

$$\|f \circ J_\lambda\|_{\dot{H}^{-s}} = \lambda^{s+d/2}\|f\|_{\dot{H}^{-s}}. \tag{2.8}$$

Equivalently, for any measure $\mu \in \dot{H}^{-s}$,

$$\|J_\lambda{}^{\#}\mu\|_{\dot{H}^{-s}} = \lambda^{s-d/2}\|\mu\|_{\dot{H}^{-s}}. \tag{2.9}$$

☛ *From now on, we will always write implicitly $s = d/2 + r$.*

Proposition 2.6 *Suppose $d \geqslant 2$,*[2] *and let μ be a compactly supported signed measure on $\mathbb{R}^d$ with total mass 0, then for any $r \in (0,1)$, μ can be seen as an element of $\dot{H}^{-s}$.*

Proof Thanks to Lemma 2.4 we just need to prove that $\mu * \phi_s$ is a square-integrable function. Suppose that μ is supported by the ball $B(R)$ of radius R centered at 0 and splits into $\mu_+ - \mu_-$ with μ_+ and μ_- positive measures each of total mass M. Then for $\rho > 0$, on $B(\rho)$ $\mu * \phi_s$ is equal to $\mu * (\mathbb{1}_{B(R+\rho)}\phi_s)$, so the L^2 norm of $\mathbb{1}_{B(\rho)}(\mu * \phi_s)$ is bounded above by $2M \cdot \|\mathbb{1}_{B(R+\rho)}\phi_s\|_{L^2} < \infty$. Thus $\mu * \phi_s$ is locally L^2. On the other hand, for $|x| = \rho > R$,

$$|(\mu * \phi_s)(x)| \leqslant M\left(\frac{1}{(\rho - R)^{d/2-r}} - \frac{1}{(\rho + R)^{d/2-r}}\right) \leqslant 2MR\frac{d/2 - r}{(r - R)^{d/2+1-r}}, \tag{2.10}$$

so $\mu * \phi_s$ is L^2 at infinity, which finishes the proof. □

Corollary 2.7 *Still suppose $d \geqslant 2$, then for $r \in (0,1)$, any signed measure with zero total mass, if it has an r-th polynomial momentum, can be seen as en element of $\dot{H}^{-s}$.*

Proof Let $\mu = \mu_+ - \mu_-$ be such a measure with its Hahn decomposition, μ_+ and μ_- each having total mass M. Then the integral Minkowski inequality gives

$$\begin{aligned}\|\mu\|_{\dot{H}^{-s}} &\leqslant \frac{1}{M}\int_{(\mathbb{R}^d)^2} \|\delta_x - \delta_y\|_{\dot{H}^{-s}} \mathrm{d}\mu_+(x)\mathrm{d}\mu_-(y) \\ &= \frac{C_r}{M} \cdot \int_{(\mathbb{R}^d)^2} |x - y|^r \mathrm{d}\mu_+(x)\mathrm{d}\mu_-(y) < \infty,\end{aligned} \tag{2.11}$$

C_r being the $\dot{H}^{-s}$ norm of any $\delta_x - \delta_y$ for $|x - y| = 1$, which is finite by the previous proposition. □

Remark 2.8 The $\dot{H}^{-s}$ norm allows us to measure the *distance* between two (sufficiently integrable) probability measures, but speaking of the $\dot{H}^{-s}$ norm of a *single* probability measure would be nonsense! Note also that, by Sobolev imbedding, one can bound above $\|\nu - \mu\|_{\dot{H}^{-s}}$, for any two probability measures μ and ν, by (up to some explicit multiplicative constant)

$$W_{1,r}(\mu, \nu) = \sup\left\{\left|\int f \mathrm{d}\mu - \int f \mathrm{d}\nu\right| ;\ \forall x, y\ |f(x) - f(y)| \leqslant |y - x|^r\right\}. \tag{2.12}$$

3 Dynamic Control

3.1 Abstract Setting

Now let us study the evolution of our particle system along time. We first give our main result in an abstract setting to alleviate its proof; the reader more comfortable with physical settings may read Theorem 3.6 instead.

[2]The proposition remains valid with $d = 1$, except that it must be demanded that $r < 1/2$.

Let H be a Hilbert space, let A be an H-affine space and let $(\widehat{X}_t)_{t\geqslant 0}$ be some jump Markov process on A with generator $\mathcal{L}$. Fix o an arbitrary point of A and define

$$\begin{array}{rccc} I: & A & \longrightarrow & H \\ & x & \mapsto & \overrightarrow{ox}. \end{array} \tag{3.1}$$

Since I is defined up to an additive constant, the operator[3] $(\mathcal{L}I): A \longrightarrow H$ does not depend on the choice of o and we can therefore define $(X_t)_{t\geqslant 0}$ as the deterministic process on A following the differential equation

$$D_t X = \big(\mathcal{L}I\big)(X_t). \tag{3.3}$$

Our goal is to control the distance between $\widehat{X}_t$ and X_t. Here what is important for us is to have a good control of large deviations for that distance. As Cramér's method cannot be applied directly because of the infinite-dimensional setting, we introduce an exponential utility function $\mathcal{U}: H \longrightarrow \mathbb{R}$ defined by:

$$\mathcal{U}(x) = e^{\|x\|} + e^{-\|x\|}. \tag{3.4}$$

The following proposition gathers the properties of $\mathcal{U}$ we will use in our work:

Immediate proposition 3.1

1. *For all* $x \in H, \mathcal{U}(x) \geqslant e^{\|x\|}$;
2. $\mathcal{U}(0) = 2$;
3. *For all* $x, h \in H, \mathcal{U}(x+h) \leqslant e^{\|h\|}\mathcal{U}(x)$;
4. $\mathcal{U}$ *is of class* $\mathcal{C}^\infty$;[4]
5. *For all* $x \in H$, $\nabla\mathcal{U}(x)$ *is positively colinear to* x;
6. *For all* $x \in H$, $|||\nabla^2\mathcal{U}||| \leqslant \mathcal{U}(x)$.

Then one can state the theorem which will be our central tool. We first need some notation to alleviate our formulas:

Definition 3.2 We denote $e_1(t) = (e^t - 1)/t$, extended by $e_1(0) = 1$, resp. $e_2(t) = (e^t - 1 - t)/t^2$, extended by $e_2(0) = 1/2$. We also denote κ_- the negative part of κ, i.e. $\kappa_- = \max\{-\kappa, 0\}$.

Theorem 3.3 *Suppose that* (3.3) *has a* κ*-contracting semigroup for some* $\kappa \in \mathbb{R}$, *in the sense that for all* $x \in A, h \in H$:

$$\big\langle D_x(\mathcal{L}I)\cdot h, h\big\rangle \leqslant -\kappa\|h\|^2. \tag{3.5}$$

[3] *Stricto sensu* $\mathcal{L}$ acts on some space of *real* functions on A, say the space of continuous bounded functions $\mathcal{C}_b(A,\mathbb{R})$. Yet we can straightforwardly extend $\mathcal{L}$ to the space $\mathcal{C}_b(A,E)$ for any Banach space E by defining the operator $\mathcal{L}^{(E)}: \mathcal{C}_b(A,E) \longrightarrow \mathcal{C}_b(A,E)$ through:

$$\forall\phi \in E'\ \forall f \in \mathcal{C}_b(A,E) \quad \big\langle \phi, \mathcal{L}^{(E)} f\big\rangle = \mathcal{L}\big(\langle\phi, f\rangle\big). \tag{3.2}$$

That is what we do here: I is a function from A to H, so $\mathcal{L}I$ actually denotes $\mathcal{L}^{(H)}I$.

[4] To prove it, note that $\mathcal{U}(x) = f(\|x\|^2)$, where $f = 2\cosh(\sqrt{\cdot})$ is (the restriction to $[0,+\infty)$ of) an analytic function on $\mathbb{R}$.

Suppose moreover that the Markov process—which we recall to be a jump process—has the amplitude of all its jumps bounded above by some $L < \infty$, and satisfies:

$$\forall x \in A \quad \mathcal{L}(\|\cdot - x\|^2)(x) \leqslant V \tag{3.6}$$

for some $V < \infty$.

Then, denoting $\widehat{X}_0$ the (random) initial value of the Markov process and X_0 the (deterministic) initial value of the differential equation (3.3), one has for any $T \geqslant 0$, for any $\lambda > 0$:

$$\begin{aligned}&\ln \mathbb{E}\big[\mathcal{U}\big(\lambda(\widehat{X}_T - X_T)\big)\big]\\ &\quad\leqslant \ln \mathbb{E}\big[\mathcal{U}\big(\lambda e^{-\kappa T}(\widehat{X}_0 - X_0)\big)\big] + \lambda^2 e_2(\lambda e^{2\kappa - T} L) e_1(-2\kappa T) V T.\end{aligned} \tag{3.7}$$

Proof The principle of the proof is to show that some time-depending functional

$$F(\widehat{X}_t) = e^{h(t)} \mathcal{U}\Big(\lambda e^{\kappa(t-T)}(\widehat{X}_t - X_t)\Big), \tag{3.8}$$

for a well-chosen function h, is a supermartingale.

☛ *To make our computations completely rigorous, throughout the proof we will assume that the expected number of collisions per unit of time is uniformly bounded, that is, that there is some $M < \infty$ such that $|(\mathcal{L}\mathbb{1}_{A'})(x)| \leqslant M$ for all $x \in A$ and any Borel subset $A' \subset A$. Then the general result can be recovered by a standard truncation argument.*

Let us fix some $t \in [0, T]$ and suppose $(\widehat{X}_{t'})_{t' \in [0,t]}$ is known. Let δt be a small amount of time devised to tend to 0; $O(\delta t^n)$ will denote any quantity bounded by some $C\delta t^n$ when δt tends to 0, where C depends only on κ, V, M, λ, T, t and $\|X_t\|$. With this notation, the law of $\widehat{X}_{t+\delta t}$ depends on $(\widehat{X}_{t'})_{t' \in [0,t]}$ only through $\widehat{X}_t$, and our goal is to show that $\mathbb{E}[F(\widehat{X}_{t+\delta t})] - F(\widehat{X}_t)$, which is $O(\delta t)$, is nonpositive—more precisely, we only need to prove that $\mathbb{E}[F(\widehat{X}_{t+\delta t})] - F(\widehat{X}_t) \leqslant O(\delta t^2)$.[5]

Set $\widehat{Y} = \widehat{X} - X$. Denote $\delta\widehat{X} = \widehat{X}_{t+\delta t} - \widehat{X}_t$, resp. $\delta X = X_{t+\delta t} - X_t$, $\delta\widehat{Y} = \widehat{Y}_{t+\delta t} - \widehat{Y}_t$, $\delta F = F(\widehat{X}_{t+\delta t}) - F(\widehat{X}_t)$. The fundamental observation is that

$$\mathbb{E}\big[\delta\widehat{X}\big] = \big(\mathcal{L}I\big)(\widehat{X}_t)\delta t + O(\delta t^2). \tag{3.9}$$

Now, admitting temporarily that h will be of class $\mathcal{C}^2$, we write:

$$\begin{aligned}\delta F &= h'(t) F(t) \delta t && (3.10)\\ &\quad + e^{h(t)} \lambda e^{\kappa(t-T)} \nabla\mathcal{U}(\lambda e^{\kappa(t-T)}\widehat{Y}_t) \cdot \big(\mathcal{L}I(\widehat{X}_t) - \mathcal{L}I(X_t) + \kappa\widehat{Y}_t\big)\delta t && (3.11)\\ &\quad + e^{h(t)}\Big[\mathcal{U}\big(\lambda e^{\kappa(t-T)}\widehat{Y}_{t+\delta t}\big) - \mathcal{U}\Big(\lambda e^{\kappa(t-T)}\{\widehat{Y}_t + [\mathcal{L}I(\widehat{X}_t) - \mathcal{L}I(X_t)]\delta t\}\Big)\Big] && (3.12)\\ &\quad + O(\delta t^2).\end{aligned}$$

In that sum we first see that the term (3.11) is nonpositive: (3.5) indeed implies, for all $x \in A$, $y \in H$,

$$\big\langle (\mathcal{L}I)(x + y) - (\mathcal{L}I)(x) + \kappa y, y\big\rangle \leqslant 0, \tag{3.13}$$

[5] Beware that "*expr.* $\leqslant O(\delta t^n)$" does not mean "*expr.* $= O(\delta t^n)$" but actually "$(\textit{expr.})_+ = O(\delta t^n)$".

which we apply here with $x = X_t$ and $y = \widehat{Y}_t$, using that $\nabla\mathcal{U}(\lambda e^{\kappa(t-T)}\widehat{Y}_t)$ is positively colinear to $\widehat{Y}_t$ (Proposition 3.1-5).

Now let us look at term (3.12). Because of (3.9), the expectation of the random variable

$$\lambda e^{\kappa(t-T)}\Big(\widehat{Y}_{t+\delta t} - \big(\widehat{Y}_t + [\mathcal{L}I(\widehat{X}_t) - \mathcal{L}I(X_t)]\delta t\big)\Big) \tag{3.14}$$

is $O(\delta t^2)$. We will use it thanks to the following

Lemma 3.4 *Let $X \in H$; let y be an H-valued random variable with zero mean. Then one has*:

$$\mathbb{E}[\mathcal{U}(X+y)] \leqslant \mathcal{U}(X)\big(1 + \mathbb{E}\big[e_2(\|y\|)\|y\|^2\big]\big). \tag{3.15}$$

Proof of the lemma Taylor's formula yields

$$\mathcal{U}(X+y) = \mathcal{U}(X) + \nabla\mathcal{U}(X)\cdot y + \left(\int_0^1 (1-\theta)\nabla^2\mathcal{U}(X+\theta y)\mathrm{d}\theta\right)\cdot(y\otimes y). \tag{3.16}$$

In that sum the third term is bounded above by

$$\|y\|^2\int_0^1 (1-\theta)\mathcal{U}(X+\theta y)\mathrm{d}\theta \tag{3.17}$$

by Proposition 3.1-6, which in turn is bounded by

$$\|y\|^2\mathcal{U}(X)\int_0^1 (1-\theta)e^{\theta\|y\|}\mathrm{d}\theta \;=\; e_2(\|y\|)\|y\|^2. \tag{3.18}$$

by Proposition 3.1-3. Taking expectation gives the result since the second term in sum (3.16) has zero mean by assumption. □

What does it give for us? Let E be the event "some collision occurs between t and $t+\delta t$". E is an event of probability $O(\delta t)$, on E, the random variable (3.14) is $O(1)$, and on ${}^cE$ it is $O(\delta t)$. Hence, denoting temporarily $*$ for that variable, $\mathbb{E}[\|*\|^2 e_2(\|*\|)]$, up to some $O(\delta t^2)$, is merely $\lambda^2 e^{2\kappa(t-T)}\mathbb{E}[\|\delta\widehat{Y}\|^2 e_2(\|\lambda e^{\kappa(t-T)}\delta\widehat{Y}\|)]$, which is bounded above by $\lambda^2 e^{2\kappa(t-T)}e_2(\lambda e^{2\kappa_- T}L)V$ uniformly in t.

Putting all things together, we get

$$\mathbb{E}[\delta F] \leqslant \Big(h'(t) + \lambda^2 e^{2\kappa(t-T)}e_2(\lambda e^{2\kappa_- T}L)V\Big)F(t)\delta t + O(\delta t^2), \tag{3.19}$$

which will be $\leqslant O(\delta t^2)$ provided

$$h'(t) \leqslant -\lambda^2 e^{2\kappa(t-T)}e_2(\lambda e^{2\kappa_- T}L)V. \tag{3.20}$$

To achieve that optimally with $h(T) = 0$, we choose

$$h(t) = \lambda^2 e_2\big(\lambda e^{2\kappa_- T}L\big)e_1\big(2\kappa(t-T)\big)V(T-t), \tag{3.21}$$

which is of class $\mathcal{C}^2$ indeed. Formula (3.7) then follows by the supermartingale property. □

Remark 3.5 Strictly speaking our proof only shows that $F(\widehat{X}_t)$ is a *local* supermartingale. But this local supermartingale is nonnegative, so it is actually a global supermartingale (see [17, Sect. IV-1.5]).

3.2 Application to Boltzmann's Model

Translation of Theorem 3.3 Let us see what Theorem 3.3 gives for the Boltzmann model. For the time being, according to Remark 0.1 we do not precise what H and A are: all you have to know is that H is some Hilbert space of measures and A the corresponding affine space, which is assumed to contain all the probability measures having enough moments.

Let $N \in \mathbb{N}^\star$. The stochastic process $(\widehat{X}_t)_{t\geqslant 0}$ on A will be the empirical measure $\widehat{\mu}_t^N$ of our microscopic process $(v_0(t), \dots, v_{N-1}(t))$ on $(\mathbb{R}^d)^N$. It is a Markov process indeed; let us denote $\mathcal{L}^N$ its generator. Regardless of N, one has as expected:

$$\forall \mu \in A \quad \big(\mathcal{L}^N I\big)(\mu) = Q(\mu, \mu), \tag{3.22}$$

so the deterministic process $(X_t)_{t\geqslant 0}$ on A will be our macroscopic process following the Boltzmann equation (1.6). Finally, for $\mu \in A, \nu \in H$, $(D_\mu(\mathcal{L}I)) \cdot \nu = 2Q(\mu, \nu)$.[6] So Theorem 3.3 becomes:

Theorem 3.6 *Let H be a Hilbert space of measures and A the corresponding affine space containing probability measures (or a subset of it, cf. Remark 3.7). Consider our microscopic and macroscopic models for some $N \in \mathbb{N}^\star$, with certain initial conditions*[7] $(v_0, \dots, v_{N-1})$, *resp.* μ_0.

Suppose that there exists some constants $\kappa \in \mathbb{R}$, $L < \infty$, $V < \infty$ such that:

1. *For all $\mu \in A, \nu \in H$,*

$$\big\langle Q(\mu, \nu), \nu \big\rangle \leqslant -\frac{\kappa}{2}\|\nu\|^2; \tag{3.23}$$

2. *For all $\mu \in A$,*

$$\mathbb{E}\big[\|\mu_{t+} - \mu_t\|^2 \big| \mu_t = \mu\big] \leqslant V; \tag{3.24}$$

3. *The effect of collisions for the microscopic model in A is always bounded by L, i.e. one has almost surely*

$$\forall t \geqslant 0 \quad \big\|\widehat{\mu}_{t+}^N - \widehat{\mu}_t^N\big\| \leqslant L. \tag{3.25}$$

Then for any $T \geqslant 0$, for any $\lambda > 0$,

$$\begin{aligned} &\ln \mathbb{E}\big[\mathcal{U}\big(\lambda(\widehat{\mu}_T - \mu_T)\big)\big] \\ &\quad \leqslant \ln \mathbb{E}\big[\mathcal{U}\big(\lambda e^{-\kappa T}(\widehat{\mu}_0 - \mu_0)\big)\big] + \lambda^2 e_2(\lambda e^{2\kappa_- T} L) e_1(-2\kappa T) V T. \end{aligned} \tag{3.26}$$

Remark 3.7 Theorem 3.6 remains valid, with the same proof, if we replace A by any subset $\widetilde{A} \subset A$ such that almost surely $\forall t \geqslant 0 \quad \widehat{X}_t^N, X_t \in \widetilde{A}$. An important example of it is that, when A contains nonpositive measures, one can always take for $\widetilde{A}$ the subset of the true probability measures of A (which subset is stable because of positivity and conservation of mass for the evolutions, cf. Sect. 1.3), for which the properties of positive measures can be used.

[6] *Stricto sensu* $Q(\mu, \cdot)$ is an affine operator from A to H, not a linear operator on H: in fact here $Q(\mu, \nu)$ denotes $\overline{Q}(\mu, \nu)$, $\overline{Q}(\mu, \cdot)$ being the linear part of $Q(\mu, \cdot)$. Identifying notations is relevant because $\overline{Q}$, like Q, is formally defined by (1.7).

[7] The initial condition for the stochastic process can be random.

Constants for the Sobolev Setting

☛ *From now on, when dealing with Boltzmann models we work in the space* $\dot{H}^{-s}(\mathbb{R}^d)$ *for some* $r \in (0,1)$. *We denote by* C_r *the* $\dot{H}^{-s}$ *norm of any* $\delta_x - \delta_y$ *for* $|x-y|=1$, *which is some finite explicit function of* d *and* r.

To apply Theorem 3.6, we have to compute the values of L, V, κ and $\mathbb{E}[\mathcal{U}(\lambda e^{-\kappa T} \times (\widehat{X}_0 - X_0))]$. Here let us just look at the first two quantities—the last two ones will be the objects of separate sections.

Recall that K denotes the energy of the N-particle system, which is conserved along the stochastic evolution—note by the way that up to translating the origin of $\mathbb{R}^d$, we can replace K by the internal energy

$$\widetilde{K} = K - \frac{|P|^2}{2N}. \tag{3.27}$$

Then at any time no particle has speed greater than $\sqrt{2K}$, so the effect of a collision between two particles on the empirical measure can be no more than $2 \cdot (8K)^{r/2} C_r N^{-1}$, which yields an admissible value for L.

Remark 3.8 To get the bound $L \leqslant 2 \cdot (8K)^{r/2} C_r N^{-1}$ we have used that the relative speed between two particles is at most $2\sqrt{2K}$, and that the effect of a collision with relative speed u is at most $2u^r N^{-1}$. Actually one can do slightly better: the relative speed between two particles is at most $2\sqrt{K}$ and the effect of a collision with relative speed u is at most $2\sqrt{2^{1-r}-1}C_r u^r/N$ (corresponding to the deviation angle $\theta = \pi/2$), so we could have taken

$$L = 2^{1+r}\sqrt{2^{1-r}-1}C_r K^{r/2} N^{-1}. \tag{3.28}$$

It is that bound that we will use in the sequel.

Anyway remember that, since K is going to be of order of magnitude $O(N)$, one has $L = O(N^{r/2-1})$ when $N \longrightarrow \infty$.

Now let us compute V: V is defined by (3.24), which is bounded above by

$$\begin{aligned} &2{C_r}^2 N^{-1} \int_{(\mathbb{R}^d)^2} |w-v|^{2r} \mathrm{d}\mu(v)\mathrm{d}\mu(w) \\ &\overset{\text{Jensen}}{\leqslant} 2{C_r}^2 N^{-1} \left(\int_{(\mathbb{R}^d)^2} |w-v|^2 \mathrm{d}\mu(v)\mathrm{d}\mu(w) \right)^r \\ &= 2^{1+2r} {C_r}^2 N^{-1} \left(\frac{\widetilde{K}}{N} \right)^r \leqslant 2^{1+2r} {C_r}^2 K^r N^{-(1+r)}; \end{aligned} \tag{3.29}$$

taking into account Remark 3.8, we could even take

$$V = (2^{1-r}-1) 2^{1+2r} {C_r}^2 K^r N^{-(1+r)}. \tag{3.30}$$

Anyway remember that $V = O(N^{-1})$ when $N \longrightarrow \infty$.

3.3 Comments on the Results

☛ *All the computations in this subsection are heuristic, so we will drop lower order terms without wondering when we can do so. $C_1, C_2, \ldots$ will denote constants depending only on κ, V, L and T, whose exact expression does not interest us.*

In the right-hand side of formula (3.26) there are two terms: the first one, $\ln \mathbb{E}[\mathcal{U}(\lambda e^{-\kappa T} \times (\widehat{X}_0 - X_0))]$, merely expresses the difference between the experimental initial condition and its continuous limit. There is obviously no surprise in getting such a term, whose study is deferred to Sect. 5: for the time being just notice the presence of the factor $e^{-\kappa T}$ in front of $\widehat{X}_0 - X_0$, which means that the effect of initial fluctuations will be quite large if $\kappa < 0$, and conversely quite small if $\kappa > 0$.

The actual dynamic effect in (3.26) lies in the term $\lambda^2 e_2(\lambda e^{2\kappa_- T} L) e_1(-2\kappa T) V T$. Let us study it in the case of our Boltzmann model, according to Sect. 3.2. We have noticed that, when N becomes large, one has $L = O(N^{r/2-1})$, resp. $V = O(N^{-1})$. So let us write $L \simeq \ell N^{r/2-1}$, resp. $V \simeq \omega N^{-1}$. Then the dynamic term of (3.26) becomes

$$\lambda^2 e_2(\lambda e^{2\kappa_- T} L) e_1(-2\kappa T) V T \simeq \lambda^2 N^{-1} e_2(\lambda e^{2\kappa_- T} \ell N^{r/2-1}) e_1(-2\kappa T) \omega T. \tag{3.31}$$

The $\lambda^2 N^{-1}$ factor hints that the good order of magnitude for λ will be $\lambda = O(N^{1/2})$. So write $\lambda = y N^{1/2}$; then (3.31) becomes

$$\lambda^2 e_2(\lambda e^{2\kappa_- T} L) e_1(-2\kappa T) V T \simeq e_2(y e^{2\kappa_- T} \ell N^{(r-1)/2}) e_1(-2\kappa T) \omega y^2 T. \tag{3.32}$$

In our case $(r-1)/2 < 0$ so, if N is sufficiently large, $y e^{2\kappa_- T} \ell N^{(r-1)/2}$ is very close to zero and the $e_2(*)$ term is very close to $e_2(0) = 1/2$, finally giving

$$\lambda^2 e_2(\lambda e^{2\kappa_- T} L) e_1(-2\kappa T) V T \simeq \frac{1}{2} e_1(-2\kappa T) \omega y^2 T. \tag{3.33}$$

For a fixed T, (3.33) shows that the dynamic term in formula (3.26) is approximately $C_1 y^2$. Moreover, as we will see in Sect. 5, the static term $\ln \mathbb{E}[\mathcal{U}(\lambda e^{-\kappa T}(\widehat{X}_0 - X_0))]$ is approximately $C_2 y^2 + C_3$. In the end, one gets

$$\ln \mathbb{E}\big[\mathcal{U}\big(y N^{1/2}(\widehat{X}_t - X_t)\big)\big] \lesssim C_4 y^2 + C_3, \tag{3.34}$$

hence by Markov's inequality and Proposition 3.1-1, for all $x > 0$,

$$\mathbb{P}\Big(y N^{1/2} \|\widehat{X}_T - X_T\| \geqslant x\Big) \lesssim e^{C_4 y^2 + C_3 - x}. \tag{3.35}$$

Optimizing formula (3.35) for fixed x/y ratio, one finally finds:

$$\forall \varepsilon \geqslant 0 \quad \mathbb{P}\big(\|\widehat{X}_T - X_T\| \geqslant \varepsilon\big) \lesssim \exp\big(C_3 - C_5 N \varepsilon^2\big). \tag{3.36}$$

So Theorem 3.6 gives a Gaussian control for the fluctuations between $\widehat{X}_T$ and X_T for any fixed value of T—provided the existence of some contractivity constant κ, which for $H = \dot{H}^{-s}$ will be proved for the Maxwellian case in Sect. 4. Moreover the order of magnitude of the fluctuations we get is $N^{-1/2}$, the typical deviation size in central limit theorems. So we may say that the bounds we have got are a kind of explicit dynamic central limit bound for the Boltzmann model.

Remark 3.9 Actually the approximations we made to get (3.34) are sensible only if y is not too large, otherwise $\lambda e^{2\kappa - T}\ell N^{r/2-1} \gtrsim 1$ and then the $e_2(*)$ term in (3.31) cannot be considered as close to $1/2$. It follows that our computations are valid only for $\lambda \lesssim N^{1-r/2}/\ell$, i.e. for $y \lesssim N^{(1-r)/2}/\ell$. Tracking that constraint throughout our reasoning, it finally turns out that (3.36) is only valid for $\varepsilon \lesssim T\omega N^{-r/2}/\ell$. So our Gaussian control does not hold up to large deviations but only till some intermediate deviations.[8] Fortunately (3.36) tells us that the probability of such intermediate deviations is bounded above by something like $e^{-C_6 N^{1-r}}$, which goes very fast to 0 anyway. Moreover, even for $\varepsilon \gg \omega T N^{-r/2}/\ell$ one can still use (3.35) with $y = N^{(1-r)/2}/\ell$ and $k \simeq yN^{1/2}\varepsilon$, which gives an exponential control of the tail of the law of $\|\widehat{X}_T - X_T\|$ applicable to large deviations.

The behaviour of formula (3.33) as T becomes large depends on the sign of κ:[9]

- If $\kappa < 0$ (the worst case), then the $e_1(-2\kappa T)$ factor becomes exponentially large as soon as $T \gtrsim 1/|\kappa|$. Thus the dynamic control given by Theorem 3.6 is relevant only for moderate values of T corresponding to durations for which each particle makes only a couple of collisions. Moreover, as we noticed in the beginning of that subsection, in that case the term due to the control of initial fluctuations will become huge as T increases. Note however that qualitatively we get a Gaussian control for *any* fixed T, only the constants in that control becoming bad.
- If $\kappa = 0$ the dynamic term of (3.26) increases proportionally to T, so our bound remains good even for moderately large values of T, but ultimately becomes uninteresting.
- If $\kappa > 0$ (the best case) then $Te_1(-2\kappa T) \longrightarrow 1/2\kappa$ when $T \longrightarrow \infty$, so the right-hand side of (3.26) remains bounded uniformly in T, implying that the N-particle model approximates well its continuous limit for *any* time.[10] Note that $\kappa > 0$ is tantamount to having an exponential convergence of (1.6) to equilibrium in A, so in that case our bound rather looks like a result of convergence to "equilibrium" for the empirical measure $\widehat{\mu}_t^N$.

4 Contractivity of the Collision Kernel

4.1 Limitations Due to Our Settings

In this section we are going to look for computing constant κ in (3.5). Unfortunately it turns out that, for the choices we have made, our results are unavoidably limited, as we quickly explain in this foreword. Let me stress however that all the issues encountered may be solved by working in a trickier space than the plain $\dot{H}^{-s}$ (cf. Remark 0.1).

[8] Note however that our control (3.28) on L was very coarse: in real situations indeed the maximal relative speed between two particles is $\sim\sqrt{\ln N}$ with very large probability (think about the Maxwell distribution), so in most cases $L \sim (\ln N)^{r/2}N^{-1}$, and then we can study deviation orders $N^{-\eta}$ with η arbitrarily close to 0. However such a study needs a control on the probability that L becomes large, in other words a control on the probability of appearance of an abnormally hot particle, which would require another article.

[9] In Sect. 4.1 we will see that for $H = \dot{H}^{-s}$, κ is actually always negative. Our the discussion is relevant nevertheless, because it remains valid for other applications of abstract Theorem 3.3, therefore highlighting the interest of choosing a Hilbert space better than $\dot{H}^{-s}$.

[10] Beware: it does not mean that *one* random particle system has large probability to stay *always* close to the continuous limit—which is trivially false by ergodicity—but that at any *given* time, *most* of the particle systems will be close to the limit.

First κ can only be negative, which is the worst case (see page 1118). Why that? Well, if κ were positive, as we said previously it would imply convergence of (1.6) to a unique equilibrium for all probability measures. Yet there are several different equilibrium probability measures for the Boltzmann evolution (see formula (1.8) and below), whose differences lie in $\dot{H}^{-s}$, which is a contradiction. So κ is nonpositive. Then we could prove, using that the model is nondegenerate, that κ cannot be zero and thus is negative. To have a chance to get positive values of κ, $\dot{H}^{-s}$ should be replaced by a Hilbert space containing only signed measures η such that $\int \eta(\mathrm{d}x),\ \int x\eta(\mathrm{d}x),\ \int |x|^2\eta(\mathrm{d}x) = 0$—but which one?

Secondly, the only chance for κ to be finite is the case of Maxwellian models (remember definition below (1.5)): this is due to a bad scale invariance property for non-Maxwellian models, cf. Remark 4.4. Though the Maxwellian case is often a useful first step for theoretists, the physical models encountered in real life do not have any reason for being so! To have a chance to get results for non-Maxwellian models, $\dot{H}^{-s}$ should be replaced by some non-homogeneous space—but non-homogeneous spaces are often less tractable than homogeneous spaces and more difficult to interpret physically.

4.2 Principle to the Computation of κ

To check hypothesis (3.23), according to Remark 3.7, we can consider our Markov processes restricted to the set of probability measures, and then by positive linearity it suffices to prove (3.23) when μ is a Dirac mass:

Immediate proposition 4.1 *If, for one arbitrary (then for all) $v \in \mathbb{R}^d$, the linear operator $Q(\delta_v, \cdot) : \dot{H}^{-s} \longrightarrow \dot{H}^{-s}$ satisfies the "contractivity" property*

$$\forall f \in \dot{H}^{-s} \quad \langle Q(\delta_v, f), f\rangle \leqslant -\frac{\kappa}{2}\|f\|^2, \tag{4.1}$$

then the restriction of Q to probability measures satisfies hypothesis (3.23).

Remark 4.2 It is not hard to see that conversely the best κ possible in (3.23) is *exactly* the best κ possible in (4.1). We do not prove it as it is not essential, but it will be implicitly used in Remark 4.4.

Lemma 4.3 *Recall definition (2.5) of ϕ_s. Note $(*\phi_s)$ the convolution operator*

$$\begin{array}{rccl} *\phi_s : & \dot{H}^{-s} & \longrightarrow & L^2 \\ & f & \mapsto & f * \phi_s. \end{array} \tag{4.2}$$

Then $Q(\delta_v, \cdot) : \dot{H}^{-s} \longrightarrow \dot{H}^{-s}$ satisfies property (4.1) if and only if

$$(*\phi_s) \circ Q(\delta_v, \cdot) \circ (*\phi_s)^{-1} : L^2 \longrightarrow L^2 \tag{4.3}$$

satisfies the same property in the space $L^2(\mathbb{R}^d)$.

Proof It follows directly from the isomorphism formula (2.6). □

Remark 4.4 Now we can understand why κ cannot be finite for a non-Maxwellian model: suppose the model satisfies (1.5) with $g \neq 0$, and for $\lambda \in (0, +\infty)$ denote by I_λ the homothety transforming v into λv, then you get

$$Q\big(\delta_0, I_{\lambda\#}\mu\big) = \lambda^g I_{\lambda\#} Q\big(\delta_0, \mu\big), \tag{4.4}$$

so if $Q(\delta_0, \cdot)$ were κ-contracting for a $\kappa < 0$ it would also be $\lambda^g\kappa$-contracting for all λ, thus 0-contracting, which is impossible.

4.3 Effective Computation

Lemma 4.5 *Let $\theta \in [0, \pi]$; define the linear operator $\check{Q}_\theta$ on measures on $\mathbb{R}^d$, such that $\check{Q}_\theta(\delta_v)$ is the uniform probability measure on the $(d-2)$-dimensional sphere*[11] *of velocities v' such that $|v' - v/2| = |v|/2$ and $\widehat{v\frac{v}{2}v'} = \theta$. Then*

$$(*\phi_s) \circ \check{Q}_\theta = \big(\cos(\theta/2)\big)^s \check{Q}_\theta \circ (*\phi_s). \tag{4.5}$$

Remark 4.6 $\check{Q}_\theta(\delta_v)$ respresents the post-collisional distribution of velocity of a particle at initial velocity v which has collided with a particle at initial velocity 0, undergoing an angular deviation θ in the collision referential, the precise direction of that deviation being random.

Proof Let us give first a neat proof working when d is even. Call $\mathcal{R}_\theta$ the set of the rotations R of $\mathbb{R}^d$ satisfying $\widehat{v0(Rv)} = \theta/2$ for *all* $v \in \mathbb{R}^d$. If d is even, $\mathcal{R}_\theta$ is non-empty and has some canonical probability measure π_θ equipping it. Then we notice that

$$\check{Q}_\theta(\mu) = \int_{\mathcal{R}_\theta} [\cos(\theta/2)R]^\# \mu \, \mathrm{d}\pi_\theta(R). \tag{4.6}$$

Because of the rotational invariance of ϕ_s, for any $R \in R_\theta$,

$$(*\phi_s) \circ R^\# = R^\# \circ (*\phi_s). \tag{4.7}$$

Similarly, the scale invariance of ϕ_s makes that for any $\lambda \in (0, \infty)$,

$$(*\phi_s) \circ I_\lambda{}^\# = \lambda^s \cdot I_\lambda{}^\# \circ (*\phi_s). \tag{4.8}$$

The result then follows by applying formulas (4.7) and (4.8) to the integral (4.6).

When d is odd unfortunately I have nothing better than a calculation—which by the way also works for d even. Choose an arbitrary $v > 0$, we will prove that $(\check{Q}_\theta\delta_v) * \phi_s = \check{Q}_\theta(\delta_v * \phi_s)$, where v also denotes the point $(v, 0, \ldots, 0) \in \mathbb{R}^d$. Since these two functions are invariant by any rotation around v, we will locate a point in $\mathbb{R}^d$ merely by its first coordinate z and its distance ρ to the z axis; we will also denote $Z = \sqrt{z^2 + \rho^2}$ its distance to 0. In the following calculations $\mathbb{S}$ denotes the unit sphere in $\mathbb{R}^{d-1}$, equipped with its Lebesgue probability measure σ, and ρ also denotes the point $(\rho, 0, \ldots, 0) \in \mathbb{R}^{d-1}$; points of $\mathbb{S}$ are denoted $y = (y_0, y_1)$ with $y_0 \in \mathbb{R}$, $y_1 \in \mathbb{R}^{d-2}$. Treating $(\check{Q}_\theta\delta_v) * \phi_s$ as a function, we find:

$$\begin{aligned}
&\big((\check{Q}_\theta\delta_v) * \phi_s\big)(z, \rho) \\
&\quad = \int_{\mathbb{S}} \Big\{\big(\cos(\theta/2)v - z\big)^2 + \big(\sin\theta y_0 v/2 - \rho\big)^2 + (\sin\theta)^2|y_1|^2v^2/4\Big\}^{-(d-s)/2} \mathrm{d}\sigma(y_0, y_1) \\
&\quad = \int_{\mathbb{S}} \Big\{Z^2 + \cos(\theta/2)^2v^2 - 2\cos(\theta/2)^2vz - \sin\theta v\rho y_0\Big\}^{-(d-s)/2} \mathrm{d}\sigma(y_0, y_1).
\end{aligned} \tag{4.9}$$

[11] That sphere degenerates into a point if $\theta \in \{0, \pi\}$.

For $\check{Q}_\theta(\delta_v * \phi_s)$ it is more complicated since that case needs computing a expression of type $\check{Q}_\theta f$, f being a function. Usually that kind of computation raises no difficulty, but here the operator $\check{Q}_\theta$ has some singularity which makes it less tractable: in $\check{Q}_\theta f$, the "mass" (in the measure sense) received by the point $(z, 0, \dots, 0)$ comes only from a $(d-2)$-dimensional sphere in $\mathbb{R}^d$—more precisely the sphere of points (z, ρ), $\rho \in \mathbb{R}^{d-1}$, with $|\rho| = \tan(\theta/2)z$. That regularity problem can be overcome by an approximation technique, yielding:

$$\big(\check{Q}_\theta f\big)(z, 0, \dots, 0) = \frac{1}{\cos(\theta/2)^d} \int_{\mathbb{S}} f\big(z, [\tan(\theta/2)z]y\big) \mathrm{d}\sigma(y) \tag{4.10}$$

—that formula also allowing to compute $\check{Q}_\theta f$ at points not located on the z axis by rotational invariance.

So

$$\begin{aligned}
&\big(\check{Q}_\theta(\delta_v * \phi_s)\big)(z, \rho) \\
&\quad = \cos(\theta/2)^{-d} \cdot \int_{\mathbb{S}} \Big\{ \big(z - \tan(\theta/2)\rho y_0 - v\big)^2 + \big(\rho + \tan(\theta/2) z y_0\big)^2 \\
&\qquad + \tan(\theta/2)^2 Z^2 |y_1|^2 \Big\}^{-(d-s)/2} \mathrm{d}\sigma(y_0, y_1) \\
&\quad = \cos(\theta/2)^{-d} \int_{\mathbb{S}} \Big\{ \big(1 + \tan(\theta/2)^2\big) Z^2 - 2v\big(z - \tan(\theta/2)\rho y_0\big) + v^2 \Big\}^{-(d-s)/2} \mathrm{d}\sigma(y_0, y_1) \\
&\quad = \cos(\theta/2)^{-s} \big((\check{Q}_\theta \delta_v) * \phi_s\big)(z, \rho).
\end{aligned} \tag{4.11}$$

□

Corollary 4.7 *Let $Q_\theta = \check{Q}_\theta + \check{Q}_{\pi-\theta} - \check{Q}_0 - \check{Q}_\pi$. Then, for any $f \in \dot{H}^{-s}$,*

$$\langle Q_\theta f, f \rangle \leqslant \big[\big(\cos(\theta/2)\big)^r + \big(\sin(\theta/2)\big)^r - 1\big] \cdot \|f\|^2. \tag{4.12}$$

Proof Observe first that $\check{Q}_0$ is the identity and that $\check{Q}_\pi = 0$, so it suffices to prove that the operator norm of $\check{Q}_\theta$ in $\dot{H}^{-s}$ is bounded above by $(\cos(\theta/2))^r$. By isomorphism formula (2.6), that is also the norm of $(*\phi_s) \circ \check{Q}_\theta \circ (*\phi_s)^{-1}$ in L^2, which is $\cos(\theta/2)^s \check{Q}_\theta$ by Lemma 4.5. So we just have to bound the norm of $\check{Q}_\theta$, *regarded as an operator in* L^2, by $\cos(\theta/2)^{-d/2}$. Now we note that one can write

$$\check{Q}_\theta f = I_{\cos(\theta/2)\#} \widetilde{Q}_\theta f, \tag{4.13}$$

where $\widetilde{Q}_\theta$ is the kernel of the Markov chain on $\mathbb{R}^d$ which sends x uniformly to the $(d-2)$-dimensional sphere of points y such that $|y| = |x|$ and $\widehat{x0y} = \theta/2$. But that Markov chain has the Lebesgue measure on $\mathbb{R}^d$ as reversible equilibrium measure, so $|||\widetilde{Q}_\theta|||_{L^2} \leqslant 1$, thus $|||\check{Q}_\theta|||_{L^2} \leqslant \cos(\theta/2)^{-d/2}$, *quod erat demonstrandum.* □

Now we are ready to state the main result of this section:

Theorem 4.8 *In a Maxwellian model, calling $\overline{\gamma}$ the common value of all the measures $\overline{\gamma}_u$, the collision kernel Q, when restricted to the probability measures, satisfies hypothesis (3.5) with*

$$\kappa = \int_0^\pi \big[1 - \cos(\theta/2)^r - \sin(\theta/2)^r\big] \mathrm{d}\overline{\gamma}(\theta). \tag{4.14}$$

Proof Note that

$$Q(\delta_0, \cdot) = \frac{1}{2}\int_0^\pi Q_\theta \mathrm{d}\overline{\gamma}(\theta) \tag{4.15}$$

and apply all the previous work of this section (Lemmas 4.1, 4.3, 4.5 and 4.7). □

Example 4.9 The "Kac" model[12] is the case where the measure $\gamma_{v,w}$ always has total mass 1 and is uniform on the sphere supporting it, i.e. it is the Maxwellian model with

$$\mathrm{d}\overline{\gamma}(\theta) = \frac{\Gamma(d-1)}{2^{d-2}\Gamma((d-1)/2)^2}(\sin\theta)^{d-2}\mathrm{d}\theta. \tag{4.16}$$

By Theorem 4.8, for that model one has $-\infty < \kappa < 0$ for any $r \in (0, 1)$.

Example 4.10 The model of Maxwellian potential corresponds to particles having a repulsive force with a radially symmetric potential decreasing like $\rho^{-(2d-2)}$ as the distance ρ between two particles increases. For that model $\mathrm{d}\overline{\gamma} \sim \theta^{-3/2}\mathrm{d}\theta$ when $\theta \longrightarrow 0$ for any d (thus the measure $\overline{\gamma}$ is not finite, however it remains possible to define both the N-particle and the limit models, cf. Remark 1.1-3), so by Theorem 4.8 one also has $-\infty < \kappa < 0$ for any $r \in (0, 1)$.

5 Initial Value

In formula (3.26) given by Theorem 3.6, as we saw, besides the dynamic term there is a term due to the fluctuations of the initial empirical measure. In this section we control these fluctuations in the case of i.i.d. initial particles.

Let μ be a probability measure on $\mathbb{R}^d$ and let $r \in (0, 1)$. We assume that μ has an r-th exponential moment, i.e. that there exists some $a > 0$ such that

$$\int_{\mathbb{R}^d} e^{a|v|^r}\mathrm{d}v < \infty. \tag{5.1}$$

In the sequel we suppose a fixed.

If v is a random variable of $\mathbb{R}^d$ with law μ, then $\delta_v - \mu$ is a random variable in $\dot{H}^{-s}$, whose law will be denoted $\mathfrak{D}_\mu$: $\mathfrak{D}_\mu$ is a centered probability measure on $\dot{H}^{-s}$. I claim that $\mathfrak{D}_\mu$ has an exponential moment with parameter a, i.e.

$$\int_{\dot{H}^{-s}} e^{a\|\nu\|}\mathrm{d}\mathfrak{D}_\mu(\nu) < \infty. \tag{5.2}$$

To prove it suffices to note that

$$\|\delta_v - \mu\| \leqslant \|\delta_v - \delta_0\| + \|\delta_{v_0} - \mu\| = C_r|v - v_0|^r + \|\delta_{v_0} - \mu\|, \tag{5.3}$$

whose a-parameter exponential is integrable because of (5.1).

[12] Actually this is not exactly the Kac model of [11], but the spirit is the same.

So the law $\mathfrak{D}_\mu$ has a finite exponential moment, hence *a fortiori* a finite variance. Let us denote it by σ^2:

$$\sigma^2 = \int_{\dot{H}^{-s}} \|\nu\|^2 \mathrm{d}\mathfrak{D}_\mu(\nu). \tag{5.4}$$

Now we have all the definitions at hand to state the main result of this section:

Theorem 5.1 *Let $v_0, \ldots, v_{N-1}$ be N i.i.d. random variables on $\mathbb{R}^d$ with law μ, and denote $\widehat{\mu}^N = N^{-1}\sum_{i=0}^{N-1}\delta_{v_i}$ their empirical measure. Then there exists an explicit constant $A(\mu)$, which is easy to bound, such that for all $\lambda \leqslant aN$:*

$$\mathbb{E}\big[\mathcal{U}\big(\lambda(\widehat{\mu}^N - \widehat{\mu})\big)\big] \leqslant 2\exp\left(\frac{\lambda^2\sigma^2}{2N} + \frac{\lambda^3 A(\mu)}{N^2 a^3}\right). \tag{5.5}$$

Before proving Theorem 5.1, let us further examine formula (5.5): the term in the exponential remains bounded when $N \longrightarrow \infty$ if λ increases as $N^{1/2}$, like in (3.31). Thus, writing $\lambda = yN^{1/2}$ like in (3.34):

$$\mathbb{E}\big[\mathcal{U}\big(yN^{1/2}(\widehat{\mu}^N - \mu)\big)\big] \leqslant 2\exp\left(\frac{\sigma^2 y^2}{2} + \frac{A(\mu)y^3}{a^3}N^{-1/2}\right). \tag{5.6}$$

Though we will not use it in the sequel, note the following

Corollary 5.2 *For $S \geqslant \sigma^2$, for all $x \geqslant 0$, for all $N \geqslant N_0 := x^2/a^2S^2$:*

$$\mathbb{P}\Big(\|\widehat{\mu}^N - \mu\| \geqslant xN^{-1/2}\Big) \leqslant \exp\left(-\frac{x^2}{2S} + \ln 2 + A(\mu){N_0}^{1/2}N^{-1/2}\right). \tag{5.7}$$

Remark 5.3 (5.7) works as soon as $N \geqslant x^2/a^2S^2$, i.e. as soon as $x \leqslant aSN^{1/2}$, so that estimate is valid up to the large deviations setting.

Proof of Theorem 5.1 The principle of the proof is exactly the same as for Theorem 3.3, except that here time will be discrete.

Let $v_0, \ldots, v_{N-1}$ be N i.i.d. random variables with law μ and set $\widehat{M}_i = \sum_{j=0}^{i-1} N^{-1} \times (\delta_{v_i} - \mu)$, then $(\widehat{M}_i)_i$ is a Markov chain and a martingale, and $\widehat{M}_N$ has the same law as $\widehat{\mu}^N$. So it suffices to prove that for all $0 \leqslant i < N$,

$$\mathbb{E}\big[\mathcal{U}(\lambda\widehat{M}_{i+1})\big|\widehat{M}_i\big] \leqslant \exp\left(\frac{\lambda^2\sigma^2}{2N} + \frac{\lambda^3 A(\mu)}{a^3N^2}\right)\mathcal{U}(\lambda\widehat{M}_i). \tag{5.8}$$

To get (5.8), thanks to Lemma 3.4 it suffices to prove that

$$\int \big(e^{\lambda N^{-1}\|\nu\|} - N^{-1}\lambda N^{-1}\|\nu\|\big)\mathrm{d}\mathfrak{D}_\mu(\nu) \leqslant \exp\left(\frac{\lambda^2\sigma^2}{2N^2} + \frac{\lambda^3 A(\mu)}{a^3N^3}\right). \tag{5.9}$$

We set

$$A(\mu) = \int \big(e^{a\|\nu\|} - a\|\nu\| - 1\big)\mathrm{d}\mathfrak{D}_\mu(\nu). \tag{5.10}$$

The function $e_2(t) = (e^t - 1 - t)/t^2$ is convex on $\mathbb{R}_+$, so

$$\forall t \geqslant 0\ \forall \theta \in [0,1] \quad e^{\theta t} - \theta t - 1 \leqslant \frac{1}{2}(1-\theta)\theta^2 t^2 + \theta^3(e^t - t - 1). \tag{5.11}$$

Consequently

$$\int \left(e^{N^{-1}\lambda\|\nu\|} - N^{-1}\lambda\|\nu\| - 1\right)\mathrm{d}\mathfrak{D}_\mu(\nu) \leqslant \left(1 - \lambda a N^{-1}\right)\frac{\lambda^2\sigma^2}{2N^2} + \frac{\lambda^3 A(\mu)}{a^3 N^3}, \tag{5.12}$$

whence (5.9). □

6 Discussion

6.1 Examples of Synthetic Results

Until now in this article I have just given separate results, mainly Theorem 3.6, formulas (3.28) and (3.30), and Theorems 4.8 and 5.1. Obviously all these results are to be put together to get synthetic results on the convergence of N-particle dynamic models to their mean field limit; yet I did not do it in the previous sections.

There are several reasons why I have postponed the presentation of such synthetic results to the last section. The most obvious one is that these global results would have been quite unreadable if put in the beginning of the article. More important, the different "bricks" of results given within the core of the paper are open to improvements different for each, some of which may work for some cases but not for others, so that there may be no ideal general result.

Let us however give some examples of formulas got by piling our theorems together—proofs will not be given since they really consist in plain gluing game:

Theorem 6.1 *Let $d \geqslant 2$, $r \in (0,1)$. Let μ_0 be a probability measure on $\mathbb{R}^d$ with finite r-exponential moments for all $r < 1$. Up to translating the origin of $\mathbb{R}^d$ we can suppose that $p := \int_{\mathbb{R}^d} v\mathrm{d}\mu_0(v) = 0$; then let $k = \frac{1}{2}\int |v|^2\mathrm{d}\mu_0(v)$. Choose some $k_1 > k$ and define*

$$\kappa = 1 - \frac{\Gamma(d-1)}{2^{d-3}\Gamma(d-1)/2)^2}\int_0^\pi \sin(\theta/2)^r \sin(\theta)^{d-2}\mathrm{d}\theta,^{13} \tag{6.1}$$

$$\ell = 2^{1+r}\sqrt{2^{1-r} - 1}\,C_r k_1^{r/2}, \tag{6.2}$$

$$\omega = (2^{1-r} - 1)2^{1+2r}{C_r}^2 k_1^r, \tag{6.3}$$

$$\sigma^2 = \int_{\mathbb{R}^d} \|\delta_v - \mu_0\|^2_{\dot{H}^{-s}}\mathrm{d}\mu_0(v). \tag{6.4}$$

Let $N \geqslant 2$; let $v_0^0, \ldots, v_{N-1}^0$ be N i.i.d. random variables with law μ_0 and let $\widehat{\mu}_0^N$ be their empirical measure; denote $\widehat{K}^N = \frac{1}{2}\sum_{i=0}^{N-1}|v_i^0|^2$. Let $\widehat{\mu}_t^N$ be the empirical measure at time t of the Markov process with generator (1.1) for the "Kac" model (4.16) and initial condition $(v_0^0, \ldots, v_{N-1}^0)$. Let $(\mu_t)_{t\geqslant 0}$ be the deterministic evolution (1.6) for the same model with initial value μ_0.

[13] Warning, κ is negative.

Then for any $a > 0$, there is a (easily bounded) constant $A(a, \mu)$ such that, for any $T > 0$, as soon as $\lambda \leqslant a e^{-|\kappa|T} N$:

$$\begin{aligned}&\ln \mathbb{E}\big[\mathbb{1}_{\widehat{K}^N \leqslant N k_1} \mathcal{U}\big(\lambda(\widehat{\mu}_t^N - \mu_t)\big)\big] \\ &\quad \leqslant \ln 2 + \frac{e^{2|\kappa|T}\lambda^2\sigma^2}{2N} + \frac{e^{2|\kappa|T}\lambda^3 A(a,\mu)}{N^2 a^3} + \frac{\lambda^2 \omega T}{N} e_1(2|\kappa|T) e_2\big(\lambda e^{2|\kappa|T} \ell N^{r/2-1}\big).\end{aligned} \tag{6.5}$$

Corollary 6.2 *For the same model, for any $y \geqslant 0$:*

$$\overline{\lim_{N \longrightarrow \infty}} \ln \mathbb{E}\big[\mathbb{1}_{\widehat{K}^N \leqslant N k_1} \mathcal{U}\big(y N^{1/2}(\widehat{\mu}_t^N - \mu_t)\big)\big] \leqslant \ln 2 + e^{2|\kappa|T} \frac{\sigma^2 y^2}{2} + e_1(2|\kappa|T) \frac{\omega T y^2}{2}. \tag{6.6}$$

Corollary 6.3 *Still for the same model, for any $x \geqslant 0$:*

$$\overline{\lim_{N \longrightarrow \infty}} \mathbb{P}\big(\|\widehat{\mu}_t^N - \mu_t\| \geqslant x N^{-1/2}\big) \leqslant 2 \exp\left(\frac{-x^2}{2\big[e^{2|\kappa|T}\sigma^2 + e_1(2|\kappa|T)\omega T\big]}\right). \tag{6.7}$$

Remark 6.4 As (6.7) is true for any value of k_1, we can make k_1 approach k in it, which allows to replace ω by $\omega_0 := (2^{1-r} - 1) 2^{1+2r} {C_r}^2 k^r$.

6.2 Optimality

Theorem 6.1 essentially gives a convergence to the continuous limit at rate $N^{-1/2}$ with Gaussian control. Qualitatively it is the best result one could hope for, because it is the same way of convergence as for central limit theorems. Quantitatively however, is the parameter in the Gaussian bound optimal?

Here we will look at what happens for Theorem 5.1 (Theorem 3.3 exhibits the same behaviour, but it is harder to see). Through Corollary 5.2, Theorem 5.1 gives some Gaussian bound in an infinite-dimensional frame. Yet its proof, whose main ingredient is the use of the utility function $\mathcal{U}$, would work as well in a finite-dimensional setting. So let us imagine that we replace $\dot{H}^{-s}$ by $\mathbb{R}^d$ and $\mathfrak{D}_\mu$ by the centered normal law with variance $\mathbf{I}_d$, denoted by $\mathcal{N}$; then σ^2 becomes $\int_{\mathbb{R}^d} |x|^2 \mathrm{d}\mathcal{N}(x) = d$. In that case $\widehat{\mu}^N$ turns into a random variable X^N on $\mathbb{R}^d$ which is centered normal with variance $N^{-1}\mathbf{I}_d$, and we get:

$$\mathbb{P}(|X^N| \geqslant x N^{-1/2}) \leqslant \exp\left(-\frac{x^2}{2d} + \ln 2 + A {N_0}^{1/2} N^{-1/2}\right) \tag{6.8}$$

for some A and N_0 not depending on N, so making $N \longrightarrow \infty$:

$$\mathcal{N}(|X| \geqslant x) \leqslant 2 e^{-x^2/2d}, \tag{6.9}$$

whereas the exact result is

$$\mathcal{N}(|X| \geqslant x) = \frac{2^{1-d/2}}{\Gamma(d/2)} \int_x^\infty y^{d-1} e^{-y^2/2} \mathrm{d}y \underset{x \longrightarrow \infty}{\approx} e^{-x^2/2}, \tag{6.10}$$

where "$\approx$" means "having equivalent logarithms".

So for $d > 1$ the parameter in the Gaussian bound is underestimeted by a factor d. Why that? Well, the proof of Theorem 5.1 uses the bound on the curvature of the utility function

$\mathcal{U}$ given by Proposition 3.1-6. But as soon as x is reasonably large, the Hessian of $\mathcal{U}$ at x is much more curved in one direction than in all the other ones, so that formula (3.15) in Lemma 3.4 becomes strongly suboptimal since the factor $\|y\|^2$ in it should morally be replaced by the sole component of the variance of y in the direction along which $\nabla^2\mathcal{U}(x)$ is most curved.

So the techniques involving $\mathcal{U}$ are poor as soon as the dimension in which the random phenomena occur becomes large. For our particle models we work in $\dot{H}^{-s}$, whose dimension is… infinite! Does that mean that our results are "infinitely bad"? Actually not, because each increment of the martingale $\widehat{M}$ (see the proof of Theorem 5.1) is determined by the value of one v_i, so the law of this increment may be seen as a probability on $\mathbb{R}^d$. More precisely, the support of $\mathfrak{D}_\mu$ is isometric to $\mathbb{R}^d$ equipped with the distance $|\cdot - \cdot|^r$, whose Hausdorff dimension is d/r, so that the "effective" dimension of $\dot{H}^{-s}$ in our theorem is d/r.

As a consequence we had better not choose r too close to 0. On the other hand, the bigger s is, the more regular the test functions in the definition of $\|\cdot\|_{\dot{H}^{-s}}$ (see Proposition 2.3) are, so the less small-scale details $\|\cdot\|_{\dot{H}^{-s}}$ catches.[14] So it should be advised to take medium values of r, e.g. $r = 1/2$.

6.3 A Numerical Computation

One important side of our work is that it gives non-asymptotic results. The idea behind it is that, to understand Boltzmann's evolution, we will not actually look at $N \longrightarrow \infty$, but rather take some *fixed* large N and say that the behaviour of the N-particle system for that N is very close to the limit evolution with very large probability. In particular, think about the case of numerical simulation: we cannot afford dealing with 10^{24} particles on our computers!

Here I will compute numerical values for the following case: the collision kernel is the one of "Kac" model for $d = 3$ and we take $\mu_0 = \frac{1}{2}(\delta_{-1} + \delta_1)$. Physically speaking, it means that we crash together two same-sized sets of frozen particles with relative speed 2. Then the collisions between particles of differents sets will tend to scatter the distribution of velocities of the particles, which will morally converge to the law (1.8) with $k = 1/2$ in a few units of time—this is the behaviour of Boltzmann's equation (1.6) indeed. The question is, which N shall we choose to be almost certain that the evolution of the particle system will be fairly close to (1.6)?

Say we take $r = 1/2$ and we want to have $\|\widehat{\mu}_T^N - \mu_T\|_{\dot{H}^{-s}}$ greater than $\varepsilon = 10^{-2}$ with probability less than $q = 10^{-1}$ for $T = 3$. As in our case $\widehat{K}^N \leqslant Nk$ almost surely, we take $k_1 = k = 1/2$. Then one computes the following numerical values, which are all rounded above:

$$-\kappa \simeq 0.600; \tag{6.11}$$

$$\ell \simeq 0.432; \tag{6.12}$$

$$\omega \simeq 0.0933; \tag{6.13}$$

$$\sigma^2 \simeq 0.0398. \tag{6.14}$$

We choose arbitrarily $a = 1$; then (5.10) gives $A(a, \mu) \simeq 0.0213$. We have to take $\lambda \gtrsim |\ln q|/\varepsilon$, so let us put $\lambda = 500$. For $N = 8 \cdot 10^5$ we find by (6.5):

$$\ln \mathbb{E}\big[\mathcal{U}\big(\lambda(\widehat{\mu}_t^N - \mu_t)\big)\big] < 2.692, \tag{6.15}$$

[14] Remember however that homogeneous Sobolev spaces have no inclusion relations.

thus

$$\mathbb{P}\big[\|\widehat{\mu}_t^N - \mu_t\| \geqslant 10^{-2}\big] < 10^{-1} \tag{6.16}$$

by Markov's inequality.

So with a discrete system of $8 \cdot 10^5$ particles one will much probably find a quite good approximation of Boltzmann mean field limit by running the particle system 3 units of time. Now simulating $8 \cdot 10^5$ particles is easy for today's computers, which shows that our bounds can actually be useful in practice. However there is little doubt that the true speed of convergence is much faster than what our computations suggest.

Remark 6.5 Here we have bypassed the problem of the $\mathbb{1}_{\widehat{K}^N \leqslant Nk_1}$ factor in (6.5) by a specific argument. How can we do for it in the general case? Well, merely note that, as soon as one wants to have a result in terms of probability, they will just have to add $\mathbb{P}(\widehat{K}^N > Nk_1)$ to the probability they get forgetting the indicator. But the event $\{\widehat{K}^N > Nk_1\}$ is a large deviations event, so as soon as μ_0 has some square-exponential moment its probability will decrease exponentially with N and thus cause no problem actually.

6.4 Comparison to Older Results

The usual method to tackle mean field limit problems relies on the concept of *propagation of chaos* devised by Kac [11]. Briefly speaking, propagation of chaos consists in studying the law of the N-particle assembly $(v_0, \ldots, v_{N-1})$ through its finite-dimensional marginals, i.e. in studying the laws of the $(v_0, \ldots, v_{k-1})$'s for k finite. One says that there is "chaos" when these finite-dimensional laws tend to product laws $u^{\otimes k}$ when $N \longrightarrow \infty$, and the goal is to prove that, if there is chaos at time 0, then this chaos "propagates" for all t.

Propagation of chaos has been proved by Sznitman [20] for spatially homogeneous Boltzmann models, and more recently by Graham and Méléard [10] for the more general *Povzner equation*. Actually, proving propagation of chaos is the same as proving the convergence of empirical distributions of the N-particle system to some deterministic distribution, but propagation of chaos emphasizes the individual behaviour of each particle, which is described by the *nonlinear particle* [21]. On the other hand, the finite-dimensional setting of that method makes that the quantitative results got thanks to it (for instance in [9, 13]) do not translate very well when one tries to control the difference between $\widehat{\mu}^N$ and μ in some metric space.

My paper was motivated by the reading of [2], in which Bolley, Guillin and Villani tackle some mean field limit problems in a quantitative way by working with W_1 Wasserstein distances for the empirical measures. They get an explicit control on the large deviations of the difference between the empirical measure of the N-particle system and its theoretical limit for positive times. Yet there are two annoying shortcomings in their work:

- First, it seems to be limited to McKean–Vlasov models, that is, systems where the interactions between particles are due to forces rather than collisions. The proofs of [2] indeed fundamentally relie on a coupling technique (popularized by Sznitman [20]), in which one defines a coupling between the real assembly of particles and a virtual assembly of N independent nonlinear particles. Such a technique has little relevance when one deals with collisions, because these events imply *two* particles at the *same* moment each time they occur, so there is no natural way of coupling with independent particles.
- Secondly, the results of [2] are good for large deviations, but the control they give for medium deviations is far too poor to get, as we would wish, some $N^{-1/2}$ convergence rate. As we tell in Appendix, that is actually an intrinsic shortcoming of W_1 distances.

After writing my article I discovered some other papers sharing certain features with mine:

- The microscopic model exposed in this paper is an example of Bird's direct simulation Monte-Carlo method [1], whose convergence for the Boltzmann equation was proved in [16, 24], with explicit estimates on L^1 distance of the marginals.
- The first having looked at the empirical distribution of the particles in Hilbert spaces to bypass the coupling problems were Fernandez and Méléard [8], who analysed the fluctuations of the particle distribution when $N \longrightarrow \infty$ in a spirit close to uniform central limit theory [7].
- Aldéric Joulin pointed out to me that my convergence theorems could be interpreted as results of concentration of measure for a Markov process with positive curvature, according to the geometric notions for Markov chains introduced by Ollivier in [15]. Hypotheses 1, 2 and 3 of Theorem 3.6 indeed correspond resp. to the hypotheses about the discrete Ricci curvature κ, the coarse diffusion constant $\sigma(x)$ and the granularity σ_∞ in Theorem 33 of [15].

6.5 Uniform in Time Bounds

The results we have given work for some *fixed* T, i.e. they control $\|\widehat{\mu}_T^N - \mu_T\|$. It may be more natural to control $\sup_{t\in[0,T]} \|\widehat{\mu}_t^N - \mu_t\|$, i.e. to say that the system is *always* close to the Boltzmann mean field limit between times 0 and T, as [2] does for McKean–Vlasov models. We do not do it here, but note that, as we have used martingale techniques, getting results valid for all $t \in [0, T]$ could be easily achieved from the previous work by using stopping times. Actually for $\kappa \leqslant 0$ it would turn out that uniform in time results are not much different from fixed time results (which is quite logical because then the control on $\|\widehat{\mu}_t^N - \mu_t\|$ is worst for $t = T$). For $\kappa > 0$ yet, when T is large the maximum of the difference between $\widehat{\mu}_t^N$ and μ_t is much less well controlled than its terminal value, as we already noticed in footnote 10 at p. 1118.

Appendix: Why Wasserstein Distances Cannot Yield $N^{-1/2}$ Convergence

This appendix aims at explaining quickly why, in quite general situations, the W_1 distance *cannot* yield a $N^{-1/2}$ rate of convergence for the empirical distribution of an assembly of N particles to its continuous limit. As it is not the main matter of this article, I will remain at a heuristic level.

My explanation relies on the transportation interpretation of W_1 distances. Recall that a *coupling* between two finite measures of same mass μ and ν on respective spaces X and Y is a measure π on $X \times Y$ whose marginals are resp. μ and ν, i.e. s.t. $\pi(A \times Y) = \mu(A)$ for all measurable $A \subset X$, resp. $\pi(X \times B) = \nu(B)$. π is also called a *transportation plan* because it describes a way to transport a mass distributed according to μ into the mass distribution ν. The set of couplings between μ and ν is always non-void; we denote it $\Pi(\mu, \nu)$. When $X = Y = \mathbb{R}^d$, for $\pi \in \Pi(\mu, \nu)$ we define the *transportation cost*

$$I[\pi] = \int_{(\mathbb{R}^d)^2} |y - x| \mathrm{d}\pi(x, y), \tag{A.1}$$

which represents the total effort you have to put in to transform μ into ν following the plan π. Then the *optimal transportation cost* is merely

$$W_1(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} c(\pi). \tag{A.2}$$

It is a deep result due to Kantorovitch [12] that definitions (2.1) and (A.2) actually coincide. For more details on all that, see [23, Sect. 1].

Now, consider a N-particle system whose empirical measure $\widehat{\mu}^N$ is expected to converge to some density measure μ, say the Lebesgue measure on $[0, 1]^d$, and look at the Wasserstein distance $W_1(\mu, \widehat{\mu}^N)$. Let π be a coupling between $\widehat{\mu}^N$ and μ. Write $\widehat{\mu}^N = N^{-1} \sum_{i=1}^N \delta_{v_i}$, and call $A_i \subset \mathbb{R}^d$ the image of $N^{-1}\delta_{v_i}$ by the transportation plan π. A_i has Lebesgue measure N^{-1}, so its observable diameter, which by isoperimetry is minimal when A_i is a ball, is at least $\sim N^{-1/d}$, therefore $\int_{A_i} |v - v_i| \mathrm{d}\mu(v) \gtrsim N^{-1} \cdot N^{-1/d}$. Thus the total transportation cost between μ and $\widehat{\mu}^N$ is $I[\pi] = \sum_i \int_{A_i} |v - v_i| \mathrm{d}\mu(v) \gtrsim N^{-1/d}$, and since that is true for any transportation plan, in the end $W_1(\mu, \widehat{\mu}^N) \gtrsim N^{-1/d}$. But that is always true, *however cleverly you might choose the v_i's* (in other words, the phenomenon we describe is not due to *fluctuations* but to *discretization*), so for $d > 2$,[15] it is hopeless getting an $N^{-1/2}$ convergence rate of $\widehat{\mu}^N$ to μ for the Wasserstein distance.

References

1. Bird, G.A.: Direct simulation and the Boltzmann equation. Phys. Fluids **13**(10), 2676–2681 (1970)
2. Bolley, F., Guillin, A., Villani, C.: Quantitative concentration inequalities for empirical measures on non-compact spaces. Probab. Theory Relat. Fields **137**, 541–593 (2007)
3. Boltzmann, L.: Weitere Studien über das Wärmegleichgenicht unter Gasmoläkuler. Sitzungsber. Akad. Wiss. **66**, 275–370 (1872). English translation: Brush, S.: Further studies on the thermal equilibrium of gas molecules. Kinetic Theory **2**, 88–175 (1966)
4. Carleman, T.: Sur la théorie de l'équation intégrodifférentielle de Boltzmann. Acta Math. **60**(1), 91–146 (1933)
5. Carlen, E.A., Gabetta, E., Toscani, G.: Propagation of smoothness and the rate of exponential convergence to equilibrium for a spatially homogeneous Maxwellian gas. Commun. Math. Phys. **199**(3), 521–546 (1999)
6. Desvillettes, L.: Boltzmann's kernel and the spatially homogeneous Boltzmann equation. Riv. Mat. Univ. Parma **6**(4), 1–22 (2001). Special issue
7. Dudley, R.M.: Uniform Central Limit Theorems. Cambridge Studies in Advanced Mathematics, vol. 63. Cambridge University Press, Cambridge (1999)
8. Fernandez, B., Méléard, S.: A Hilbertian approach for fluctuations on the McKean-Vlasov model. Stoch. Process. Their Appl. **71**(1), 33–53 (1997)
9. Fournier, N., Méléard, S.: Monte-Carlo approximations and fluctuations for 2D Boltzmann equations without cutoff. Markov Process. Relat. Fields **7**(1), 159–191 (2001). Inhomogeneous random systems (Cergy-Pontoise, 2000)
10. Graham, C., Méléard, S.: Stochastic particle approximations for generalized Boltzmann models and convergence estimates. Ann. Probab. **25**(1), 115–132 (1997)
11. Kac, M.: Foundations of kinetic theory. In: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955. Berkeley and Los Angeles, vol. III, pp. 171–197. University of California Press, Berkeley (1956)
12. Kantorovitch, L.: On the translocation of masses. C.R. (Dokl.) Acad. Sci. URSS (N.S.) **37**, 199–201 (1942)
13. Malrieu, F.: Logarithmic Sobolev inequalities for some nonlinear PDE's. Stoch. Process. Their Appl. **95**(1), 109–132 (2001)
14. Maxwell, J.: On the dynamical theory of gases. Philos. Trans. R. Soc. Lond. **157**, 49–88 (1867)
15. Ollivier, Y.: Ricci curvature of Markov chains on metric spaces. J. Funct. Anal. **256**(3), 810–864 (2009)
16. Pulvirenti, M., Wagner, W., Zavelani Rossi, M.B.: Convergence of particle schemes for the Boltzmann equation. Eur. J. Mech. B Fluids **13**(3), 339–351 (1994)

[15] For $d = 1$, Wasserstein distances *do* typically yield $N^{-1/2}$ fluctuations, but that case is physically trivial when you study Boltzmann gases. For the critical dimension $d = 2$ (for which the phenomena of discretization and fluctuations have the same order of magnitude $N^{-1/2}$), it turns out that the typical rate of convergence of empirical measures is $N^{-1/2} \ln N$, so there is in fact no central limit theorem either.

17. Revuz, D., Yor, M., Continuous Martingales and Brownian Motion. Grundlehren der mathematischen Wissenschaften, vol. 293. Springer, Berlin (1999)
18. Schwarz, L.: Méthodes Mathématiques pour les Sciences Physiques. Hermann, Paris (1997)
19. Spohn, H.: Large Scale Dynamics of Interacting Particles. Texts and Monographs in Physics. Springer, Berlin (1991)
20. Sznitman, A.-S., Équations de type de Boltzmann, spatialement homogènes. Z. Wahrscheinlichkeitstheor. Verw. Geb. **66**(4), 559–592 (1984)
21. Sznitman, A.-S.: Topics in propagation of chaos. In: École d'Été de Probabilités de Saint-Flour XIX—1989. Lecture Notes in Math., vol. 1464, pp. 165–251. Springer, Berlin (1991)
22. Villani, C.: A review of mathematical topics in collisional kinetic theory. In: Handbook of Mathematical Fluid Dynamics, vol. I, pp. 71–305. North-Holland, Amsterdam (2002)
23. Villani, C.: Topics in Optimal Transportation. Graduate Studies in Mathematics, vol. 58. Am. Math. Soc., Providence (2003)
24. Wagner, W.: A convergence proof for Bird's direct simulation Monte Carlo method for the Boltzmann equation. J. Stat. Phys. **66**(3–4), 1011–1044 (1992)