

(Ir)reversibility and entropy

Cédric VILLANI
University of Lyon
& Institut Henri Poincaré
11 rue Pierre et Marie Curie
75231 Paris Cedex 05, FRANCE

La cosa più meravigliosa è la felicità del momento

L. Ferré

Time's arrow is part of our sense experience and we experience it every day : broken mirrors do not come back together, human beings don't rejuvenate, rings grow unceasingly in tree trunks, we have memories of past events and not of future events. In sum, time always flows in the same direction ! Nonetheless, the fundamental laws of classical physics don't favor any time direction and conform to a rigorous symmetry between past and future. It is possible, as discussed in the article by T. Damour in this same volume, that irreversibility is inscribed in other physical laws, for example on the side of general relativity or quantum mechanics. Since Boltzmann, **statistical physics** has advanced another explanation : time's arrow translates a constant flow of less probable events toward more probable events. Before continuing with this interpretation, which constitutes the guiding principle of the whole exposition, I note that the flow of time is not necessarily based on a single explanation.

At first glance, Boltzmann's suggestion seems preposterous : it isn't because an event is *probable* that it is actually achieved, for time's arrow seems inexorable and seems not to tolerate any exception. The response to this objection lies in a catchphrase : **separation of scales**. If the fundamental laws of physics are exercised on the microscopic, particulate (atoms, molecules,...) level, phenomena that we can sense or measure involve a considerable number of particles. The effect of this number is even greater when it enters into combinatoric computations : if N , the number of atoms participating in an experiment, is of order 10^{10} , this is already considerable, but $N!$ or 2^N are supernaturally large, invincible numbers.

The innumerable debates between physicists that have been pursued for more than a century, and that are still pursued today, give witness to the subtlety and depth of Maxwell's and Boltzmann's arguments, banners of a small scientific revolution that was accomplished in the 1860's and 1870's, and which saw the birth of the fundamentals of the modern kinetic theory of gases, the universal concept of statistical entropy and the notion of macroscopic irreversibility. In truth, the arguments are so subtle that Maxwell and Boltzmann themselves sometimes went astray, hesitating on certain interpretations, alternating naive errors with profound concepts ; the greatest scientists at the end of the nineteenth century, e.g. Poincaré and Lord Kelvin, were not to be left behind. We find an overview of these delays in the book by Damour already mentioned ; for my part, I am content to present a "decanted" version of Boltzmann's theory. At the end of the text we will evoke the way in which Landau shattered Boltzmann's paradigm, discovering an apparent

irreversibility where there seemed not to be any and opening up a new mine of mathematical problems.

In retracing the history of the statistical interpretation of time's arrow, we will have occasion to make a voyage to the heart of profound problems that have agitated mathematicians and physicists for more than a century.

The notations used in this exposition are generally classical; I denote $\mathbb{N} = \{1, 2, 3, \dots\}$ and $\log =$ natural logarithm.

1 Newton's inaccessible realm

We will adopt here a purely classical description of our physical universe, in accordance with the laws enacted by Newton : the ambient space is Euclidean, time is absolute and acceleration is equal to the product of the mass by the resultant of the forces.

In the case of the description of a gas, these hypotheses are questionable : according to E.G.D. Cohen, the quantum fluctuations aren't negligible on the mesoscopic level. The probabilistic nature of quantum mechanics is still debated ; we nevertheless accept that the resulting increased uncertainty due to taking these uncertainties into account can but arrange our affairs, at least qualitatively, and we thus concentrate on the classical and deterministic models, "à la Newton".

1.1 The solid sphere model

In order to fix the ideas, we consider a system of ideal spherical particles bouncing off one another : let there be N particles in a box Λ . We let $X_i(t)$ denote the position at time t of the center of the i -th particle. The rules of motion are stated as follows :

- We suppose that initially the particles are well separated ($i \neq j \implies |X_i - X_j| > 2r$) and separated from the walls ($d(X_i, \partial\Lambda) > r$ for all i).
- While these separation conditions are satisfied, the movement is uniformly rectilinear : $\ddot{X}_i(t) = 0$ for each i , where we denote $\ddot{X} = d^2X/dt^2$, the acceleration of X .
- When two particles meet, their velocities change abruptly according to Descartes' laws : if $|X_i(t) - X_j(t)| = 2r$, then

$$\begin{cases} \dot{X}_i(t^+) = \dot{X}_i(t^-) - 2 \left\langle \dot{X}_i(t^-) - \dot{X}_j(t^-), n_{ij} \right\rangle n_{ij}, \\ \dot{X}_j(t^+) = \dot{X}_j(t^-) - 2 \left\langle \dot{X}_j(t^-) - \dot{X}_i(t^-), n_{ji} \right\rangle n_{ji}, \end{cases}$$

where $n_{ij} = (X_i - X_j)/|X_i - X_j|$ denotes the unit vector joining the centers of the colliding balls.

- When a particle encounters the boundary, its velocity changes also : if $|X_i - x| = r$ with $x \in \partial\Lambda$, then

$$\dot{X}_i(t^+) = \dot{X}_i(t^-) - 2 \left\langle \dot{X}_i(t^-), n(x) \right\rangle n(x),$$

where $n(x)$ is the exterior normal to Λ at x , supposed well defined.

These rules are not sufficient for completely determining the dynamics : we can't exclude *a priori* the possibility of triple collisions, simultaneous collisions between particles and the boundary, or again an infinity of collisions occurring in a finite time. However, such events are of probability zero if the initial conditions are drawn at random with respect to Lebesgue measure (or Liouville measure) in phase space [40, appendix 4.A]; we thus neglect these eventualities. The dynamic thus defined, as simple as it may be, can then be considered as a caricature of our complex universe if the number N of particles is very large. Studied for more than a century, this caricature has still not yielded all its secrets ; many still remain.

1.2 Other Newtonian models

Beginning with the emblematic model of hard spheres, we can define a certain number of more or less complex variants :

- replace dimension 3 by an arbitrary dimension $d \geq 2$ (dimension 1 is likely pathological) ;
- replace the boundary condition (elastic rebound) by a more complex law [40, chapter 8] ;
- or, instead, eliminate the boundaries, always delicate, by posing a system in the whole space \mathbb{R}^d (but we may then add that the number of particles must then be infinite so as keep a nonzero global mean density) or in a torus of side L , $\mathbb{T}_L^d = \mathbb{R}^d / (L\mathbb{Z}^d)$, which will be my choice of preference in the sequel ;
- replace the contact interaction of hard spheres by another interaction between point particles, e.g. associated with an interaction potential between two bodies : $\phi(x - y) =$ potential exerted at point x by a material point situated at y .

Among the notable interaction potentials in dimension 3 we mention (within a multiplicative constant) :

- the **Coulomb** potential : $\phi(x - y) = 1/|x - y|$;
- the **Newtonian** potential : $\phi(x - y) = -1/|x - y|$;
- the **Maxwellian** potential : $\phi(x - y) = 1/|x - y|^4$.

The Maxwellian interaction was artificially introduced by Maxwell and Boltzmann in the context of the statistical study of gases ; it leads to important simplifications in certain formulas. There exists a taxonomy of other potentials (Lennard-Jones, Manev...). The hard spheres correspond to the limiting case of a potential that equals 0 for $|x - y| > r$ and $+\infty$ for $|x - y| < 2r$.

Suppose, more generally, that the interaction takes place on a scale of order r and with an intensity a . We end up with a **system of point particles with interaction potential**

$$\ddot{X}_i(t) = -a \sum_{j \neq i} \nabla \phi \left(\frac{X_i - X_j}{r} \right), \quad (1)$$

for each $i \in \{1, \dots, N\}$; we thus suppose that $X_i \in \mathbb{T}_L^d$. Here again, the dynamic is well defined except for a set of exceptional initial conditions and it is associated

with a **Newtonian flow** \mathcal{N}_t , which associates the position at time $s + t$ with the configuration at time s ($t \in \mathbb{R}$ can be positive or negative).

1.3 Functions of distribution

Even if the Newtonian model (1) is accepted, it remains *inaccessible* to us : first because we can't perceive the individual particles (too small), and because their number N is large. By well designed experiments, we can measure the pressure exerted on a small surface, the temperature about a point, the mean density, etc. None of these quantities is expressed directly in terms of the X_i , but rather in terms of averages

$$\frac{1}{N} \sum_i \chi(X_i, \dot{X}_i), \quad (2)$$

where χ is a scalar function.

It may seem an idle distinction : in concentrating χ near the particle i , we retrieve the missing information. But quite clearly this is impossible : in practice χ is of *macroscopic* variation, e.g. of the order of the size of the box. Besides, the information contained in the averages (2) doesn't distinguish particles, so that we have to replace the vector of the (X_i, \dot{X}_i) by the **empirical measure**

$$\hat{\mu}_t^N = \frac{1}{N} \sum_{i=1}^N \delta_{(X_i(t), \dot{X}_i(t))}. \quad (3)$$

The terminology “empirical” is well chosen : it's the measure that is observed by means (without intending a pun) of measures.

To resume : our knowledge of the particle system is achieved only through the behavior of the empirical measure in a *weak topology* that models the macroscopic limit of our experiments — laboratory experiments as well as sensory perceptions.

Frequently, on our own scale, the empirical measure appears continuous :

$$\hat{\mu}_t^N(dx dv) \simeq f(t, x, v) dx dv.$$

We often use the notation $f(t, \cdot) = f_t$. The density f is the **kinetic distribution** of the gas. The study of this distribution constitutes the kinetic theory of gases ; the founder of this science is undoubtedly D. Bernoulli (around 1738), and the more famous contributors to it are Maxwell and Boltzmann. A brief history of this kinetic theory can be found in [40, chapter 1] and in the references included there.

We continue with the study of the Newtonian system. We can imagine that certain experiments allow for simultaneous measurement of the parameters of various particles, thus giving access to correlations between particles. This leads us to define, for example,

$$\hat{\mu}_t^{2;N}(dx_1 dv_1 dx_2 dv_2) = \frac{1}{N(N-1)} \sum_{i \neq j} \delta_{(X_{i_1}(t), \dot{X}_{i_2}(t), X_{i_2}(t), \dot{X}_{i_2}(t))},$$

or more generally

$$\hat{\mu}_t^{k;N}(dx_1 dv_1 \dots dx_k dv_k) = \frac{(N-k-1)!}{N!} \sum_{(i_1, \dots, i_k) \text{ distinct}} \delta_{(X_{i_1}(t), \dot{X}_{i_1}(t), \dots, X_{i_k}(t), \dot{X}_{i_k}(t))}.$$

The corresponding approximations are **distribution functions in k particles** :

$$\hat{\mu}_t^{k;N}(dx_1 dv_1 \dots dx_k dv_k) \simeq f^{(k)}(t, x_1, v_1, \dots, x_k, v_k).$$

Evidently, by continuing up until $k = N$, we find a measure $\hat{\mu}^{N;N}(dx_1 \dots dv_N)$ concentrated at the vector of particle positions and velocities (the mean over all permutations of the particles). But in practice we never go to $k = N$: k remains very small (going to 3 would already be a feat), whereas N is huge.

1.4 Microscopic randomness

In spite of the determinism of the Newtonian model, hypotheses of a probabilistic nature on the initial data have already been made, by supposing that they aren't configured to end up in some unusual catastrophe such as a triple collision. We can now generalize this approach by considering a probability distribution on the set of initial positions and velocities :

$$\mu_0^N(dx_1 dv_1 \dots dx_N dv_N),$$

which is called a **microscopic probability measure**. In the sequel we will use the abbreviated notation

$$dx^N dv^N := dx_1 dv_1 \dots dx_N dv_N.$$

It is natural to choose μ_0^N symmetric, i.e. invariant under coordinate permutations. The data μ_0^N replace the measure $\hat{\mu}_0^{N;N}$ and generalize it, giving rise to a flow of measures, obtained by the action of the flow :

$$\mu_t^N = (\mathcal{N}_t)_\# \mu_0^N,$$

and the marginals

$$\mu_t^{k;N} = \int_{(x_1, v_1, \dots, x_k, v_k)} \mu_t^N.$$

If the sense of the empirical measure is transparent (it's the "true" particle density), that of the microscopic probability measure is less evident. We imagine that the initial state has been prepared by a great combination of circumstances about which we know little : we can only make suppositions and guesses. Thus μ_0^N is a probability measure on the set of possible initial configurations. A physical statement involving μ_0^N will, however, scarcely make sense if we use the precise form of this distribution (we can't verify it, since we don't observe μ_0^N); but it will make good sense if a μ_0^N -almost certain property is stated, or indeed with μ_0^N -probability of 0.99 or more.

Likewise, the form of $\mu_t^{1:N}$ has scarcely any physical meaning. But if there is a phenomenon of concentration of measure due to the hugeness of N , then it may be hoped that

$$\mu_0^N [\text{dist} (\hat{\mu}_t^N, f_t(x, v) dx dv) \geq r] \leq \alpha(N, r),$$

where dist is a well chosen distance on the space of measures and $\alpha(N, r) \rightarrow 0$ when $r \rightarrow \infty$ even faster than the magnitude of N (for example $\alpha(N, r) = e^{-cNr}$). We will then have

$$\begin{aligned} \text{dist} (\mu_t^{1:N}, f_t(x, v) dx dv) &= \text{dist} \left(\int \hat{\mu}_t^N d\mu_t^N, f_t(x, v) dx dv \right) \\ &\leq \int \text{dist} (\hat{\mu}_t^N, f_t dx dv) \\ &\leq \int_0^\infty \alpha(N, r) dr =: \eta(N). \end{aligned}$$

If $\eta(N) \rightarrow 0$ when $N \rightarrow \infty$ it follows that, with very high probability, $\mu_t^{1:N}$ is an excellent approximation to $f(t, x, v) dx dv$, which itself is a good approximation to $\hat{\mu}_t^N$.

1.5 Micromegas

In this section we introduce two very different statistical descriptions : the macroscopic description $f(t, x, v) dx dv$ and the microscopic probabilities $\mu_t^N(dx^N dv^N)$. Of course, the quantity of information contained in μ^N is considerably more important than that contained in the macroscopic distribution : the latter informs us about the state of a typical particle, whereas a draw following the distribution μ_t^N informs us about the state of *all* particles. Considering that if we have 10^{20} degrees of freedom, we will have to integrate 99999999999999999999 of them. For handling such vertiginous dimensions, we will require a fundamental concept : entropy.

2 The entropic world

The concept and the name entropy were introduced by Clausius in 1865 as part of the theory — then under construction — of thermodynamics. A few years later Boltzmann revolutionized the concept by giving it a statistical interpretation based on atomic theory. In addition to this section, the reader can consult e.g. Balian [9, 10] about the notion of entropy in physical statistics.

2.1 Boltzmann's formula

Let a physical system be given, which we suppose is completely described by its microscopic state $z \in \mathcal{Z}$. Experimentally we only gain access to a partial description of that state, say $\pi(z) \in \mathcal{Y}$, where \mathcal{Y} is a space of macroscopic states. I won't give precise hypotheses on the spaces \mathcal{Z} and \mathcal{Y} , but with the introduction

of measure theory we will implicitly assume that these are “Polish” (separable complete metric) spaces.

How can we estimate the amount of information that is lost when we summarize the microscopic information by the macroscopic? Assuming that \mathcal{Y} and \mathcal{Z} are denumerable, it is natural so suppose that the uncertainty associated with a state $y \in \mathcal{Y}$ is a function of the cardinality of the pre-image, i.e. $\#\pi^{-1}(y)$.

If we carry out two independent measures of two different systems, we are tempted to say that the uncertainties are additive. Now, with obvious notation, $\#\pi^{-1}(y_1, y_2) = (\#\pi_1^{-1}(y_1)) (\#\pi_2^{-1}(y_2))$. To pass from this multiplicative operation to an addition, we take a multiple of the logarithm. We thus end up with Boltzmann’s celebrated formula, engraved on his tombstone in the Central Cemetery in Vienna :

$$S = k \log W, \tag{4}$$

where $W = \#\pi^{-1}(y)$ is the number of microscopic states compatible with the observed macroscopic state y . Here k is a physical constant, more precisely Planck’s constant.

In numerous cases, the space \mathcal{Z} of microscopic configurations is continuous, and in applying Boltzmann’s formula it is customary to replace the counting measure by a privileged measure : for example by Liouville measure if we are interested in a Hamiltonian system. Thus W in (4) can be the *volume* of microscopic states that are compatible with the macroscopic state y .

If the space \mathcal{Y} of macroscopic configurations is likewise continuous, then this notion of volume must be handled prudently : the fiber $\pi^{-1}(y)$ is typically of volume zero and thus of scarce interest. One is tempted to postulate, for a given topology,

$$S(y) = \text{p.f.}_{\varepsilon \rightarrow 0} \log |\pi^{-1}(B_\varepsilon(y))|,$$

where $B_\varepsilon(y)$ is the ball of radius ε centered at y and p.f. denotes the finite portion, that is we excise the divergence in ε , if indeed it has a global behavior.

If this last point is not at all evident, the universality is nonetheless verified in the particular case that interests us where the microscopic state \mathcal{Z} is the space of configurations of N particles, i.e. \mathcal{Y}^N , and where we begin by taking the limit $N \rightarrow \infty$. In this limit, as we will see, the mean entropy per molecule tends to a finite value and we can subsequently take the limit $\varepsilon \rightarrow 0$, which corresponds to an arbitrarily precise *macroscopic* measure. The result in, within a sign, nothing other than Boltzmann’s famous *H function*.

2.2 The entropy function H

We apply the preceding considerations with a macroscopic space made up of k different states : a macroscopic state is thus a vector (f_1, \dots, f_k) of frequencies with, of course, $f_1 + \dots + f_k = 1$. It is supposed that the measure is absolute (no error) and that $Nf_j = N_j$ is entire for all j . The number of microscopic states associated with this macroscopic state then equals

$$W = \frac{N!}{N_1! \dots N_k!}$$

(If N_j positions are prepared in the j -th state and if we number the positions from 1 to N , then there are $N!$ ways of arranging the N balls in the N positions and it's subsequently impossible to distinguish between permutations on the interior of any single box.)

According to Stirling's formula, when $N \rightarrow \infty$ we have $\log N! = N \log N - N + \log \sqrt{2\pi N} + o(1)$. It follows easily that

$$\begin{aligned} \frac{1}{N} \log W &= - \sum_i \frac{N_i}{N} \log \frac{N_i}{N} + O\left(\frac{k \log N}{N}\right) \\ &= - \sum_i f_i \log f_i + o(1). \end{aligned}$$

We note that we can also arrive at the same result without using Stirling's formula, thanks to the so-called method of types [41, section 12.4].

If now we increase the number of experiments, we can formally make k tend to ∞ , while making sure that k remains small compared to N . Let us suppose that we have at our disposal a reference measure ν on the macroscopic space \mathcal{Y} , and that we can separate this space into "cells" of volume (measure) $\delta > 0$, corresponding to the different states. When $\delta \rightarrow 0$, if the system has a statistical distribution $f(y)$ with respect to the measure ν , we can reasonably think that $f_i \simeq \delta f(y_i)$, where y_i is a representative point of cell number i . But then

$$\sum_i f_i \log \frac{f_i}{\delta} \simeq \delta \sum_i f(y_i) \log f(y_i) \simeq \int f \log f d\nu,$$

where the last approximation comes from the second sum being a Riemann sum of the integral.

We have ended up with **Boltzmann's H function** : being given a reference measure ν on a space \mathcal{Y} and a probability measure μ on \mathcal{Y} ,

$$H_\nu(\mu) = \int f \log f d\nu, \quad f = \frac{d\mu}{d\nu}. \quad (5)$$

If ν is a probability measure, or more generally a measure of finite mass, it is easy to extend this formula to all probabilities μ by setting $H_\nu(\mu) = +\infty$ if μ isn't absolutely continuous with respect to ν . If ν is a measure of infinite mass, more precautions must be taken; we could require at the very least the finiteness of $\int f (\log f)_- d\nu$.

We then note that if the macroscopic space \mathcal{Y} bears a measure ν , then the microscopic space $\mathcal{Z} = \mathcal{Y}^N$ bears a natural measure $\nu^{\otimes N}$.

We are now ready to state the precise mathematical version of the formula for the function H : given a family $\{\varphi_j\}_{j \in \mathbb{N}}$ of bounded and uniformly continuous functions, then

$$\begin{aligned} \lim_{k \rightarrow \infty} \lim_{\varepsilon \rightarrow 0} \lim_{N \rightarrow \infty} \frac{1}{N} \log \nu^{\otimes N} \left[\left\{ (y_1, \dots, y_N) \in \mathcal{Y}^N; \quad \forall j \in \{1, \dots, k\}, \right. \right. \\ \left. \left. \left| \int \varphi_j d\mu - \frac{1}{N} \sum_i \varphi_j(y_j) \right| \leq \varepsilon \right\} \right] = -H_\nu(\mu). \end{aligned} \quad (6)$$

We thus interpret N as the number of particles; the φ_j as a sequence of observables for which we measure the average value; and ε as the precision of the measurements. This formula summarizes in a concise manner the essential information contained in the function H .

If ν is a probability measure, statement (6) is known as **Sanov's theorem** [43] and is a leading result in the theory of large deviations. Before giving the interpretation of (6) in this theory, note that once we know how to treat the case where ν is a probability measure we easily deduce the case where ν is a measure of finite mass; however, I have no knowledge of any rigorous discussion in the case where ν is of infinite mass, even though we may expect that the result remains true.

2.3 Large deviations

Let ν be a probability measure and suppose that we independently draw random variables y_i according to ν . The empirical measure $\hat{\mu} = N^{-1} \sum \delta_{y_j}$ is then a random measure, almost certainly convergent to ν as $N \rightarrow \infty$ (it's Varadarajan's theorem, also called the fundamental law of statistics [49]). Of course it's possible that appearances deceive and that we think we are observing a measure μ distinct from ν . This probability decreases exponentially with N and is roughly proportional to $\exp(-N H_\nu(\mu))$; in other words, the Boltzmann entropy dictates the rarity of conditions that lead to the "unexpected" observation μ .

2.4 Information

Information theory was born in 1948 with the remarkable treatise of Shannon and Weaver [94] on the "theory of communication" which is now a pillar for the whole industry of information transmission.

In Shannon's theory, somewhat disembodied for its reproduction and impassionate discussion, the quantity of information carried by the decoding of a random signal is defined as a function of the reciprocal of the probability of the signal (which is rare and precious). Using the logarithm allows having the additivity property, and Shannon's formula for the mean quantity gained in the course of decoding is obtained: $E \log(1/p(Y))$, where p is the law of Y . This of course gives Boltzmann's formula again!

2.5 Entropies at all stages

Entropy is not an intrinsic concept; it depends on the observer and the degree of knowledge that can be acquired through experiments and measures. The notion of entropy will consequently vary with the degree of precision of the description.

Boltzmann's entropy, as has been seen, informs us of the rarity of the kinetic distribution function $f(x, v)$ and the quantity of microscopic information remaining to be discovered once f is known.

If to the contrary we are given the microscopic state of all the microscopic particles, no hidden information remains and thus no more entropy. But if we

are given a probability μ^N on the microscopic configurations, then the concept of entropy again has meaning : the entropy will be lower when the probability μ^N is concentrated and informative in itself. We thus find ourselves with a notion of **microscopic entropy**, $S_N = -H_N$,

$$H_N = \frac{1}{N} \int f^N \log f^N dx^N dv^N,$$

which is typically conserved by the Newtonian dynamic in consequence of Liouville's theorem. We can verify that

$$H_N \geq H(\mu^{1:N}),$$

with equality when μ^N is a tensor product and there are thus no correlations between particles. The idea is that the state of the microscopic particles is easier to obtain by multiparticle measurements than particle by particle — unless of course when the particles are independent !

In the other direction, we can also be given a *less precise* distribution than the kinetic distribution : this typically concerns a hydrodynamic description, which involves only the density field $\rho(x)$, the temperature $T(x)$ and mean velocity $u(x)$. The passage from the kinetic formalism to hydrodynamic formalism is accomplished by simple formulas :

$$\begin{aligned} \rho(x) &= \int f(x, v) dv; & u(x) &= \rho(x)^{-1} \int f(x, v) v dv; \\ T(x) &= \frac{1}{d \rho(x)} \int f(x, v) \frac{|v - u(x)|^2}{2} dv. \end{aligned}$$

With this description is associated a notion of **hydrodynamic entropy** :

$$S_h = - \int \rho \log \frac{\rho}{T^{d/2}}.$$

This information is always lower than kinetic information. We have, finally, a hierarchy : first microscopic information at the low level, then “mesoscopic” information from the Boltzmann distribution function, finally “macroscopic information” contributed by the hydrodynamic description. The relative proportions of these different entropies constitute excellent means for appraising the physical state of the systems considered.

2.6 The universality of entropy.

Initially introduced within the context of the kinetic theory of gases, entropy is an abstract and evolving mathematical concept, which plays an important role in numerous areas of physics, but also in branches of mathematics having nothing to do with physics, such as information theory and in other sciences.

Some mathematical implications of the concept are reviewed in my survey *H-Theorem and beyond : Boltzmann's entropy in today's mathematics* [106].

3 Order and chaos

Intuitively, a microscopic system is ordered if all its particles are arranged in a coordinated, *correlated* way. On the other hand, it is chaotic if the particles, doing just as they please, act entirely independently from one another. Let us reformulate this idea : a distribution of particles is chaotic if each of the particles is oblivious to all the others, in the sense that a gain of information obtained for a given particle brings no gain in information about any other particle. This simple notion, key to Boltzmann's equation, presents some important subtleties that we will briefly mention.

3.1 Microscopic chaos

To say that random particles that are oblivious to each other is equivalent to saying that their joint law is tensorial. Of course, even if the particles are unaware of each other initially, they will enter into interaction right away and the independence property will be destroyed. In the case of hard spheres, the situation is still worse : the particles are obliged to consider each another since the spheres can't interpenetrate. Their independence is thus to be understood asymptotically when the number of particles becomes very large ; and experiments seeking to measure the degree of independence will involve but a finite number of particles. This leads naturally to the definition that follows.

Let \mathcal{Y} be a macroscopic space and, for each N , let μ^N be a probability measure, assumed symmetric (invariant under coordinate permutations). We say that (μ^N) is chaotic if there exists a probability μ such that $\mu^N \simeq \mu^{\otimes N}$ in the sense of the weak topology of product measures. Explicitly, this means that for each $k \in \mathbb{N}$ and for all choices of the continuous functions $\varphi_1, \dots, \varphi_k$ bounded on \mathcal{Y} , we have

$$\int_{\mathcal{Y}^N} \varphi_1(y_1) \dots \varphi_k(y_k) \mu^N(dy_1 \dots dy_N) \xrightarrow{N \rightarrow \infty} \left(\int \varphi_1 d\mu \right) \dots \left(\int \varphi_k d\mu \right). \quad (7)$$

Of course, the definition can be quantified by introducing an adequate notion of distance, permitting us to measure the gap between μ^N and $\mu^{\otimes N}$. We can then say that a distribution μ^N is more or less chaotic. We again emphasize : what matters is the independence of a small number k of particles taken from among a large number N .

It can be shown (see the argument in [99]) that it is equivalent to impose property (7) for all $k \in \mathbb{N}$, or simply for $k = 2$. Thus chaos means precisely that *2 particles drawn randomly from among N are asymptotically independent when $N \rightarrow \infty$* . The proof proceeds by observing the connections between chaos and empirical measure.

3.2 Chaos and empirical measure

By the law of large numbers, chaos automatically implies an asymptotic determinism : with very high probability, the empirical measure approaches the sta-

tistical distribution of an arbitrary particle when the total number of particles becomes gigantic.

It turns out that, conversely, *the correlations accommodate very badly a macroscopic prescription of density*. Before giving a precise statement, we will illustrate this concept in a simple context. Consider a box with two compartments, in which we distribute a very large number N of *indistinguishable* balls. A highly correlated state would be a one in which all the particles occupy the same compartment : if I draw two balls at random, the state of first ball informs me completely about the state of the second. But of course, from the moment when the respective numbers of balls in the compartments are fixed and both nonzero, such a state of correlation is impossible. In fact, if the particles are indistinguishable, when two are drawn at random, the only information gotten is obtained by exploiting the fact that they are distinct, so that knowledge of the state of the first particle reduces slightly the number of possibilities for the state of the second. Thus, if the first particle occupies state 1, then the chances of finding the second particle in state 1 or 2 respectively aren't $f_1 = N_1/N$ and $f_2 = N_2/N$, but $f'_1 = (N_1 - 1)/N$ and $f'_2 = N_2/N$. The joint distribution of a pair of particles is thus very close to the product law.

By developing the preceding argument, we arrive at an elementary but conceptually profound general result, whose proof can be found in Sznitman's course [99] (see also [40, p. 91]) : *microscopic chaos is equivalent to the determinism of the empirical measure*. More precisely, the following statements are equivalent :

- (i) (μ^N) is μ -chaotic ;
- (ii) the empirical measure $\hat{\mu}^N$ associated with μ^N converges in law toward the deterministic measure μ

By "empirical measure $\hat{\mu}^N$ associated with μ^N " we understand the measure of the image of μ^N under the mapping $(y_1, \dots, y_N) \mapsto N^{-1} \sum \delta_{y_i}$, which is a measure of random probability. Convergence in law means that, for each continuous bounded function Φ on the space of probability measures, we have

$$\int \Phi \left(\frac{1}{N} \sum \delta_{y_i} \right) \mu^N(dy_1 \dots dy_N) \xrightarrow{N \rightarrow \infty} \Phi(\mu).$$

In informal language, given a statistical quantity involving $\hat{\mu}^N$, we can obtain an excellent approximation for large N by replacing, in the expression for this statistic, $\hat{\mu}^N$ by μ .

The notion of chaos thus presented is a weak and susceptible to numerous variants ; the general idea being that μ^N must be close to $\mu^{\otimes N}$. The stronger concept of **entropic chaos** was introduced by Ben Arous and Zeitouni [13] : there $H_{\mu^{\otimes N}}(\mu^N) = o(N)$ is imposed. A related notion was developed by Carlen, Carvalho, Leroux, Loss and Villani [32] in the case where the microscopic space isn't a tensor product, but rather a sphere of large dimension ; the measure $\mu^{\otimes N}$ is replaced by the restriction of the product measure to the sphere. Numerous other variants remain to be discovered.

3.3 The reign of chaos

In Boltzmann's theory, it is postulated that *chaos is the rule* : when a system is prepared, it is *a priori* in a chaotic state. Here are some possible arguments :

- if we can act on the macroscopic configuration, we will not have access to the microscopic structure and it is very difficult to impose correlations ;
- the laws that underlie the microscopic laws are unknown to us and we may suppose that they involve a large number of factors destructive to correlations ;
- the macroscopic measure observed in practice seems always well determined and not random ;
- if we fix the macroscopic distribution, the entropy of a chaotic microscopic distribution is larger than the entropy of a non chaotic microscopic distribution.

We explain the last argument. If we are given a probability μ on \mathcal{Y} , then the product probability $\mu^{\otimes N}$ is the maximum entropy among all the symmetric probabilities μ^N on \mathcal{Y}^N having μ as marginal. In view of the large numbers N in play, this represents a phenomenally larger number of possibilities.

The microscopic measure μ_0^N can be considered as an object of Bayesian nature, an *a priori* probability on the space of possible observations. This choice, in general arbitrary, is made here in a canonical manner by maximization of the entropy : in some way we choose the distribution that leaves the most possibilities open and makes the observation the most likely. We thus join the scientific approach of maximum likelihood, which has proved its robustness and effectiveness — while skipping the traditional quarrel between frequentists and Bayesians !

The problem of the propagation of chaos consists of showing that our chaos hypothesis, made on the initial data (it's not entirely clear how), is propagated by the microscopic dynamic (which is well defined). The propagation of chaos is essential for two reasons : first, it shows that independence is asymptotically preserved, providing statistical information about the microscopic dynamic ; secondly, it guarantees that *the statistical measure remains deterministic*, which allows hope for the possibility of a **macroscopic equation** governing the evolution of this empirical measure or its approximation.

3.4 Evolution of entropy

A recurrent theme in the study of dynamical systems, at least since Poincaré, is the search for invariant measures ; the best known example is Liouville measure for Hamiltonian systems. This measure possesses the remarkable additional property of tensorizing itself.

Suppose that we have a microscopic dynamic on \mathcal{Y}^N and a measure ν on the space \mathcal{Y} such that $\nu^{\otimes N}$ is an invariant measure for the microscopic dynamic ; or more generally that there exists a ν -chaotic invariant measure on \mathcal{Y}^N . What happens with the preservation of microscopic volume in the limit $N \rightarrow \infty$?

A simple consequence of preservation of volume is conservation of macroscopic information $H_{\nu^{\otimes N}}(\mu_t^N)$, where μ_t^N is the measure image of μ_0^N through the

microscopic evolution. In fact, since μ_t^N is preserved by the flow (by definition) and $\nu^{\otimes N}$ likewise, the density $f^N(t, y_1, \dots, y_N)$ is constant along the trajectories of the system, and it follows that $\int f^N \log f^N d\nu^{\otimes N}$ is likewise constant.

Matters are more subtle for macroscopic information. Of course, if the various particles evolve independently from one another, the measure μ_t^N remains factored for all time, and we easily deduce that the macroscopic entropy remains constant. In general, the particles interact with one another, which destroys independence; however if there is propagation of chaos in a sufficiently strong sense, the independence is restored as $N \rightarrow \infty$, and we consequently have determinism for the empirical measure. All the typical configurations for the microscopic initial measure μ_0^N give way, after a time t , to an empirical measure $\hat{\mu}_t^N \simeq \mu_t$, where μ_t is well determined. But it is possible that *other* microscopic configurations are compatible with the state μ_t , configurations that haven't been obtained by evolution from typical initial configurations.

In other words : if we have a propagation of macroscopic determinism between the initial time and the time t , and that the microscopic dynamic preserves the reference measure produced, then we expect that the volume of the admissible microscopic states do not decrease between time 0 and time t . Keeping in mind the definition of entropy, we would have $e^{NS(t)} \geq e^{NS(0)}$, where $S(t)$ is the value of the entropy at time t . We thus expect that the entropy doesn't decrease over the course of the temporal evolution :

$$S(t) \geq S(0).$$

But then why not reverse the argument and say that chaos at time t implies chaos at time 0, by reversibility of the microscopic dynamic? This argument is in general inadmissible in that an exact notion of the chaos propagated isn't specified. The initial data, prepared "at random" with just one kinetic distribution constraint, is supposed chaotic in a less strong sense; this depends on the microscopic evolution.

The notion of scale of interaction plays an important role here. Certain interactions take place on a macroscopic scale, other on a microscopic scale, which is to say that all or part of the interaction law is coded in the parameters that are invisible on the macroscopic level. In this last case, the notion of chaos conducive to the propagation of the dynamic risks not being visible on the macroscopic scale and we can expect a degradation of the notion of chaos.

From there, the discussion must involve the details of the dynamic, and our worst troubles begin.

4 Chaotic equations

After the introduction of entropy and chaos, we can return to the Newtonian systems of section 1, for which the phase space is composed of positions and speeds. A **kinetic equation** is an evolution equation bearing on the distribution $f(t, x, v)$; the important role of the velocity variable v justifies the terminology *kinetic*. By

extension, in the case where there are external degrees of freedom (orientation of molecules for example), by extension we still speak of kinetic equations.

As descendents of Boltzmann, we pose the problem of deducing the macroscopic evolution starting from the underlying microscopic model. This problem is in general of considerable difficulty. The fundamental equations are those of Vlasov, Boltzmann, Landau and Balescu-Lenard, published respectively in 1938, 1867, 1936 and 1960 (the more or less logical order of presentation of these equations doesn't entirely follow the order in which they were discovered...)

4.1 Vlasov's equation

Also called Boltzmann's equation without collisions, Vlasov's equation [112] is a mean field equation in the sense that all particles interact with one another. To deduce it from Newtonian dynamics, we begin by translating Newton's equation as an equation in the empirical measure; for this we write the evolution equation of an arbitrary observable :

$$\begin{aligned} & \frac{d}{dt} \frac{1}{N} \sum_i \varphi(X_i(t), \dot{X}_i(t)) \\ &= \frac{1}{N} \sum_i \left[\nabla_x \varphi(X_i, \dot{X}_i) \cdot \dot{X}_i + \nabla_v \varphi(X_i, \dot{X}_i) \cdot \ddot{X}_i \right] \\ &= \frac{1}{N} \sum_i \left[\nabla_x \varphi(X_i, \dot{X}_i) \cdot \dot{X}_i + \nabla_v \varphi(X_i, \dot{X}_i) \cdot \left(a \sum_j F(X_i - X_j) \right) \right]. \end{aligned}$$

This can be rewritten

$$\frac{\partial \hat{\mu}^N}{\partial t} + v \cdot \nabla_x \hat{\mu}_t^N + aN (F * \hat{\mu}_t^N) \cdot \nabla_v \hat{\mu}_t^N = 0. \quad (8)$$

If now we suppose that $aN \simeq 1$ and we make the approximation

$$\hat{\mu}_t^N(dx dv) \simeq f(t, x, v) dx dv,$$

we obtain Vlasov's equation

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f + \left(F *_{x} \int f dv \right) \cdot \nabla_v f = 0. \quad (9)$$

We note well that $\hat{\mu}_t^N$ in (8) is a weak solution of Vlasov's equation, so that the passage to the limit is conceptually very simple : it is simply a stability outcome of Vlasov's equation.

Quite clearly I have gone a bit far, for this equation is nonlinear. If $\hat{\mu} \simeq f$ in the sense of the weak topology of measures, then $F * \hat{\mu}$ converges to $F * \int f dv$ in a topology determined by the regularity of F , and if this topology is weaker than uniform convergence, nothing guarantees that $(F * \hat{\mu})\hat{\mu} \simeq (F * f)f$.

If F is in fact bounded and uniformly continuous, then the above argument can be made rigorous. If F is furthermore L -Lipschitz, then we can do better and

establish a stability estimate in weak topology : if (μ_t) and (μ_t'') are two weak solutions of Vlasov's equation, then

$$W_1(\mu_t, \mu_t') \leq e^{2 \max(1, L)|t|} W_1(\mu_0, \mu_0'),$$

where W_1 is the Wasserstein distance of order 1,

$$W_1(\mu, \nu) = \sup : \left\{ \int \varphi d\mu - \int \varphi d\nu; \quad \varphi \text{ 1-Lipschitz} \right\}.$$

Estimates of this sort are found in [95, Chapter 5] and date back to the 1970s (Dobrushin [48], Braun and Hepp [24], Neunzert [87]). Large deviation estimates can also be established as in [20].

However, for **singular interactions**, the problem of the Vlasov limit remains open, except for a result of Jabin and Hauray [64], which essentially assumes that (a) $F(x-y) = O(|x-y|^{-s})$ with $0 < s < 1$; and (b) the particles are initially well separated in phase space, so that

$$\inf_{j \neq i} \left(\left(|X_i(0) - X_j(0)| + \left| \dot{X}_i(0) - \dot{X}_j(0) \right| \right) \right) \geq \frac{c}{N^{\frac{1}{2d}}}.$$

Neither of these conditions is satisfied : the first lacks the Coulomb case of singularity order, while the second excludes the case of random data. However, it remains the sole result available at this time... To go further, it would be nice to have a sufficiently strong notion of chaos so as to be able to control the number of pairs (i, j) such that $|X_i(t) - X_j(t)|$ is small. In the absence of such controls, Vlasov's equation for singular interactions remains an act of faith.

This act of faith is very effective if the Vlasov-Poisson model, in which $F = -\nabla W$, where W a fundamental solution of $\pm\Delta$, is the universally accepted classic model in plasma physics [42, 71] as well as in astronomy [15]. In the first instance a particle is an electron, in the second a star ! The only difference lies in the sign : repulsive interaction for electrons, attractive for stars. We shouldn't be astonished to see stars considered in this way as microscopic objects : they are effectively so on the scale of a galaxy (which can tally 10^{12} stars...).

The theory of the Vlasov-Poisson equation itself remains incomplete. We can distinguish presently two principal theories, both developed in the entire space. That of Pfaffelmoser, exposted for example in [51], supposes that the initial data f_i is C^1 with compact support ; that of Lions-Perthame, reviewed in [23], supposes only that $|f_i| + |\nabla f_i|$ decreases sufficiently fast at infinity. Pfaffelmoser's theory has been adapted in periodic context in space [12], which isn't the case for the Lions-Perthame theory ; however, the hypothesis of compact support in velocity in the Pfaffelmoser theory is profoundly unsatisfying and work remains to be done in this area.

4.2 Boltzmann's equation

Vlasov's equation loses its relevance when the interactions have a short reach. A typical example is that of rarefied gas, for which the dominant interactions are binary and are uniquely produced in the course of "collisions" between particles.

Boltzmann's equation was established by Maxwell [80] and Boltzmann [21, 22]; it describes a situation where the interactions are of short reach and where each particle undergoes $O(1)$ impacts per unit of time. Much more subtle than the situation of Vlasov's mean field, the Boltzmann situation is nonetheless simpler than the hydrodynamic one where the particles undergo a very large number of collisions per unit of time.

We start by establishing the equation informally. The movement of a particle occurs with alternation of rectilinear trajectories and collisions, during the course of which its velocity changes so abruptly that we can consider the event as instantaneous and localized in space. We first consider the emblematic case of hard spheres of radius r : a collision occurs when two particles, with respective positions x and y and with respective velocities v and w , are found in a configuration where $|x - y| = 2r$ and $(w - v) \cdot (y - x) < 0$. We then speak of a *precollisional configuration*. We let $\omega = (y - x)/|x - y|$.

We now come to the central point in all Boltzmann's argument: *when two particles encounter each other, with very strong probability they will (almost) not be correlated*: think of two people who encounter each other for the first time. We can consequently apply the hypothesis of molecular chaos to such particles, and we find that the probability of an encounter between these particles is proportional to

$$\begin{aligned} f^{2;N}(t, x, v, x + 2r\omega, w) &\simeq f^{1;N}(t, x, v) f^{1;N}(t, x + 2r\omega, w) \\ &\simeq f^{1;N}(t, x, v) f^{1;N}(t, x, w), \end{aligned}$$

provided thus that $(w - v) \cdot \omega < 0$. We likewise need to take into account the relative velocities in order to evaluate the influence of the particles of velocity w on the particles of velocity v : the probability of encountering a particle of velocity w in a unit of time is proportional to the product of $|v - w|$ and the effective section (in dimension 3 this is the apparent area of the particles, or πr^2) and by the cosine of the angle between $v - w$ and ω (the extreme case is where $v - w$ is orthogonal to ω , which is to say that the two particles but graze each other, clearly an event of probability zero. Each of these collisions removes a particle of velocity v , and we thus have a negative term, the *loss term*, proportional to

$$- \iint f(t, x, v) f(t, x, v_*) |(v - v_*) \cdot \omega| dv_* d\omega.$$

The velocities after the collision are easily calculated:

$$v' = v - (v - v_*) \cdot \omega \omega; \quad v'_* = v_* + (v - v_*) \cdot \omega \omega \quad (10)$$

These velocities matter little in the final analysis.

However, we also need to take account of all the particles of velocity v that have been created by collisions between particles of arbitrary velocities. By microscopic reversibility, these velocities are of the form (v', v'_*) , and our problem is to take account of all the possible pairs (v', v'_*) , which in this problem of computing the *gain term* are the *pre-collisional* velocities. We thus again apply the hypothesis of

pre-collisional chaos and obtain finally the expression of the **Boltzmann equation for solid spheres** :

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f = Q(f, f), \quad (11)$$

where

$$\begin{aligned} & Q(f, f)(t, x, v) \\ &= B \int_{\mathbb{R}^3} \int_{S^2_-} |(v - v_*) \cdot \omega| \left(f(t, x, v') f(t, x, v'_*) - f(t, x, v) f(t, x, v_*) \right) dv_* d\omega, \end{aligned}$$

Here S^2_- denotes the pre-collisional configurations $\omega \cdot (v - v_*) < 0$, and B is a constant. By using the change of variable $\omega \rightarrow -\omega$ we can symmetrize this expression and arrive at the final expression

$$\begin{aligned} & Q(f, f)(t, x, v) \\ &= B \int_{\mathbb{R}^3} \int_{S^2} |(v - v_*) \cdot \omega| \left(f(t, x, v') f(t, x, v'_*) - f(t, x, v) f(t, x, v_*) \right) dv_* d\omega. \end{aligned} \quad (12)$$

The operator (12) is the **Boltzmann collision operator for solid spheres**. The problem now consists of justifying this approximation.

To do this, in the 1960s Grad proposed a precise mathematical limit : have r tend toward 0 and at the same time N toward infinity, so that $Nr^2 \rightarrow 1$, which is to say that the total effective section remains constant. Thus a given particle, moving among all the others, will typically encounter a finite number of them in a unit of time. One next starts with a microscopic probability density $f_0^N(x^N, v^N) dx^N dv^N$, which is allowed to evolve by the Newtonian flow \mathcal{N}_t , and one attempts to show that the first marginal $f^{1:N}(t, x, v)$ (obtained by integrating all the variables except the first position variable and the first velocity variable) converges in the limit to a solution of the Boltzmann equation.

The Boltzmann-Grad limit is also often called the **weak density limit** [40, p. 60] : in fact, if we start from the Newtonian dynamic and fix the particle size, we will dilate the spatial scale by a factor $1/\sqrt{N}$ and the density will be of the order $N/N^{3/2} = N^{-1/2}$.

At the beginning of the 1970's, Cercignani [37] showed that Grad's program could be completed if one proved a number of *plausible* estimates ; shortly thereafter, independently, Lanford [69] sketched the proof of the desired result.

Lanford's theorem is perhaps the most important mathematical result in kinetic theory. In this theorem, we are given microscopic densities f_0^N such that for each k the densities $f_0^{k:N}$ of the k particle marginals are continuous, satisfy Gaussian bounds at high velocities and converge uniformly outside the collisional configurations (those where the positions of two distinct particles coincide) to their limit $f_0^{\otimes k}$. The conclusion is that there exists a time $t_* > 0$ such that $f_t^{k:N}$ converges almost everywhere to $f_t^{\otimes k}$, where f_t is a solutions of Boltzmann's equation, for all time $t \in [0, t_*]$.

Lanford's estimates were later rewritten by Spohn [95] and by Illner and Pulvirenti [61, 62] who replaced the hypothesis of small time by a smallness hypothesis

on the initial data, permitting Boltzmann’s equation to be treated as a perturbation of free transport. These results are reviewed in [40, 90, 95].

The technique used by Lanford and his successors goes through the **BBGKY hierarchy** (Bogoliubov-Born-Green-Kirkwood-Yvon), the method by which the evolution of the marginal for a particle $f^{1:N}$ is expressed as a function of the marginal for two particles $f^{2:N}$; the evolution of a two-particle marginal $f^{2:N}$ as a function of a three-particle marginal $f^{3:N}$, and so forth. This procedure is especially uneconomical (in the preceding heuristic argument, we only use $f^{1:N}$ and $f^{2:N}$, but there is no known alternative).

Each of the equations of the hierarchy is then solved via Duhamel’s formula, applying successively the free transport and collision operators, and by summing over all the possible collisional history. The solution at time t is thus formally expressed, as with an exponential operator, as a function of the initial data and we can apply the chaos hypothesis on (f_0^N) .

We then pass to the limit $N \rightarrow \infty$ in each of the equations, after having verified that we can neglect pathological “recollisions”, where a particle again encounters a particle that it had just encountered the first time, and which is thus not unknown to it. This point is subtle : in [40, appendix 4.C] an a dynamic that is *a priori* simpler than that of solid spheres, due to Uchiyama, is discussed, with only four velocities in the plane, for which the recollision configurations can’t be neglected, and the kinetic limit doesn’t exist.

It remains to identify the result with the series of tensor products of the solution to Boltzmann’s equation and conclude by using a uniqueness result.

Spohn [95, section 4.6] shows that one can give more precise information on the microscopic distribution of the particles : on the small scale, this follows a homogeneous **Poisson law** in phase space. This is consistent with the intuitive idea of molecular chaos.

Lanford’s theorem settled a controversy that had lasted since Boltzmann himself; but it leaves numerous questions in suspense. In the first place, it is limited to a small time interval (on which only about 1/6 of the particles have had time to collide... but the conceptual impact of the theorem is nonetheless important). The variant of Illner and Pulvirenti lifts this restriction of small time, but the proof does not lend itself to a bounded geometry. As for lifting the smallness restriction, at the moment it is but a distant dream.

Next, to this day the theorem has only been proved for a system of solid spheres; long-range interactions are not covered. Cercignani [36] notes that the limit of Boltzmann-Grad for such interactions poses subtle problems, even from the formal viewpoint.

Finally, the most frustrating thing is that Lanford avoided discussion of **pre-collisional chaos**, the notion that particles that are about to collide are not correlated. This notion is very subtle(!), because just after the collision, correlations have inevitably taken place. In other words, we have *pre-collisional chaos, but not post-collisional*.

What does pre-collisional chaos mean exactly? For the moment we don’t have a precise definition. It’s certainly a stronger notion than chaos in the usual sense; it involves too a de-correlation hypothesis that is seen on a set of codimension

1, i.e. configurations leading to collisions. We would infer that it is a notion of chaos where we have replaced the weak topology by a uniform topology; but that can't be so simple, since chaos in a uniform topology also implies post-collisional chaos, which is incompatible with pre-collisional chaos! In fact, the continuity of the two-particle marginal along a collision would imply

$$\begin{aligned} f(t, x, v) f(t, x, v_*) &\simeq f^{(2;N)}(t, x, v, x + 2r\omega, v_*) \\ &= f^{(2;N)}(t, x, v', x + 2r\omega, v'_*) \simeq f^{(1;N)}(t, x, v'^{(1;N)}(t, x, v'_*)). \end{aligned}$$

Passing to the limit we would have

$$f(t, x, v') f(t, x, v'_*) = f(t, x, v) f(t, x, v_*),$$

and as we will see in section 5.3 this implies that f is Gaussian in the velocity variable, which is of course false in general. Another argument for showing that post-collisional chaos must be incompatible with pre-collisional chaos consists of noting that if we have post-collisional chaos, the reasoning leading to the Boltzmann equation can be used again by expressing two-particle probabilities in terms of post-collisional probabilities... and we then obtain Boltzmann's equation in reverse :

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f = -Q(f, f).$$

As has been mentioned, Lanford's proof applies only to solid spheres; but Boltzmann's equation is used for a much larger range of interactions. The general form of the equation, say in dimension d , is the same as in (11) :

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f = Q(f, f), \quad (13)$$

but now

$$Q(f, f) = \int_{\mathbb{R}^d} \int_{S^{d-1}} (f' f'_* - f f_*) \tilde{B}(v - v_*, \omega) dv_* d\omega \quad (14)$$

where $\tilde{B}(v - v_*, \omega)$ depends only on $|v - v_*|$ and $|(v - v_*) \cdot \omega|$. There exist several representations of this integral operator (see [103]); it is often convenient to change variables by introducing another angle, $\sigma = (v' - v'_*)/|v - v_*|$, so that the formulas (10) must be replaced by

$$v' = \frac{v + v_*}{2} + \frac{|v - v_*|}{2} \sigma, \quad v'_* = \frac{v + v_*}{2} - \frac{|v - v_*|}{2} \sigma. \quad (15)$$

We must then replace the collision kernel \tilde{B} by B so that

$$B d\sigma = \tilde{B} d\omega.$$

Explicitly, we find

$$\frac{1}{2} \tilde{B}(z, \omega) = \left| 2 \frac{z}{|z|} \cdot \omega \right|^{d-2} B(z, \sigma).$$

The precise form of B (or, in an equivalent way, of \tilde{B}) is obtained by a classical scattering computation that goes back to Maxwell and which can be found in [38] : for an impact parameter $p \geq 0$ and a relative velocity $z \in \mathbb{R}^3$, the deviation angle θ equals

$$\theta(p, z) = \pi - 2p \int_{s_0}^{+\infty} \frac{ds/s^2}{\sqrt{1 - \frac{p^2}{s^2} - 4\frac{\phi(s)}{|z|^2}}} = \pi - 2 \int_0^{\frac{p}{s_0}} \frac{du}{\sqrt{1 - u^2 - \frac{4}{|z|^2}\phi\left(\frac{p}{u}\right)}},$$

where s_0 is the positive root of

$$1 - \frac{p^2}{s_0^2} - 4\frac{\phi(s_0)}{|z|^2} = 0.$$

So B is implicitly defined by

$$B(|z|, \cos \theta) = \frac{p}{\sin \theta} \frac{dp}{d\theta} |z|. \quad (16)$$

We write either $B(|z|, \cos \theta)$ or $B(z, \sigma)$, it being understood that the deviation angle θ is the angle formed by the vectors $v - v_*$ and $v' - v'_*$.

When $\phi(r) = 1/r$, we recover Rutherford's formula for the Coulomb deviation :

$$B(|v - v_*|, \cos \theta) = \frac{1}{|v - v_*|^3 \sin^4(\theta/2)}.$$

When $\phi(r) = 1/r^{s-1}$, $s > 2$, the collision kernel isn't computed explicitly, but it can be shown that (always in dimension 3)

$$B(|v - v_*|, \cos \theta) = b(\cos \theta) |v - v_*|^\gamma, \quad \gamma = \frac{s-5}{s-1}. \quad (17)$$

Furthermore, the function b , defined implicitly, is locally smooth with a *non integrable angular singularity* when $\theta \rightarrow 0$:

$$\sin \theta b(\cos \theta) \sim K\theta^{-1-\nu}, \quad \nu = \frac{2}{s-1}. \quad (18)$$

This singularity corresponds to collisions with large impact parameter p , where there is scant deflection. It is inevitable once the forces are of infinite range : in fact

$$\int_0^\pi B(|z|, \cos \theta) \sin \theta d\theta = |z| \int_0^\pi p \frac{dp}{d\theta} d\theta = |z| \int_0^{p_{\max}} p dp = \frac{|z| p_{\max}^2}{2}. \quad (19)$$

In the particular case $s = 5$, the collision kernel depends no longer on the relative velocity, but only on the *deviation angle* : we speak of Maxwellian molecules. By extension, we say that $B(v - v_*, \sigma)$ is a Maxwellian collision kernel if it depends only on the angle between $v - v_*$ and σ . The Maxwellian molecules are above all a phenomenological model, even if the interaction between a charged ion and a

neutral particle in a plasma is regulated by such a law [42, Vol. 1, p. 149]. The potentials in $1/r^{s-1}$ for $s > 5$ are called hard potentials, for $s < 5$ soft potentials. Often the angular singularity $b(\cos \theta)$ is truncated to small values of θ .

The Boltzmann equation is important in modeling rarefied gases, as explained in [39]. Nonetheless, because of its eventful history and its conceptual content, as well as the impact of Boltzmann’s treatise [22], this equation has exerted a fascination that goes far beyond its usefulness. The first mathematical works dedicated to it are those of Carleman [26, 27], followed by Grad [57]. Besides the article by Lanford [69] already mentioned, a result that has had a great impact is the weak stability theorem of DiPerna-Lions [47]. The equation is well understood in the spatially homogeneous setting for hard potentials with angular truncation, see e.g. [84]; and in the setting close to equilibrium, see e.g. [60]. We refer to the reference treatises [38, 40, 103] for a number of other results.

4.3 Landau’s equation

Boltzmann’s collisional integral loses its meaning for Coulomb interactions because of the extremely slow decrease of the Coulomb potential. The grazing collisions, with large impact parameter, then become dominant.

In 1936, Landau [67] established, using formal arguments, an asymptotic of Boltzmann’s kernel in this setting. Letting λ_D be the shielding distance (below which the Coulomb potential is no longer visible because of the global neutrality of the plasma), and r_0 the typical collision distance (distance of two particles whose interaction energy is comparable to the molecular excitation energy), the parameter $\Lambda = 2\lambda_D/r_0$ is the plasma parameter, and in the limit $\Lambda \rightarrow \infty$ (justified for “classical” plasmas), the Boltzmann operator can be formally replaced by a diffusive operator called Landau’s operator :

$$Q_B(f, f) \simeq \frac{\log \Lambda}{2\pi\Lambda} Q_L(f, f), \quad (20)$$

$$Q_L(f, f) = \nabla_v \cdot \left(\int_{\mathbb{R}^3} a(v - v_*) [f(v_*) \nabla_v f(v) - f(v) \nabla_v f(v_*)] dv_* \right), \quad (21)$$

$$a(v - v_*) = \frac{L}{|v - v_*|} \Pi_{(v-v_*)^\perp}, \quad (22)$$

where L is a constant and Π_{z^\perp} denotes orthogonal projection onto z^\perp .

The Landau approximation is now well understood mathematically in the context of a limit called **grazing collision asymptotics**; [3] can be consulted for a detailed discussion of this problem.

The Landau operator, both diffusive and integral, presents a remarkable structure. It is easily generalized to arbitrary dimensions $d \geq 2$, and the coefficient $L/|z|$ can be changed to $L|z|^{\gamma+2}$, where γ is the exponent appearing in (17). The models of hard potential type with $\gamma > 0$ have been completely studied in the spatially homogeneous case [45]; but it is definitely the case $\gamma = -3$ in dimension 3 that is physically interesting. In this case we only know how to prove the existence of weak solutions in the spatially homogeneous case (by adapting [1, Section

7] and the existence of strong solutions for perturbations of equilibrium [59]. This situation is entirely unsatisfactory.

4.4 The Balescu-Lenard equation

In 1960, Balescu [7] directly established a kinetic equation that describes the Coulomb interactions in a plasma; he thus recovers an equation published in another form by Bogoliubov [19] and simplified by Lenard. The reference [96] can be consulted for information on the genesis of the equation, and [8] for a synthetic presentation. The collision kernel in this equation takes the same form as (21), the difference is in the expression of the matrix $a(v - v_*)$, which now depends both on v and ∇f :

$$a_{BL}(v, v - v_*, \nabla f) = B \int_{|k| \leq K_0} \frac{k \otimes k}{|k|^4} \frac{\delta_{k \cdot (v - v_*)}}{|\epsilon(k, k \cdot v, \nabla f)|^2} dk, \quad (23)$$

$$\epsilon(k, k \cdot v, \nabla f) = 1 + \frac{b}{|k|^2} \int_{\mathbb{R}^3} \frac{k \cdot \nabla f(u)}{k \cdot (v - u) - i0} du.$$

This equation can also be obtained beginning with the study of long duration fluctuations in Vlasov's equation [71, Section 51].

The Balescu-Lenard equation is scarcely used because of its complexity. Under reasonable hypotheses, the Landau equation provides a good approximation [8, 70]. The procedure is adaptable interactions other than the Coulomb interaction, but in contrast with the limit of grazing collisions, it still provides the expression (21), the only change being in the coefficient L of (22), which is proportional to

$$\int_{\mathbb{R}^3} |k| |\hat{W}(k)|^2 dk,$$

where W is the interaction potential. This equation is briefly reviewed in [95, Chapter 6].

The mathematical theory of the Balescu-Lenard equation is wide open, both with regard to establishing it and to studying its qualitative properties; one of the rare rigorous papers on the subject is the linearized study of R. Strain [96]. Even though little used, the Balescu-Lenard equation is nonetheless the most respected of the collisional models in plasmas and it's an intermediary that allows justification for using the Landau collision operator to represent long duration fluctuations in particle systems; its theory represents a formidable challenge.

5 Boltzmann's theorem H

In this section we will start with Boltzmann's equation and examine several of its most striking properties. Much more information can be found in my long review article [103].

5.1 Modification of observables by collisions

The statistical properties of a gas are manifested, in the kinetic model, by the evolution of observables $\iint f(t, x, v) \varphi(x, v) dx dv$. Still assuming conditions with periodic limits and all the required regularity, we may write

$$\begin{aligned}
\frac{d}{dt} \iint f \varphi dx dv &= \iint (\partial_t f) \varphi dx dv \\
&= - \iint v \cdot \nabla_x f \varphi dx dv + \iint Q(f, f) \varphi dx dv \\
&= \iint v \cdot \nabla_x \varphi f dx dv + \iiint \tilde{B} (f' f'_* - f f_*) \varphi dx dv dv_* d\omega,
\end{aligned} \tag{24}$$

where we are still using the notation $f' = f(t, x, v')$, etc.

In the term with the integral in $f' f'_*$ we now make the pre-postcollisional change of variables $(v, v_*) \longrightarrow (v', v'_*)$, for all $\omega \in S^{d-1}$. This change of variable is unitary (Jacobian determinant equal to 1) and preserves \tilde{B} (its properties are traces of the microreversibility). After having renamed the variables, we obtain

$$\frac{d}{dt} \iint f \varphi dx dv = \iint v \cdot \nabla_x \varphi f dx dv + \iiint \tilde{B} f f_* (\varphi' - \varphi) dv dv_* d\omega dx. \tag{25}$$

This is, incidentally, the form in which Maxwell wrote Boltzmann's equation from 1867 on... We deduce from (25) that $\iint f dx dv$ is constant (happily!!), and we get an important quantity, the effective momentum transfer cross section $M(v - v_*)$ defined by

$$M(v - v_*) (v - v_*) = \int \tilde{B}(v - v_*, \omega) (v' - v) d\omega.$$

Even when \tilde{B} is a divergent integral, the quantity M may be finite, expressing the fact that the collisions modify the velocities in a statistically reasonable way. Readers may refer to [103] [2,3] for more details on the treatment of grazing singularities of \tilde{B} .

Boltzmann would improve Maxwell's procedure by making better use of the symmetries of the equation. First, by making the pre-postcollisional change of variables in the whole second term of (24) we obtain

$$\iiint \tilde{B} (f' f'_* - f f_*) \varphi dv dv_* d\omega dx = - \iiint \tilde{B} (f' f'_* - f f_*) \varphi' dv dv_* d\omega dx. \tag{26}$$

Instead of exchanging the pre- and postcollisional configurations, we may exchange the particles together : $(v, v_*) \longmapsto (v_*, v)$, which also clearly has a unitary Jacobian determinant. This gives us two new forms from (26) :

$$\iiint \tilde{B} (f' f'_* - f f_*) \varphi_* dv dv_* d\omega dx = - \iiint \tilde{B} (f' f'_* - f f_*) \varphi'_* dv dv_* d\omega dx. \tag{27}$$

By combining the four forms appearing in (26) and (27), we obtain

$$\begin{aligned} \frac{d}{dt} \iint f \varphi dv dx &= \iint f (v \cdot \nabla_x \varphi) dx dv \\ &\quad - \frac{1}{4} \iiint \tilde{B} (f' f'_* - f f_*) (\varphi' + \varphi'_* - \varphi - \varphi_*) dx dv dv_* d\omega. \end{aligned} \quad (28)$$

As a consequence of (28), we note in the first place that $\iint f \varphi$ is preserved if φ satisfies the functional equation

$$\varphi(v') + \varphi(v'_*) = \varphi(v) + \varphi(v_*) \quad (29)$$

for each choice of velocities v, v_* and of the parameter ω . Such functions are called **collision invariants** and reduce, under extremely weak hypotheses, to just linear combinations of the functions

$$1, \quad v_j \quad (1 \leq j \leq d), \quad \frac{|v|^2}{2}.$$

Readers may consult [40] in this regard. This is again natural : it is the macroscopic reflection of conservation of mass, the amount of motion and kinetic energy in during microscopic interactions.

5.2 Theorem H

We now come to the discovery that will put Boltzmann among the greatest names in physics. We choose $\varphi = \log f$ and assume all the regularity needed for carrying through the calculations ; in particular

$$\iint f v \cdot \nabla_x (\log f) dv dx = \iint v \cdot \nabla_x (f \log f - f) dv dx = 0.$$

Identity (28) thus becomes, taking into account the additive properties of the logarithm,

$$\frac{d}{dt} \iint f \log f dx dv = -\frac{1}{4} \iiint \tilde{B} (f' f'_* - f f_*) (\log f' f'_* - \log f f_*). \quad (30)$$

The logarithm function being increasing, the above expression is always negative ! Moreover, knowing that B is vanishes only on a set of measure zero, we see that the expression (30) is strictly negative whenever $f' f'_*$ is unequal to $f f_*$ almost everywhere, which is true for generic distributions. Thus, modulo the rigorous justification of the integrations by parts and a change of variables, we have proved that, *in Boltzmann's model, the entropy increases with time.*

The impact of this result is crucial. First, the heuristic microscopic reasoning of section 3.4 has been replaced by a simple argument that leads directly to the limit equation. Next, even if it is a manifestation of the **second law of thermodynamics**, the increase in the entropy in Boltzmann's model is deduced by

logical reasoning and not by a postulate (a law) which one accepts or not. Finally, of course, in doing so, Boltzmann displayed an **arrow of time** associated with his equation.

Not only is this macroscopic irreversibility not contradictory with microscopic reversibility, but it is in fact intimately linked to it : as has already been explained, it's the conservation of microscopic volume in phase space that guarantees the non decrease of entropy. For the rest, as L. Carleson was already astonished to find in 1979 while examining simplified models of Boltzmann's equation [35], we have that when the parameters are adjusted in such a way to achieve microscopic reversibility, we obtain theorem H at the same time. The phenomenon is well known in the context of the physics of granular media [105] : there the microscopic dynamic is dissipative (non reversible), including a loss of energy due to friction, and the macroscopic dynamic does not satisfy Theorem H !

From the informational point of view, the increase in entropy means that the system always runs toward macroscopic states that are more and more probable. This probabilistic idea is exacerbated by the formidable power of the combinatorics : we suppose for example that we are considering a gas with $N \simeq 10^{16}$ particles (which is roughly what we find in $1mm^3$ of gas under ordinary conditions!), and between time $t = t_1$ and time $t = t_2$ the entropy increases only 10^{-5} . The volume of microscopic possibilities is then multiplied by $e^{N[S(t_2)-S(t_1)]} = e^{10^{11}} \gg 10^{10^{10}}$. This phenomenal factor far exceeds the number of protons in the universe (10^{100} ?) or the number of 1000-page books that could be written by combining all the alphabetic characters of all the languages in the world...

The intuitive interpretation of Theorem H is thus rather eloquent : the high entropy states occupy, at the microscopic level, a place so monstrously larger than the states of low entropy that the microscopic system goes to them automatically. As we have seen, the logical reasoning justifying this scenario is complex and indirect, involving the propagation of chaos and macroscopic determinism — and to this day only a small portion of the program has been rigorously achieved.

5.3 Vanishing of entropy production

The increase in entropy admits a complement that is no less profound, frequently stated as a second part of Theorem H : the *characterization of cases of equality*, i.e. states for which the production of entropy vanishes.

We have seen in (30) that the entropy production equals

$$\int \text{PE}(f(x, \cdot)) dx, \quad (31)$$

where PE is the functional of “local production of entropy”, acting on the kinetic distributions $f = f(v)$:

$$\text{PE}(f) = \iiint \tilde{B}(v - v_*, \omega) (f(v)f(v_*) - f(v)f(v_*)) \log \frac{f(v)f(v_*)}{f(v)f(v_*)} dv dv_* d\omega. \quad (32)$$

For all reasonable models, we have $\tilde{B}(z, \omega) > 0$ almost everywhere, and it follows that the entropy production doesn't vanish for a distribution satisfying the functional equation

$$f(v)f(v'_*) = f(v)f(v_*) \quad (33)$$

for (almost) all v, v_*, ω . By taking the logarithm in (33) we recover equation (29), which shows that f must be the exponential of a collision invariant. In view of the form of these latter, and taking into account the integrability constraint of f , we obtain $f(v) = e^{a+b \cdot v + c|v|^2/2}$, which can be rewritten

$$f(v) = \rho \frac{e^{-\frac{|v-u|^2}{2T}}}{(2\pi T)^{d/2}}, \quad (34)$$

where $\rho \geq 0$, $T > 0$ and $u \in \mathbb{R}^d$ are constants. It is therefore a particular Gaussian, with covariance matrix proportional to the identity.

Maxwell already noticed that (34) cancels Boltzmann's collision operator : $Q(f, f) = 0$. Such a distribution is called **Maxwellian** in his honor. However, it is Boltzmann who first gave a convincing argument for the distributions (34) being *the only* solutions of the equation $PE(f) = 0$, and consequently the only solutions of $Q(f, f) = 0$. Let's honor him by sketching a variant of his original proof.

We begin by averaging (33) over all angles $\sigma = (v' - v'_*)/|v - v_*|$; the left side $|S^{d-1}|^{-1} \int f' f'_* d\sigma$ is then the mean of the function $\sigma \rightarrow f(c + r\sigma) f(c - r\sigma)$, where $c = (v + v_*)/2$ and $r = |v - v_*|$. This thus depends only on c and r or, in an equivalent way, only on $m = v + v_*$ and $e = (|v|^2 + |v_*|^2)/2$, respectively the amount of motion and the total energy involved in a collision. After passing to the logarithm, we find for $\varphi = \log f$ the identity

$$\varphi(v) + \varphi(v_*) = G(m, e). \quad (35)$$

The operator $\nabla_v - \nabla_{v_*}$, applied to the left hand side of (35), yields $\nabla\varphi(v) - \nabla\varphi(v_*)$. When we apply the same operator to the right hand side, the contribution of m disappears, and the contribution of e is collinear with $v - v_*$. We thus conclude that $F = \nabla\varphi$ satisfies

$$F(v) - F(v_*) \text{ is collinear with } v - v_* \text{ for all } v, v_*$$

and it is easy to deduce that $F(v)$ is an affine transformation, whence the conclusion. (Here is a crude method for showing the affine character of F , assuming regularity : we start by writing a Taylor expansion and noting that the Jacobian matrix of F is a multiple of the identity at each point, say $\partial_i F^j(v) z_i = \lambda(v) z_j$; then by differentiating with respect to v_k we deduce that $\partial_{ik} F^j = 0$ if $i \neq j$, and it follows that all the coefficients $\partial_i F^j$ cancel, after which we easily see that DF is a multiple of the identity.)

As a consequence of (31) and (34), the distributions $f(x, v)$ that cancel the production of Boltzmann entropy are precisely the distributions of the form

$$f(x, v) = \rho(x) \frac{e^{-\frac{|v-u(x)|^2}{2T(x)}}}{(2\pi T(x))^{d/2}}. \quad (36)$$

They are called **local Maxwellians** or else **hydrodynamic states**. In accordance with the kinetic description, these states are characterized by a considerable reduction in complexity, since they depend on but three fields : the density field ρ , the field of macroscopic velocities u and the temperature field T . These are the fields that enter into the hydrodynamic equations, thus the above terminology.

This discovery establishes a bridge between the kinetic and hydrodynamic descriptions : in a process where collisions are very numerous (weak Knudsen number), the finiteness of entropy production forces the dynamic to be concentrated near distributions that annul the production. This remark makes way for a vast program of hydrodynamic approximation of Boltzmann's equation, to which Hilbert alludes in his Sixth Problem. Readers can consult [54, 55, 93]. If the Boltzmann equation can be approached both by compressible and incompressible equations, we should note that it doesn't lead to the whole range of hydrodynamic equations, but only to those for perfect gases, i.e. those that conform to a law where pressure is proportional to ρT .

6 Entropic convergence : forced march to oblivion

If Maxwell discovered the importance of Gaussian velocity profiles, he didn't, as Boltzmann remarks, prove that these profiles are actually induced by the dynamic. Boltzmann wanted to complete this program, and for this recover the Maxwellian profiles not only as equilibrium distributions, but also as limits of the kinetic equation asymptotically as time becomes large ($t \rightarrow \infty$). This conceptual leap aimed at basing equilibrium statistical mechanics on its nonequilibrium counterpart — usually much more delicate — is still topical in innumerable contexts.

I have written a good bit on this topic and readers can consult the survey article [103, chapter 3], the course [108], the research article [46] or the research memoir [109]. In the sequel, in order to fix the ideas, I will suppose that the position variable lives in the torus \mathbb{T}^d .

6.1 Global Maxwellian

We have already encountered local Maxwellians that annul the collision operator. To likewise annul the transport operator $v \cdot \nabla_x$, it is natural to look for Maxwellians whose parameters ρ, u, T are homogeneous, constants independent of position. A single set of these parameters is compatible with the laws of conservation of mass, momentum and energy. The distribution thus obtained is called **global Maxwellian** :

$$M_{\rho u T} = \rho \frac{e^{-\frac{|v-u|^2}{2T}}}{(2\pi T)^{d/2}}.$$

Without loss of generality, even with a change in Galilean reference or physical scale, we may suppose that $\rho = 1$, $u = 0$ and $T = 1$, and we will denote the corresponding distribution by M .

This distribution is thus an equilibrium for Boltzmann's equation. Moreover, it is easy to verify that it is the distribution that maximizes entropy under the

constraints of fixed mass, linear momentum and total energy. This selection criterion foreshadows the classical theory of equilibrium statistical mechanics and Gibbs' famous canonical ensembles (Gibbs measure).

6.2 The entropic argument

Boltzmann now uses Theorem H for given a more solid justification to the global Maxwellian : he notes that

- entropy increases strictly unless it is in a hydrodynamic state
- the global Maxwellian, stationary, is the only hydrodynamic solution of Boltzmann's equation.

The image that emerges is that the entropy will continue to increase as much as possible since the distribution never remains “stuck” on a hydrodynamic solution ; the entropy will end up approaching the maximal entropy of the global Maxwellian, and convergence results.

In this regard we can make two remarks : the first is that the Lebesgue measure, which we have taken as the reference measure in Boltzmann's entropy, may be replaced by the Maxwellian : in fact

$$H(f) - H(M) = \iint f \log \frac{f}{M} dv dx = H_M(f),$$

where we have used the fact that $\log M$ is a collision invariant. The second remark is that the difference in entropies allows us to quantify the difference between the Gaussian and equilibrium, for example by virtue of the Csiszàr–Kullback–Pinsker inequality : $H_M(f) \geq \|f - M\|_{L^1}^2 / 2 \|M\|_{L^1}$.

Boltzmann's reasoning is essentially correct and it's not difficult to transform it into a rigorous argument by showing that sufficiently regular solutions of the Boltzmann equation approach Maxwellian equilibrium. In the context of spatially homogeneous solutions, T. Carleman expedited this reasoning in 1932 [26].

However, Boltzmann didn't have the means for making his argument qualitative; it would be necessary to wait almost a century before anyone dared to pose the problem of the speed of convergence toward Gaussian equilibrium, especially pertinent since the range of validity of Boltzmann's equation isn't eternal and is limited in time by phenomena such as the Poincaré recurrence theorem.

6.3 The probabilistic approach of Mark Kac

At the beginning of the 1950s, Kac [66] attempted to understand convergence toward equilibrium for the Boltzmann equation and began by simplifying the model. Kac ignores positions, simplifies the collision geometry extravagantly and invents a stochastic model *where randomness is present in the interaction* : when two particles interact, one draws at random the parameters describing the collision. Positions being absent, the particles all interact, each with all the others, and thus a “mean field” model is produced. This simplified probabilistic model is for Kac an opportunity for formalizing mathematically the notion of propagation of

chaos in the mean field equations, which had proved so fertile and would be taken up later by Sznitman [98] and many others.

Suspicious of Boltzmann’s equation, Kac wants to explain the convergence by microscopic probabilistic reasoning on the level of an N -particle system; he attempts to obtain spectral gap estimates that are uniform in N . His approach seems naive nowadays in that it underestimates the difficulty of treating dimension N ; nonetheless, the problem of determining the optimal spectral gap, resolved a half-century later, has proved to be very interesting [31, 65, 79]. For this subject readers can equally well consult [104, section 6] and [32], where there is interest in the entropic version of this “microscopic” program.

In 1966 McKean [81] resumed Kac’s work and drew a parallel with the problems of the central limit theorem. He introduced the tools of information theory to the subject, in particular Fisher information [41], which measures the difficulty in estimating a parameter such as the velocity of the particles. The program would be completed by Tanaka [100], who discovered new contracting distances, and would culminate with the work of Carlen, Carvalho, Gebetta, Lu, Toscani on “the central limit theorem for the Boltzmann equation” [29, 30, 33, 34]. This theory encompasses both basic convergence theorems on the combinatorics of interactions between particles and common tools with the central limit theorem (weak distances...), as well as counterexamples demonstrating extremely slow convergence to equilibrium.

This stochastic program allows us to dispense with Theorem H ; in fact it has also permitted the updating of several Lyapunov functionals: Tanaka’s contracting distance (of optimal transport), Fisher’s information. Nonetheless, from a technical point of view, the whole theory remains essentially confined to Maxwellian interactions (in which $\tilde{B}(v - v_*, \omega)$ depends only on the angle between $v - v_*$ and ω) and to spatially homogeneous gases. Chapter 4 of [103] is dedicated to particular properties, highly elegant moreover, of these interactions.

For gaining generality and for studying inhomogeneous situations or non Maxwellian interactions, the only robust approach known to this day is based on Theorem H .

6.4 Cercignani’s conjecture

Boltzmann’s Theorem H is general and relevant, so that it’s natural to look for its quantitative refinements. At the beginning of the 1980s, C. Cercignani asked if one could *underestimate* the production of local entropy as a function of the “non-Gaussianism” of the kinetic distribution, ideally by a multiple of the information $H_M(f)$. It was not until a decade later that Carlen and Carvalho [28], Desvillettes [44], without responding to Cercignani, could nonetheless present quantitative lower bounds for the production of entropy.

A more precise answer to this problem is obtained in my articles [101] and [103] (the first in collaboration with G. Toscani). Without loss of generality we suppose that $\int f dv = 1$, $\int f v dv = 0$, $\int f |v|^2 dv = d$; the general case being deducible by change of scale or reference frame. We begin by mentioning a surprisingly simple example, taken from [104], which is applied in a non physical situation: if

$B(v - v_*, \sigma) \geq K(1 + |v - v_*|^2)$, then

$$\text{PE}(f) \geq \left(K_B \frac{|S^{d-1}|}{8} \frac{d-1}{1+2d} \right) T_f^* H_M(f), \quad (37)$$

where

$$T_f^* = \inf_{e \in S^{d-1}} \int f(v) (v \cdot e)^2 dv.$$

The quantity T_f^* quantifies the non concentration of f near a hyperplane; it underestimates based on information on entropy or regularity, even automatically for radially symmetric distributions. The conciseness of the result masks a surprising proof technique whereby f is regularized by an auxiliary diffusion semigroup; under the effect of this this semigroup, the variation in the production of Boltzmann entropy is essentially underestimated by the production of Landau entropy, which is estimated in terms of Fisher information before integrating along the semigroup; see [104] or [108] for the details.

The hypothesis of quadratic growth in the relative velocity isn't physically realistic; it is nonetheless optimal in the sense that there exist counterexamples [16] for kernels with growth $|v - v_*|^\gamma$, for each $\gamma < 2$. One can then work on inequality (37) in order to derive from it a weaker underestimate that applies to realistically effective sections, such as the model of solid spheres; the principal difficulty lies in controlling the quantity of entropy production induced by the small relative velocities ($|v - v_*| \leq \delta$); the logarithms make this control delicate; see [104] for the details. In the end, for each $\varepsilon > 0$ we obtain the inequality

$$\text{PE}(f) \geq K_\varepsilon(f) [H(f) - H(M^f)]^{1+\varepsilon}, \quad (38)$$

where M^f is the Maxwellian associated with f , i.e. that with parameters ρ, u, T corresponding to the density, mean velocity and temperature of f . The constant $K_\varepsilon(f)$ depends only on ε , on the regularity C^r of f for r sufficiently large, on a moment $\int f|v|^s dv$ for s sufficiently large, and on a lower bound $f \geq K e^{-A|v|^q}$. The question as to whether these hypotheses can be relaxed remains open.

6.5 Conditional convergence

Inequality (38) concerns a function $f = f(v)$ but does not include spatial dependence; this is inevitable since the variable x does not enter into the study of the global production of entropy. Of course, (38) immediately implies (modulo the regularity bounds) convergence in $O(t^{-\infty})$ for the spatially homogeneous equation, i.e. the distance between the distribution and equilibrium tends to 0 faster than t^{-k} for each k ; yet this inequality doesn't resolve the inhomogeneous problem. The obstacle to be overcome is the *degeneracy of entropy production on hydrodynamic states*. A key to the study over long time of Boltzmann's equation thus consists in showing that not too much time is spent in a hydrodynamic, or approximately hydrodynamic, state. To avoid this trap, we can only depend on the transport, represented by the operator $v \cdot \nabla_x$. Grad [56] had understood in 1965, in a moreover rather obscure paper : "the question is whether the deviation from a local

Maxwellian, which is fed by molecular streaming in the presence of spatial inhomogeneity, is sufficiently strong to ultimately wipe out the inhomogeneity” (...) “a valid proof of the approach to equilibrium in a spatially varying problem requires just the opposite of the procedure that is followed in a proof of the *H*-Theorem, viz., to show that the distribution function does not approach too closely to a local Maxwellian.”

In 2000s, Desvillettes and I [46] rediscovered this principle formulated by Grad and we established a version of the **instability of hydrodynamic approximation** : if the system becomes, at a given moment, close to being hydrodynamic without being in equilibrium, then transport phenomena cause it to leave this hydrodynamic state. This is quantified, under the hypothesis of strong regularity, by studying second variations of the square of the norm, $\|f - M^f\|^2$, between $f = f(t, x, v)$ and the associated *local* Maxwellian

$$M^f(t, x, v) = \rho(t, x) \frac{e^{-\frac{|v-u(t,x)|^2}{2T(t,x)}}}{(2\pi T(t, x))^{d/2}},$$

$$\rho(t, x) = \int f(t, x, v) dv, \quad u(t, x) = \frac{1}{\rho(t, x)} \int f(t, x, v) v dv,$$

$$T(t, x) = \frac{1}{d\rho(t, x)} \int f(t, x, v) |v - u(t, x)|^2 dv.$$

In some way M^f is the best possible approximation of f by a hydrodynamic state, and the study of the variations $\|f - M^f\|$ allows us to verify that f cannot long remain close to a hydrodynamic state.

By adjoining (in an especially technical way, with the help of numerous functional inequalities) the quantitative Theorem *H* with the instability of hydrodynamic approximation, we end up with **conditional convergence** : a solution of the Boltzmann equation satisfying uniform regularity bounds converges toward equilibrium in $O(t^{-\infty})$. This result is constructive in the sense that the time constants involved depend only on the regularity bounds, on the form of the interaction and on the boundary conditions. The convergence resides in a system of inequalities that simultaneously involve the entropy and the distance to the hydrodynamic states. For example, one of them is written

$$\frac{d^2}{dt^2} \|f - M^f\|_{L^2}^2 \geq K \int |\nabla T|^2 dx - \frac{C}{\delta^{1-\varepsilon}} (\|f - M^f\|_{L^2}^2)^{1-\varepsilon} - \delta[H(f) - H(M)]. \quad (39)$$

In order to understand the contribution of such an inequality, we suppose that f becomes hydrodynamic at some moment : then $f = M^f$ and (38) is useless. But if the temperature is inhomogeneous and if δ in (39) is very small, then we are left with $(d^2/dt^2)\|f - M^f\|_{L^2}^2 \geq \text{const.}$, which certainly keeps f from remaining close to M^f for very long. Once f has exited the hydrodynamic approximation, we can reapply (38), and so forth. This reasoning only works when the temperature is inhomogeneous, but we can find other inequalities involving macroscopic velocity gradients and the density. We thus pass from a “passive” argument to an “active”

argument, where the increase in entropy is forced by differential inequalities rather than by the identification of a limit.

We end this section with several commentaries on the hypotheses. The regularity theory of the Boltzmann equation allows reduction of the general bounds to very particular bounds, e.g. it is known that the kinetic distribution is automatically minorized by a multiple of $e^{-|v|^q}$ if, for example, the equation is set on the torus and the solution is regular. It is also known that bounds on the moments of low order allow having bounds on arbitrarily high moments, etc. But regularity in the general context remains a celebrated open problem. The conditional convergence result shows that it's the final obstacle separating us from quantitative estimates of convergence to equilibrium ; it likewise unifies the already known results on convergence : both the case of spatially homogeneous distributions and that of distributions close to equilibrium are situations in which we have an almost complete regularity theory.

In studying convergence toward equilibrium for the Boltzmann equation, we observe a subtle interaction between the collision operator (nonlinear, degeneratively dissipative) and the transport operator (linear, conservative). Neither of the two, taken separately, would be sufficient for inducing convergence, but the combination of the two succeeds. This situation arises rather frequently and recalls the hypoellipticity problem in the theory of partial differential equations. By analogy, the *hypo-coercivity* problem is the study of the convergence properties for potentially degenerate equations.

A somewhat systematic study of these situations, both for linear and nonlinear equations, was made in my memoir [107]. The general strategy consists of constructing Lyapunov functionals adapted to the dynamic, while adjoining by a natural functional (like entropy) a well chosen term of lesser order. A case study is the “ $A^*A + B$ theorem”, inspired by Hörmander’s sum of squares theorem, which gives sufficient conditions on the commutators between operators A and B , with B antisymmetric, for the evolution $e^{-t(A^*A+B)}$ to be hypo-coercive. In the simpler variant, one of these conditions reminiscent of Hörmander’s Lie algebra condition, is the coercivity of $A^*A + [A, B]^*[A, B]$.

Hypercoercivity theory has now taken on a life of its own and there are already a number of striking results ; it continues to expand, especially in the nonlinear context. This is true as well both for kinetic theory, as in the paper [58] that will be mentioned in the next section, and outside of kinetic theory, as in the paper by Liverani and Olla on the hydrodynamic limits of certain particle systems [73].

In a nonlinear context, the principal result remains [109, theorem 51] ; this general statement allows for simplification of the proof of the conditional convergence theorem for the Boltzmann equation, and includes new interactions and limit conditions. See [109, part III] for more details.

6.6 Linearized system

The rate of convergence to equilibrium can be determined by a linearized study. We begin by flushing out a classical logical mistake : the linearized study can in

no case be a substitute for the nonlinear study, since linearization is only valid beginning from where the distribution is very close to equilibrium.

The linearized study of convergence requires overcoming three principal difficulties :

- quantitatively estimating the spectral gap for the linearized collision operator ;
- performing a spectral study of the linearization in a space appropriate for the nonlinear problem, so as to achieve a “connection” between the nonlinear study and the linearized study ;
- take into account the hydrodynamic degeneracy from a hypercoercive perspective : in fact, the linearized equation is just as degenerate as the nonlinear equation.

All of these difficulties have been resolved in the last decade by C. Mouhot and his collaborators Baranger, Gualdani and Mischler [11, 58, 82], at least in the emblematic case of solid spheres. Thus the recent article [58] establishes a conditional convergence result with exponential rate $O(e^{-\lambda t})$ instead of $O(t^{-\infty})$, and the rate λ is estimated in a constructive manner.

Exponential convergence is not a universal characteristic of Boltzmann’s equation : we don’t expect it for hard potentials or even for the moderately soft. To fix our ideas, let’s suppose that the collision kernel behaves like $|v - v_*|^\gamma b(\cos \theta)$. In the case where $b(\cos \theta) \sin^{d-2} \theta$ is integrable (often by angular truncation for the grazing collisions), the linearized collision operator only admits a spectral gap for $\gamma \geq 0$. An abundance of grazing collisions permits extending this condition, as Mouhot and Strain [83] showed : if $b(\cos \theta) \sin^{d-2} \theta \simeq \theta^{-(1+\nu)}$ for $\theta \rightarrow 0$ (important grazing collisions), then the linearized collision operator only allows a spectral gap for $\gamma + \nu \geq 0$. The regularity theory is presently under development for such equations (work of Gressman-Strain, Alexandre-Morimoto-Ukai-Xu-Yang), and we can wager that within a few years the linearized theory will cover all these cases.

For too soft potentials (or for the Landau’s Coulomb collision model), there is no spectral gap and the best result we can expect is fractional exponential convergence $O(e^{-\lambda t^\beta})$, $0 < \beta < 1$. Such estimates can be found in the paper of Guo and Strain [60].

6.7 Qualitative evolution of entropy

A recurrent theme in this whole section is the degeneracy related to hydrodynamic states, which disturbs the convergence to equilibrium. In the beginning years of this century, Desvillettes and I suggested that this degeneracy is reflected in oscillations in the production of entropy. Never previously observed, these oscillations have been identified in very precise numerical simulations by F. Filbet. Below I reproduce a striking curve, obtained with Boltzmann’s equation in a one-dimensional periodic geometry.

In this diagram, the logarithm of the function H has been drawn as a function of time ; the global rectilinear decrease thus corresponds to an exponential convergence toward the equilibrium state. The kinetic information has also been

FIG. 1 – Logarithmic evolution of the kinetic function H and of the hydrodynamic function H for the Boltzmann equation in a periodic box.

FIG. 2 – The same thing in a smaller box.

separated in to hydrodynamic information and “purely kinetic” information :

$$\int f \log \frac{f}{M} = \left(\int \rho \log \frac{\rho}{T^{d/2}} \right) + \int f \log \frac{f}{Mf};$$

the second quantity (purely kinetic information) is the curve that is seen just below the curve of the function H . When the two curves are distant from each other, the distribution is almost hydrodynamic; when they are close, the distribution is almost homogeneous. Starting from the hydrodynamic distribution, it deviates immediately, in conformance with the instability principle for the hydrodynamic approximation. We subsequently distinguish clearly between oscillations between rather hydrodynamic states, associated with a slowing down in entropy production, and the more homogeneous states; these oscillations are important, given the logarithmic nature of the diagram. Filbet, Mouhot and Pareschi [50] present other curves and attempt to explain the oscillation frequency in a certain asymptotic process.

Here the Boltzmann equation nicely reveals its double nature, relevant for both transfer of information via collisions and fluid mechanics via the transport operator. It’s often the marriage between the two aspects that proves delicate.

The relative importance of transport and collisions can be modulated by the boundary conditions; in the periodic context it comes down to the size of the box. A large box will permit important spatial variations, giving the hydrodynamic effects free rein, as in the above simulation. Nonetheless, we clearly see that even in this case, and contrary to an idea well ingrained even with specialists, the asymptotic process isn’t hydrodynamic, in the sense that the ratio between hydrodynamic entropy and total kinetic entropy doesn’t increase significantly as time passes, oscillating rather between minimum and maximum values.

We can ask ourselves what happens in a rather small box. Such a simulation is presented below.

The conclusion that we can draw from this figure is precisely opposite to our intuition, according to which the hydrodynamic effects dominate in the long run : quite the contrary, starting from a hydrodynamic situation, we quickly arrive at a situation that is almost homogeneous (visually we have the impression that at time

$\simeq 0.7$ the hydrodynamic information represents scarcely more than 1% of the total information!). The inhomogeneous effects then resume their rights (at time $t = 1$ the information is divided into parts of the same order), after which it becomes resolutely homogeneous. In this example, the homogenization has proceeded faster than convergence to equilibrium. We'll return to this figure, which has caused some perplexity, in section 8.

6.8 Two non conventional problems

I end this section by mentioning two peculiar problems linked to time's arrow in Boltzmann's equation that are perhaps just curiosities. The first is the classification of eternal solutions of Boltzmann's equation : I tried to in my doctoral thesis that, at least for the spatially homogeneous Boltzmann equation with Maxwellian molecules, there do not exist eternal solutions with finite energy. The second would be to instead look for self-similar solutions with finite energy that don't converge to Maxwellian equilibrium. For the first problem, [111] can be consulted for partial results ; the conjecture is still viable, and Bobylev and Cercignani [18] have been able to show that there doesn't exist any eternal solution having finite moments of all orders. As to the second problem, it has been resolved by the same authors [17] using Fourier transform techniques.

7 Isentropic relaxation : living with ones memories

We now consider Vlasov's equation with interaction potential W :

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f - \left(\nabla W * \int f dv \right) \cdot \nabla_v f = 0. \quad (40)$$

Unlike Boltzmann's equation, equation (40) does not impose time's arrow and remains unchanged under the action of time reversal. The constancy of entropy corresponds to a preservation of microscopic information. The solution of Vlasov's equation at time t theoretically permits reconstructing the initial condition without loss of precision, simply by solving Vlasov's equation after having reversed the velocities.

Additionally, whereas Boltzmann's equation allows but a very small number of equilibria (the Maxwellians determined by the conservation laws), Vlasov's equation allows a considerable number of them. For example, *all* the homogeneous distributions $f^0 = f^0(v)$ are stationary. There exist yet many other stationary distributions, for example the family of Bernstein-Greene-Kruskal waves [14]. For all these reasons, there is nothing *a priori* that would lead us to conjecture a well determined behavior over the long term and there is no indication at all of time's arrow. However, in 1946, L. Landau, released several years earlier from the soviet communist prisons where his frank speech had led him, suggested a very specific long term behavior for Vlasov's equation. It is based on an analysis of the linearized equation near a homogeneous equilibrium. Landau's prediction provoked a

shock and a conceptual change which still today raises lively discussions [92]; in its sequel it began to be suspected that convergence toward equilibrium is not necessarily tied to an increase in entropy. This section is devoted to a survey of the question of isentropic convergence, while emphasizing the perturbation process, which is the only one for which there are solid elements. More details can be found in my course [110].

7.1 Linearized analysis

We study Vlasov's equation near a homogeneous equilibrium $f^0(v)$. If we set $f(t, x, v) = f^0(v) + h(t, x, v)$, the equation becomes

$$\frac{\partial h}{\partial t} + v \cdot \nabla_x h + F[h] \cdot \nabla_v f^0 + F[h] \cdot \nabla_v h = 0, \quad (41)$$

where

$$F[h](t, x, v) = - \iint \nabla W(x - y) h(t, y, v) dv$$

is the force induced by the distribution h .

By neglecting the quadratic term $F[h] \cdot \nabla_v h$ in (41), we obtain the **linearized Vlasov equation** near a homogeneous equilibrium :

$$\frac{\partial h}{\partial t} + v \cdot \nabla_x h + F[h] \cdot \nabla_v f^0 = 0. \quad (42)$$

Before examining (42), we consider the case without interaction ($W = 0$), i.e. the **free transport** $\partial_t f + v \cdot \nabla_x f = 0$. This equation is solved in $\mathbb{T}_x^d \times \mathbb{R}_v^d$ by $f(t, x, v) = f_i(x - vt, v)$, where f_i is the initial distribution. We change to Fourier variables by putting

$$\tilde{g}(k, \eta) = \iint g(x, v) e^{-2i\pi k \cdot x} e^{-2i\pi \eta \cdot v} dx dv;$$

the free transport solution is then written

$$\tilde{f}(t, k, \eta) = \tilde{f}_i(k, \eta + kt). \quad (43)$$

When $k \neq 0$, this expression tends to 0 when $t \rightarrow \infty$, with speed determined by the regularity of f_i in the velocity variable (Riemann-Lebesgue principle). All these nonzero spatial modes thus relax toward 0; it's the **homogenizing** action of free transport.

Equation (42) is not so easily solved; nonetheless, if we put $\rho(t, x) = \int h(t, x, v) dv$, we then confirm that $\hat{\rho}(t, k)$ satisfies all the independent equations for distinct values of k . This remarkable *decoupling* property for the modes is the foundation for Landau's analysis. For each k we have a Volterra equation for the k -th mode :

$$\hat{\rho}(t, k) = \tilde{f}_i(k, kt) + \int_0^t K^0(k, t - \tau) \hat{\rho}(\tau, k) d\tau,$$

where

$$K^0(t, k) = -4\pi^2 \hat{W}(k) \tilde{f}^0(kt) |k|^2 t.$$

The stability of Volterra equations is a classical problem. If u satisfies $u(t) = S(t) + \int_0^t K(t - \tau) u(\tau) d\tau$, then the rate of decrease of u is dictated by the worse of two rates : the rate of decrease of S of course, and on the other hand the width of the largest band $\{0 \leq \Re \xi \leq \Lambda\}$ that doesn't intersect any solution of the equation $K^L = 1$, where K^L is the Laplace transform of K . If $\Lambda > 0$, we thus have exponential stability for the linearization.

Adapted to our context, this result leads to the **Penrose stability criterion**, for which a multidimensional version will be stated. For each $k \in \mathbb{Z}^d$, we define $f_k^0 : \mathbb{R} \rightarrow \mathbb{R}_+$ by

$$f_k^0(r) = \int_{k^\perp} f^0 \left(\frac{k}{|k|} r + z \right) dz;$$

in short, f_k^0 is the marginal of f^0 in the k -th direction. Penrose's criterion [88] requires that for each $k \in \mathbb{Z}^d$,

$$\forall \omega \in \mathbb{R}, \quad (f_k^0)'(\omega) = 0 \implies \hat{W}(k) \int \frac{(f_k^0)'(v)}{v - \omega} dv < 1.$$

If this criterion (essentially optimal) is satisfied, then there is exponential stability for the linearization : the force decreases exponentially fast, as do all inhomogeneities of the spatial density $\int h dv$.

The Penrose stability criterion has been verified in numerous situations : in the case of a Coulomb interaction when the marginals of f^0 increase to the left of 0, decrease to the right (in other words, if $(f_k^0)'(z)/z < 0$ for $z \neq 0$); in particular if f^0 is a decreasing function of $|v|$, a Gaussian for example. Again in the Coulomb case, in dimension 3 or more, the criterion is verified if f^0 is isotropic. In the case of Newtonian attraction, things are more complex : for example, for a Gaussian distribution, the stability depends on the mass and the temperature of the distribution. This reflects the celebrated Jeans instability, according to which the Vlasov equation is linearly unstable for lengths greater than

$$L_J = \sqrt{\frac{\pi T}{G \rho^0}},$$

where G is the constant of universal gravitation, ρ^0 the mass of the distribution f^0 and T its temperature. It's this instability which is responsible for the tendency of massive bodies to regroup itself in "clusters" (galaxies, clusters of galaxies, etc.).

In summary, the linearized Vlasov equation about a stable homogeneous equilibrium (in the sense of Penrose) predicts an exponential dampening of the force, in an apparently irreversible manner. This discovery brought back the problematic of time's arrow in the theory of Vlasov's equation.

The study of the linearized Vlasov equation can be found in advanced treatises on plasma physics, like [71]; however, the treatment there is systematically obscured by the use of contour integrals in the complex plane, which provide the

FIG. 3 – For a Vlasov evolution, the logarithmic ratio between the norms of the energy following the nonlinear equation to that following the linear equation, for different perturbation amplitudes.

inversion of the Laplace transform. This has been avoided in the presentation of [85, Section 3], based on the simple Fourier transform; or in the short version [110].

7.2 Nonlinear Landau dampening

The linearization effected by Landau perhaps isn't an innocent operation, and for half a century doubts have been expressed on its validity. In 1960, Backus [6] remarked that replacing $\nabla_v(f^0 + h)$ by $\nabla_v f^0$ in the force term would be conceptually simple if $\nabla_v h$ remained small throughout all time; but if we replace h by the solution of the linearized equation, we see that its velocity gradient grows linearly in time, becoming arbitrarily large. This, suggests Backus, “destroys the validity of the linear theory”. Backus's argument is questionable because $\nabla_v h$ is multiplied by $F[h]$ which one expects to see decrease exponentially; nevertheless heuristic considerations [86] suggest the failure of the linear approximation at the end of time $O(1/\sqrt{\delta})$, where δ is the size of the initial perturbation. Thereafter the curve (drawn by F. Filbet) represents the logarithm of the quotient of the energy computed using the nonlinear equation and that obtained from the linear equation, for different values of the perturbation amplitude; it is clearly seen that even for δ small, we end up in a process where the nonlinear effects cannot be neglected.

There are other reasons for distrusting the linearization. First, the eliminated term, $F[h] \cdot \nabla_v h$, is of higher degree in terms of derivatives of h with respect to velocity. Next, the linearization eliminates conservation of entropy, and favors the particular state f^0 , which voids the discussion of reversibility.

In 1997, Isichenko [63] muddied the waters by arguing that convergence toward equilibrium can't in general be more rapid than $O(1/t)$ for the nonlinear equation. It is also contradicted by Caglioti and Maffei [25], who constructed exponentially damped solutions of the nonlinear equation. Numerical simulations (see below) aren't very reliable over very long time and there is felt need for theorem.

FIG. 4 – Evolution of the norm of the force field, for electrostatic interactions (left) and gravitational interactions (right). In the electrostatic case, the rapid oscillations are called Langmuir waves.

In 2009, Mouhot and I established such a result [85]. If the interaction poten-

tial W isn't too singular, in the sense that $\hat{W}(k) = O(1/|k|^2)$ (this hypothesis allows just Coulomb and Newton interactions!), and if f^0 is an analytic homogeneous equilibrium satisfying Penrose's stability condition, then there is nonlinear dynamic stability : starting with initial data f_i , analytic and such that $\|f_i - f^0\| = O(\delta)$ when δ is very small, we have decrease of the force in $O(e^{-2\pi\lambda|t|})$ for all $\lambda < \min(\lambda_0, \lambda_i, \lambda_L)$, where λ_0 is the width of the band of complex analyticity of f^0 about \mathbb{R}_v^d , λ_i is the width of the band of complex analyticity of f_i in the variable v , and λ_L is the rate of the Landau convergence. In brief, linear damping implies nonlinear damping, with an arbitrarily small loss in rate of convergence.

The theorem also establishes the weak convergence of $f(t, \cdot)$ to an asymptotically homogeneous state $f_\infty(v)$. More precisely, the equation being invariant under time reversal, there is an asymptotic profile $f_{+\infty}$ for $t \rightarrow +\infty$, and another profile $f_{-\infty}$ for $t \rightarrow -\infty$. If Vlasov's equation is viewed as a dynamical system, there is then a remarkable behavior : the homoclinic/heteroclinic trajectories are so numerous that they fill an entire neighborhood of f^0 in analytic topology.

The nonlinear damping of Vlasov's equation is based on containment and mixing. Containment is indispensable : it is known that Landau doesn't take place in all space, even for the linearized equation [52, 53]; in our case it is automatic because the phase space is the torus. Mixing takes place because of the differential velocity phenomenon : particles with different velocities move with different velocities in phase space ; here too it is almost a tautology. An example of a non mixing system is the harmonic oscillator, where the trajectories borne by variables with different action move with constant angular velocity. Some of the other ingredients underlying the nonlinear study are :

- a reinterpretation of the problem in terms of regularity : instead of showing that there is damping, it is shown that $f(t, x, v)$ is "as regular" as the free transport solution, uniformly in time ;
- "scattering" estimates : a particle placed in an exponentially decreasing force field follows a free transport asymptotic trajectory in a sense that can be quantified precisely ;
- the stabilizing role of retarded response, *in echoes*, of the plasma : when one of the modes of the plasma is perturbed, the reaction of the other modes isn't instantaneous, but follows with a slight retardation, because the effect of the modes is compensated outside of some instants of resonance ;
- a Newton diagram that takes advantage of the fact that the linearized Vlasov equation is in some way completely integrable ; the convergence speed of this diagram allows compensating for the loss of decrease that accompanies the solution of the linearization.

All these ingredients are described in more detail in [110]. The special place of the Newton diagram and of the complete integrability form an unexpected bridge with KAM (Kolomogorov–Arnold–Moser) theory. In some way the Vlasov nonlinear Vlasov equation, in the perturbative process, inherits some of the good properties of the completely integrable linearized Vlasov equation.

From the physical point of view, information goes toward the small kinetic scales : the oscillations of the distribution function are amplified when time be-

comes large and become invisible. Lynden-Bell [75, 76] clearly understood this and used a striking formula for explaining : “a [galactic] system whose density has achieved a steady state will have information about its birth still stored in the peculiar velocities of its stars.”

These oscillations, clearly visible in the figures below, are both a nuisance from the technical point of view and the fundamental physical mechanism that produces the impression of irreversibility. We note the difference with the mechanism called *radiation*, in which the energy is emitted on macroscopic scale and goes off to infinity : here to the contrary the energy literally passes into the void...

FIG. 5 – A section of the distribution function (in relation to a homogenous equilibrium) for gravitational Landau damping, at two different times.

7.3 Sliding regularity

The nonlinear damping theorem is based on a recent reinterpretation, in terms of regularity, that deserves some comments. We begin by talking about the cascade associated with free transport, represented on the diagram below :

FIG. 6 – Evolution of energy in the space of frequencies along free transport or of a perturbation of this latter, the marks indicating the localization of energy in phase space.

This image, which is derived from formula (43), shows that the frequencies that matter vary over time : there is an overall movement toward high kinetic frequencies, and this movement is even faster than the frequency is high. More precisely, the spatial mode of frequency k oscillates in velocity with period of order $O(1/|k|t)$. The challenge of Landau damping is to show that this cascade, although distorted, is globally preserved by the effect of the interactions that *couple* the different modes.

These strong oscillations preclude any hope of obtaining bounds that are uniform in time, e.g. analytically regular in the usual sense. A key idea in [85] consists of concentrating on the Fourier modes that matter for the free transport solution, and thus to follow the cascade over the course of time. This concept is called **sliding regularity** and comes with a degradation of the regularity bounds in velocity, but simultaneously with an improvement of regularity in position, once velocity averages have been formed. Our interpretation of Landau damping is thus a transfer of regularity away from the variable v and toward the variable x , the regularity of the force improving, which implies that its amplitude dies off.

The analytic norm used in [85] is a bit complex : it has good algebraic properties that allow following the errors obtained by composition, it adapts well to the geometry of the problem, and follows free transport for measuring the sliding

regularity :

$$\|f\|_{\mathcal{Z}_\tau^{\lambda,(\mu,\gamma);p}} = \sum_{k \in \mathbb{Z}^d} \sum_{n \in \mathbb{N}^d} \frac{\lambda^n}{n!} e^{2\pi\mu|k|} (1+|k|)^\gamma \left\| (\nabla_v + 2i\pi\tau k)^n \hat{f}(k, v) \right\|_{L^p(dv)} \quad (44)$$

(here \hat{f} denotes the Fourier transform in the position variable, not in velocity). The exponent λ quantifies analytic regularity in velocity, the exponents μ and γ (by default $\gamma = 0$) quantify the regularity in position, and the parameter τ is to be taken as a time lag. Readers are referred to [85] for a study of the remarkable properties of this type of norm, and also for comparing results for more naive norms for which the nonlinear damping theorem can be stated.

The principal result of [85] consists in proving a uniform bound of type

$$\|f(t, \cdot) - f^0\|_{\mathcal{Z}_t^{\lambda,\mu;1}} = O(\delta).$$

This bound implies Landau damping, yet contains much more information : e.g. it shows that the higher spatial frequencies relax more quickly ; it also implies nonlinear orbital stability under the Penrose condition, a problem that until now has resisted all the classical methods.

7.4 Nonlinear echoes and critical regularity

The celebrated plasma wave echo experiment [77, 78] describes the interaction of two waves generated by distinct spatial perturbations. If a first perturbation is sent at the initial time with a frequency k , there ensue oscillations with kinetic frequency $|k|t$, oscillations that don't attenuate over time but rather become more and more frenzied. If now at time τ a second perturbation with frequency ℓ is made to intervene, then oscillations with kinetic frequency $|\ell|(t - \tau)$ are generated. The two oscillation trains will be invisible to each other, due to averaging, except when they have the same kinetic frequency ; this is produced in a time t such that $kt + \ell(t - \tau) = 0$, or

$$t = \frac{\ell\tau}{k + \ell}; \quad (45)$$

where it is understood that k and ℓ are collinear and opposite in direction, with $|\ell| > |k|$. In a certain sense, in the long time asymptotic, the reaction to the second perturbation τ is achieved at a time u that is strictly greater than t . This delay is critical for explaining the stability of the nonlinear Vlasov equation. To get an idea of this gain, compare the inequality $u(t) \leq A + \int_0^t \tau u(\tau) d\tau$, which implies for u a growth essentially in $O(e^{t^2})$, to the inequality $u(t) \leq A + t u(t/2)$, which implies a very slow growth in $O(t^{\log t})$.

As a caricature of the estimates for the Vlasov—Poisson equation the family of inequalities

$$\varphi_k(t) \leq a(kt) + \frac{ct}{k^2} \varphi_{k+1} \left(\frac{kt}{k+1} \right)$$

can be proposed. Here $\varphi_k(t)$ represents roughly the norm of the k -th mode of the spatial density at time t ; $a(kt)$ represents the effect of the source (we ignore

the linear term represented by a Volterra equation), the coefficient t translates the fact that the coupling occurs through the velocity gradient of f , and that the gradient grows linearly with time; $1/k^2$ is the Fourier transform of the interaction potential; we note in this regard that the interaction between modes is even more dangerous than the potential is singular; we keep only the interaction between the k -th and the $(k+1)$ -st mode; finally, the argument of the $(k+1)$ -st mode isn't t but $kt/(k+1)$, which represents a slight retardation with respect to t , as in the echoes formula. An explicit solution shows that

$$\varphi_k(t) \lesssim a(kt) \exp((ckt)^{1/3}).$$

These estimates can be adapted to the original Vlasov–Poisson equation; we then find, in the solution of the linearized equation about a non stationary solution, a loss of regularity/decay that is fractional exponential. Under good hypotheses (as strong as the Penrose condition in the gravitational case, even stronger in the Coulomb case) we find essentially $\exp((kt)^{1/3})$; in the more general case the growth remains like a fractional exponential in kt . As it remains sub-exponential, this loss of regularity can be compensated by the exponential decay issuing from the linear problem.

The loss of regularity depends essentially on the interaction, whereas the linear gain depends foremost on the regularity of the data: exponential for analytic data, polynomial for C^r data, fractional exponential for Gevrey data. The preceding discussion thus suggests that it is possible to extend the nonlinear damping theorem to Gevrey data. E.g., in the gravitational case, the critical exponent $1/3$ corresponds to a critical regularity Gevrey-3. We recall that a function is called Gevrey- ν if its successive derivatives don't grow faster than $O(n!^\nu)$. Even if losing arbitrary little over ν , it is equivalent to requiring that its Fourier transform decay as a fractional exponential $\exp(c|\xi|^{1/\nu})$.

7.5 Speculations

The nonlinear Landau damping theorem opens a large number of questions. First, its extensions to geometries other than \mathbb{T}^d is a real challenge, because then we lose the magical Fourier transform. The extension of inhomogeneous equilibria is still a distant dream; in fact, the linear stability of Bernstein–Greene–Kruskal waves is still not known!

Next, we have seen that it is known how to deal with damping under Gevrey regularity; but that the extension to lower regularities such as C^r regularity is an open problem. We have already emphasized the parallel with KAM theory, in which we know how to treat problems of class C^r ; but in KAM the loss of regularity is only polynomial, and here it is much more severe. Certain variants of the KAM problem lead to a fractional exponential loss of regularity, and then it's likewise an open problem to work with regularity lower than Gevrey. In the immediate future, the only progress in C^r regularity suggested by [85] is the possibility of proving damping on time scales much larger than the nonlinear scale ($O(1/\delta)$ instead of $O(1/\sqrt{\delta})$, see [85, section 13]; this development seems to depend on

an original conjecture concerning the optimal constant occurring in certain interpolation inequalities). In section 8 we'll again discuss strategy permitting us to conceptually bypass this limitation of very high regularity.

Whatever the optimal regularity, it is not possible to obtain a Landau damping in the space with natural energy associated with the physical conservation laws. In fact, Lin and Zeng [72] show that nonlinear Landau damping is false if there are strictly less than two derivatives, in an appropriate sense.

Finally, even if Landau damping is but a perturbative phenomenon, it should be noted that its conceptual importance remains considerable because, at the present moment, it's the only little island that we are succeeding in exploring in the ocean of open problems related to isentropic relaxation. By its discovery, Landau raised awareness that physical systems may relax without there being any irreversibility and increase in entropy. In the 1960s, Lynden-Bell [75, 76] invoked this conceptual advance for solving the galactic relaxation problem, which appears in an approximately quasi-stationary state, whereas the relaxation times associated with the galactic Vlasov equation are vastly greater than the age of the universe. Subsequently, the **violent relaxation** principle — relaxation of the force field over certain times characteristic of the dynamic — has been generally accepted by astrophysicists, without there being any theoretical explanation to promote it. We have here a major scientific challenge.

8 Weak dissipation

Between the Boltzmann model that gives preference to collisions and that of Vlasov, which completely neglects them, we find a particularly interesting compromise in Landau's (or the Fokker—Planck—Landau) model, weakly dissipative :

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f + F[f] \cdot \nabla_v f = \varepsilon Q_L(f, f), \quad (46)$$

where Q_L is the Landau operator (21).

In classical plasma physics, the coefficient ε equals $(\log \Lambda)/(2\pi\Lambda)$, where Λ is the plasma parameter, ordinarily very large (between 10^2 and 10^{30}). In a particular approach, the coefficient ε is a variation with respect to the limit of the mean field, proportional to $\log N/N$. The irreversible entropic effects modeled by the collision operator are only significantly apparent over large time $O(1/\varepsilon)$. Besides, regularizing effects are sensed instantly, even when they are weak. Interest in the study is thus multiple :

- it's a more realistic physical model than the “pure” Vlasov equation without collisions ;
- it permits quantification, as a function of the small parameter ε , of the relative velocities of the homogenization (Landau damping) and entropic convergence phenomena ;
- it permits bypassing the obstacle of Gevrey regularity that confronts the study of the non collision model.

Everything remains to be done here and I will merely sketch a long-range program.

8.1 A plausible scenario

Starting with a perturbation of homogeneous equilibrium with very rapid velocity decay, we should, in the course of temporal evolution per (46), remain close to a homogeneous process; this is in the spirit of results of Arkeryd, Esposito and Pulvirenti [5] on the weakly inhomogeneous Boltzmann equation. In the homogeneous context, the operator of the right hand side undoubtedly has the same regularization properties as a Laplacian in velocities, at least locally (the regularization properties become very weak at large velocities, but a very strong velocity decay is imposed). Assuming that this remains true in a weakly inhomogeneous context, we are left with a hypoelliptic equation that will regularize in all the variables, surely more quickly in velocity than in the position variable.

The hypoelliptic regularization in the Gevrey class has been but little studied, but using dimensional arguments we might think that in this context there is regularization in the Gevrey- $1/\alpha$ class, with velocity $O(\exp((\varepsilon t)^{-\alpha/(2-\alpha)}))$ in v , and $O(\exp((\varepsilon t)^{-3\alpha/(2-3\alpha)}))$ in x .

From another direction, in the Gevrey- $1/\alpha$ class, for $\alpha > 1/3$ we must have decay toward the homogeneous process in $O(\exp -t^\alpha)$.

By combining the two effects we obtain homogenization on a $O(\varepsilon^{-\zeta})$ time scale, with $\zeta < 1$, which is a more rapid rate than the rate of increase of the entropy in $O(\varepsilon^{-1})$.

Balancing accounts, the coefficient ζ we might hope for is disappointing, of order $8/9$. Among the steps used, the weakest link seems to be Gevrey regularization in x , which is extremely costly and perhaps not optimal since this regularity isn't necessary in linear analysis. This motivates the development of a version of the nonlinear damping theorem in low regularity in x . If this regularity occurs, the coefficient becomes much better, of order $1/6$...

8.2 Reexamining simulations

With this interpretation, we can now return to figure 6 : using a small spatial box reinforces the effect of the operator $v \cdot \nabla_x$ at the expense of the collision operator, so that we are in a weakly dissipative process (the force field is zero). Then over long time homogenization happens more quickly than entropic relaxation. This doesn't explain everything, for two reasons : first, in this figure the initial condition is strongly (and not weakly) inhomogeneous; then the Boltzmann operator doesn't regularize. Nonetheless we may well want to believe that it's the homogenization by Landau damping that primarily manifests itself in this picture, before the collisions do their work in increasing entropy. (How to describe the temporary departure from the homogeneous process seems a mystery.)

9 Metastatistics

Here I use the word "metastatistics" to talk about statistics on the distribution function, which itself has a statistical content. This section will be short because we have scarcely more than speculations on the matter.

The Hewitt–Savage theorem, a reincarnation of the Krein–Milman theorem, describes the symmetric probability measures in a large number of variables as convex combinations of chaotic measures :

$$\mu^N = \int_{P(\mathcal{Y})} \mu^{\otimes N} \Pi(d\mu),$$

where Π is a probability measure on $P(\mathcal{Y})$, the space of probability measures on the macroscopic space. In brief, a microscopic uncertainty may be decomposed on two levels : besides the chaos with fixed macroscopic profile, there is the uncertainty about the macroscopic profile, which is to say the choice of profile μ that occurs with probability measure Π .

Now is there a natural probability measure Π on the space of admissible profiles? Ideally, such a measure will be **invariant under the dynamic**. In the context of the Boltzmann equation, the question really isn't posed : only trivial measures, borne by Maxwellian equilibria, remain in contention. However, in the context of Vlasov's equation, the construction of nontrivial invariant measures is a fascinating problem. Such measures reflect the Hamiltonian nature of Vlasov's equation, studied for simplified interactions by Ambrosio and Gangbo [4].

A rather serious candidate for the status of invariant measure is Sturm's **entropic measure** [97], issuing from optimal transport theory, formally of the form $P = e^{-\beta H_\nu}$. Its complex definition has until now impeded success in proving its invariance. It shouldn't be very difficult to modify the construction by appending an energy term. Sturm measure is defined on a compact space, and there are perhaps subtleties in extending it to a kinetic context where the velocity space isn't bounded. But the worst difficulty comes no doubt from the singularity of the typical measures : it is expected that \mathbb{P} -almost every measure is totally foreign to Lebesgue measure, and is supported by a set of codimension 1. This seems to close the door to every statistical study of damping based on regularity, and increases the mystery.

Robert [91] and others have attempted to build a statistical theory of Vlasov's equation, starting from the notion of entropy, trying to predict the probable asymptotic state of dynamic evolution. The theory has gained some success, but it remains controversial. Furthermore, since the asymptotic state is obtained by a weak limit, the question arises of knowing whether an equality or inequality should be imposed on the constraints involving nonlinear functional density. For this topic readers can consult [102].

Then, this theory doesn't take into account the underlying evolution equation, postulating a certain universality with respect to the interaction. Isichenko [63] has remarked that the long-term asymptotic state, if it exists, must depend on fine details of the initial distribution and of the interaction, whereas the measures constructed by statistical theory only depend on invariants : energy, entropy, or other functionals of the form $\iint A(f) dx dv$. This objection has found substance with the counterexamples constructed in [85, section 14], which show that the transformation $f(x, v) \rightarrow f(x, -v)$ can modify the final asymptotic state, while it preserves all the known invariants of the dynamic. The objection is perhaps surmountable, because these counterexamples are constructed in analytic regularity,

i.e. in a class that must be invisible to a statistical treatment ; but these counterexamples show the subtlety of the problem, and reinforce the feeling of difficulty in the construction of invariant measures.

10 Paradoxes lost

In this last section I will review a series of more or less famous paradoxes involving time's arrow and kinetic equations, and present their commonly accepted resolutions. A certain number of them involve infinity, a classic source of paradoxes such as "Hilbert's hotel" with an infinite number of rooms, where it is always possible to find a place for a new arrival even if the hotel is already full. On our scale, this paradox reflects our incapacity to account for the appearance or disappearance of a particle in relation to the gigantic number that make up our universe. The limit $N \rightarrow \infty$ (or the asymptotic $N \gg 1$, if like Boltzmann one prefers to avoid manipulating infinities) being the basis for statistical mechanics, it's not surprising that this paradox should arise.

In all the sequel, when we speak of positive or negative time, or of pre-collisional or post-collisional configuration, we are referring to the absolute microscopic time of Newton's equations.

10.1 The Poincaré—Zermelo paradox

In 1895 Poincaré [89] cast doubt on Boltzmann's theory because it seemed to contradict the fundamental properties of dynamical systems. A little later Zermelo [113] developed this point and noted that the inexorable increase in entropy prohibited the return of the system to the initial state, which is however predicted by the recurrence theorem (within an arbitrarily small error).

The same objection can be applied to the Landau damping problem : if the distribution tends to a homogeneous equilibrium, it will never return close to its initial state.

From the mathematical point of view, this reasoning clearly does not apply, since the Boltzmann equation, like that of Landau, involves an infinite number of degrees of freedom ; it's only for a fixed number of particles that the recurrence theorem applies. From the physical point of view, the answer is a bit more subtle. On the one hand, the recurrence time diverges when the number N of particles tends to infinity, and this divergence is likely monstrously rapid ! For a system of macroscopic size, albeit small, the recurrence theorem simply never applies, for it involves times well greater than the age of the universe. On the other hand, the validity of the Boltzmann and Landau equations isn't eternal : for N fixed, the quality of the approximation will degrade with time, because chaos (simple or pre-collisional) is preserved only approximately. By the time that Poincaré recurrence takes place, the Boltzmann equation will have long ceased to be valid !!

10.2 Microscopic conservation of the volume

Poincaré's recurrence theorem is based on conservation of the volume in microscopic phase space (preservation of Liouville measure). The entropy is a direct function of the volume of microscopic admissible states ; how can it increase if the volume of microscopic states is constant ?

The answer to this question may seem surprising : it can be argued that the increase of entropy doesn't occur *despite* the conservation of microscopic volume, but *because* of this conservation ; more precisely, it's what keeps entropy from decreasing. In fact, let us start at the initial time from all the typical configurations associated with a distribution f_i . After a time t , these typical configurations have evolved and are now associated with a distribution f_t , the transition from f_i to f_t being governed by the Boltzmann equation. The typical configurations associated with f_t are thus at least as numerous as the typical configurations associated with f_i , which clearly means that entropy can't decrease.

In a microscopic irreversible model, we will typically have a contraction of microscopic phase space, linked to a dissipative phenomenon. The preceding argument doesn't apply then, and one can imagine that the entropy decreases, at least for certain initial data. It is indeed what happens, for example, in models of granular gases undergoing inelastic collisions.

10.3 Spontaneous appearance of time's arrow

How, starting from a microscopic equation that doesn't favor any time direction, can the Boltzmann equation predict an inexorable evolution toward positive times ?

The response is simple : there isn't any inexorable evolution toward positive times, and the double direction of time is preserved. There is simply a particular choice of the initial data (instant of preparation of the experiment), that has fixed a particular time, say $t = 0$. Starting from there, one has a double arrow of time ; entropy increases for positive times, and decreases for negative times.

10.4 Loschmidt's paradox

Loschmidt's paradox [74] formalizes the apparent contradiction that exists of a reversible microscopic dynamic and an irreversible evolution of the entropy. Let us suppose that we start from a given initial configuration and that at time t we stop the gas and reverse the velocities of all the particles. This operation doesn't change the entropy, and starting from this new initial data we can let the dynamic act anew. By microscopic reversibility, at the end of time $2t$ we will have returned to the point of departure ; but the entropy will not have ceased to increase, whence the contradiction.

This paradox can be resolved in several ways, all of which come down to the same finding : *the degradation of the notion of chaos* between the initial time and time $t > 0$. On the mathematical level it is only known how to prove the weak convergence of $\mu_t^{1;N}$ to $f(t, \cdot)$ as $N \rightarrow \infty$, whereas the convergence is supposed

uniform at the initial time. In fact, it is conjectured that the data (μ_t^N) satisfies the property (still to be defined...) of *pre-collisional chaos*, whereas the initial data is supposed to satisfy a complete chaos property. When the velocities are reversed, the hypothesis of pre-collisional chaos is transformed into post-collisional chaos, and the relevant equation is no longer the Boltzmann equation, but the “inverse” Boltzmann equation, in which a negative sign is placed before the collision operator. Entropy then increases toward negative times and no longer toward positive times, and all contradiction disappears.

To state matters in a more informal way : at the initial time the particles are all strangers to one another. After a time t , the particles that have just encountered each other know each other still, while those who are about to encounter each other don't know each other : the particles have a memory of the past and not of the future. When the velocities are reversed, the particles have a memory of the future and not of the past, and time begins to flow backwards !

Legend has it that Boltzmann, confronted with the velocity reversal paradox, responded : “Go ahead and reverse them !” Behind the jest is hidden a profound observation : reversal of velocities is an operation that is inaccessible to us because it requires microscopic knowledge of the system ; and the notion of entropy emerges precisely from what we can only act upon macroscopically. Beginning in the 1950s, spin echo experiments allowed us to see the paradox from another angle [10].

10.5 Nonuniversal validity of Boltzmann's equation

This paradox is a variant of the preceding. Having understood that Boltzmann's equation doesn't apply after reversal of velocities, we will exploit this fact to put Herr Boltzmann in default. We redo the preceding experiment and choose as initial data the distribution obtained after reversal of velocities at time t . We let time act, and the relevant equation certainly isn't Boltzmann's equation.

This paradox effectively shows that there are microscopic configurations that don't lead to Boltzmann's equation. Nevertheless, and it's thus that Boltzmann argued, these configurations are rare : precisely, they cause the appearance of correlations between pre-collisional velocities. This isn't rarer than correlations between post-collisional velocities, but it's rarer than not having correlations at all ! The Boltzmann equation is approximately true if we depart from a typical configuration, which is to say drawn according to a “strongly chaotic” law, but it doesn't hold for all initial configurations. Once these grand principles are stated, the quantitative work remains to be done.

10.6 Arbitrary Boltzmann procedure

To establish the Boltzmann equation, the encounter probabilities of particles are expressed in terms of the pre-collisional probabilities, which are arbitrary. If instead post-conditional probabilities had been used, a different equation would have been obtained, with a negative sign before the collision operator ! Why then have confidence in Boltzmann ?

The answer is still the same, of course, and depends on which side of the origin one is placed : for negative times, these are pre-collisional probabilities that are almost factored, whereas for positive times, these are post-collisional probabilities.

10.7 Maxwell's demon

Maxwell conceived a thought experiment in which a malicious demon positions himself in a box with two compartments and adroitly manipulates a small valve so that there is a flow of balls going from the right compartment toward the left but not the other way. The system thus evolves toward increased order, and the entropy decreases.

Of course, this can't be considered an objection to the law of increasing entropy, and the experiment is intended to make us think : first, the demon should be part of the model and himself subjected to reversible mechanical laws, taking into account the energy needed for recognizing that a particle is approaching and for evaluating its velocity, for the mental work done, etc. If a complete account is made, we will have obtained, we can be sure, the second law of thermodynamics.

We note in this regard that recently experiments with Maxwell's demon have been realized with granular gases : as I myself saw with stupefaction in a film made of an experiment, initially there is a receptacle with two vertically separated compartments and an opening above that allows communication, the two compartments are filled with inelastic particles in approximately equal number, the whole thing is agitated automatically, and little by little one of the compartments is emptied in favor of the other. An underlying principle is that in the fuller compartment the abundance of collisions results in cooling by dissipation of energy ; and the particles jump less high, rendering it more and more difficult for them to leave the full compartment. We find again on this occasion the principle — already mentioned — according to which a dissipative (irreversible) dynamic does not necessarily lead to an increase of entropy, but to the opposite.

10.8 Convergence and reversibility

This paradox is a variant of the Loschmidt paradox ; it applies both to the theme of Boltzmann entropic relaxation and to nonlinear Landau damping : how can there be convergence when $t \rightarrow +\infty$ if there is reversibility of the dynamic ?? The answer is of childish simplicity : there is also convergence when $t \rightarrow -\infty$. For Vlasov, this was accomplished with the same equation, and we thus have a phenomenon of generalized homoclinic/heteroclinic. For Boltzmann, the equation changes according as to whether times are considered which are prior or subsequent to the chaotic data.

10.9 Stability and reversibility

This paradox is more subtle and applies to nonlinear Landau damping : asymptotic stability and reversibility of the dynamic automatically imply an instability, which seems contradictory.

We detail the argument. If we have stability in time $t \rightarrow +\infty$, let $f_\infty(v)$ be an asymptotically stable profile, which we assume to be even. We take a solution $\tilde{f}(t, x, v)$, inhomogeneous, which converges toward $f_\infty(v)$. We then choose as initial data $f(T, x, -v)$ with T very large; we thus have data arbitrarily close to $f_\infty(-v) = f_\infty(v)$, and which brings us back after time T to the data $\tilde{f}(0, x, v)$, rather removed from $f_\infty(v)$. In other words, the distribution f_∞ is *unstable*. How is this compatible with stability??

The answer, as explained e.g. in [25], lies in the topology: in the theorem of asymptotic stability (nonlinear Landau damping), the convergence over large time occurs in the sense of the weak topology, with frenetic oscillations in the velocity distribution, which is compensated locally. When we say that a distribution f^0 is stable, that means that if we depart close to f^0 in the sense of the strong topology (e.g. analytic or Gevrey), then we remain close to f^0 in the sense of the weak topology. The asymptotic stability combined with the reversibility thus imply *instability in the sense of the weak topology*, which is perfectly compatible with stability in the sense of the strong topology.

10.10 Conservative relaxation

This problem is of a rather general nature. Vlasov's equation comes with a preservation of the amount of microscopic uncertainty (conservation of entropy). Moreover the distribution at time $t > 0$ allows reconstructing exactly the distribution at time $t = 0$: it suffices to solve Vlasov's equation after reversal of the velocities. We can say that Vlasov's equation forgets nothing; now, convergence consists precisely in forgetting the episodes of the dynamic evolution!

The answer again lies in the weak convergence and the oscillations. Information will be lodged in these oscillations, information which is invisible because in practice we never measure the complete distribution function, but averages of this distribution function (recall the quote of Lynden-Bell reproduced at the end of section 7.2). Every observable will converge toward its limit value, and there will be a "forgetfulness". The force field, obtained as mean of the kinetic distribution, converges toward 0 without this being contradictory to preservation of information: the information leaves the spatial variables so as to go into the kinetic variables. In particular, the spatial entropy $\int \rho \log \rho$ (where $\rho = \int f dv$) tends toward 0, whereas the total kinetic entropy $\int f \log f$ is conserved (but doesn't converge! information is conserved for all time, but because of the weak convergence there is a loss of information in the passage to the limit $t \rightarrow \infty$).

Similarly, in nonlinear Landau damping, the energy of interaction — which is $\int W(x - y) \rho(x) \rho(y) dx dy$ — tends toward 0, and it is converted into kinetic energy (which can increase or decrease as a function of the interaction).

10.11 The echo experiment

In this famous experiment [77, 78] a plasma, prepared in a state of equilibrium, is excited at the initial time by a spatial frequency impulse k . At the end of a time τ , after relaxation of the plasma, it is excited anew by a spatial frequency ℓ ,

collinear and in the direction opposite to k , and of greater amplitude. We then wait and observe spontaneous response from the electric field of the plasma, called **echo**, which is produced with spatial frequency $k + \ell$ and around the time $t_e = (|\ell|/|k + \ell|)\tau$.

This experiment shows that the kinetic distribution of the plasma has kept track of past impulses : even if the force field has died off to the point of becoming negligible, the kinetic oscillations of the distribution remain present and evolve over the course of time. The first impulse subsists in the form of very rapid oscillations of period $(|k|t)^{-1}$, the second in the form of oscillations of period $(|\ell|(t - \tau))^{-1}$. A calculation, found e.g. in [110, section 7.3] shows that the distribution continues to oscillate rapidly in velocity, and the associated force remains negligible, up until the two trains of oscillations compensate almost exactly, which is manifested by an echo.

Références

- [1] Alexandre, R., Desvillettes, L., Villani, C. and Wennberg, B. : Entropy dissipation and long-range interactions. *Arch. Rat. Mech. Anal.* **152**, 327—355 (2000).
- [2] Alexandre, R. and Villani, C. : On the Boltzmann equation for long-range interaction and the Landau approximation in plasma physics. *Comm. Pure Appl. Math.* **55**, 1, 30—70 (2002).
- [3] Alexandre, R. and Villani, C. : On the Landau approximation in plasma physics. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **21**, 1, 61—95 (2004).
- [4] Ambrosio, L. and Gangbo, W. : Hamiltonian ODE's in the Wasserstein space of probability measures. *Comm. Pure Appl. Math.* **51**, 18—53 (2007).
- [5] Arkeryd, L., Esposito, R. and Pulvirenti, M. : The Boltzmann equation for weakly inhomogeneous data. *Comm. Math. Phys.* **111**, 3, 393—407 (1987).
- [6] Backus, G. : Linearized plasma oscillations in arbitrary electron distributions. *J. Math. Phys.* **1**, 178—191, 559 (1960).
- [7] Balescu, R. : Irreversible processes in ionized gases. *Phys. Fluids* **3**, 52—63 (1960).
- [8] Balescu, R. *Statistical Mechanics of Charged Particles*. Wiley-Interscience, New York, 1963.
- [9] Balian, R. : Entropy, a protean concept. *Poincaré Seminar 2*, 119—145 (2003). Birkhäuser Verlag, Basel.
- [10] Balian, R. : Information in statistical physics. *Studies in Hist. and Phil. of Mod. Phys.* **36**, 323—353 (2005).
- [11] Baranger, C. and Mouhot, C. : Explicit spectral gap estimates for the linearized Boltzmann and Landau operators with hard potentials. *Rev. Matem. Iberoam.* **21**, 819—841 (2005).

- [12] Batt, J. and Rein, G. : Global classical solutions of the periodic Vlasov-Poisson system in three dimensions. *C. R. Acad. Sci. Paris Sér. I Math.* **313**, 6, 411—416 (1991).
- [13] Ben Arous, G. and Zeitouni, O. : Increasing propagation of chaos for mean field models. *Ann. Inst. H. Poincaré Probab. Statist.* **35**, 1, 85—102 (1999).
- [14] Bernstein, I.B., Greene, J.M. and Kruskal, M.D. : Exact nonlinear plasma oscillations. *Phys. Rev.* **108**, 3, 546—550 (1957).
- [15] Binney, J. and Tremaine, S. : *Galactic Dynamics*, 2nd ed. Princeton Series in Astrophysics. Princeton University Press, 2008.
- [16] Bobylev, A.V. and Cercignani, C. : On the rate of entropy production for the Boltzmann equation. *J. Statist. Phys.* **94**, 3-4, 603—618 (1999).
- [17] Bobylev, A.V. and Cercignani, C. : Exact eternal solutions of the Boltzmann equation. *J. Statist. Phys.* **106**, 5—6, 1019—1038 (2002).
- [18] Bobylev, A.V. and Cercignani, C. : Self-similar solutions of the Boltzmann equation and their applications. *J. Statist. Phys.* **106**, 5—6, 1039—1071 (2002).
- [19] Bogoliubov, N.N. : Problems of a dynamical theory in statistical physics. *Studies in Statistical mechanics* 1, 1—118 (1962). Translation from the 1946 Russian original.
- [20] Bolley, F., Guillin, A. and Villani, C. : Quantitative concentration inequalities for empirical measures on non-compact spaces. *Probab. Theory Related Fields* **137**, 3-4, 541—593 (2007).
- [21] Boltzmann, L. : Weitere Studien über das Wärmegleichgewicht unter Gas-molekülen. *Sitzungsberichte Akad. Wiss.* **66** (1872), 275—370. Traduction : Further studies on the thermal equilibrium of gas molecules, in *Kinetic Theory* **2**, 88—174, Éd. S.G. Brush, Pergamon, Oxford (1966).
- [22] Boltzmann, L. : *Lectures on gas theory*. University of California Press, Berkeley, 1964. English translation by Stephen G. Brush. Reprint of the 1896—1898 edition. Dover Publications, 1995.
- [23] Bouchut, F. : Introduction à la théorie mathématique des équations cinétiques. In Session “L’Etat de la Recherche” de la SMF (1998). “Kinetic equations and asymptotic theory”, F. Bouchut, F. Golse, M. Pulvirenti, coll. “Series in Appl. Math.”, Gauthier-Villars, 2000.
- [24] Braun, W. and Hepp, K. : The Vlasov dynamics and its fluctuations in the $1/N$ limit of interacting classical particles. *Commun. Math. Phys.* **56**, 125—146 (1977).
- [25] Caglioti, E. and Maffei, C. : Time asymptotics for solutions of Vlasov—Poisson equation in a circle. *J. Statist. Phys.* **92**, 1—2, 301—323 (1998).
- [26] Carleman, T. : Sur la théorie de l’équation intégrodifférentielle de Boltzmann. *Acta Math.* **60**, 369—424 (1932).
- [27] Carleman, T. : *Problèmes Mathématiques dans la Théorie Cinétique des Gaz*. Almqvist & Wiksell, 1957.

- [28] Carlen, E.A. and Carvalho, M.C. : Strict entropy production bounds and stability of the rate of convergence to equilibrium for the Boltzmann equation. *J. Statist. Phys.* **67**, 3—4, 575—608 (1992).
- [29] Carlen, E.A., Carvalho, M. C. and Gabetta, E. : Central limit theorem for Maxwellian molecules and truncation of the Wild expansion. *Comm. Pure Appl. Math.* **53**, 370—397 (2000).
- [30] Carlen, E.A., Carvalho, M. C. and Gabetta, E. : On the relation between rates of relaxation and convergence of Wild sums for solutions of the Kac equation. *J. Funct. Anal.* **220**, 362—387 (2005).
- [31] Carlen, E.A., Carvalho, M.C. and Loss, M. : Determination of the spectral gap for Kac's master equation and related stochastic evolution. *Acta Math.* **191**, 1—54 (2003).
- [32] Carlen, E.A., Carvalho, M.C., Le Roux, J., Loss, M. and Villani, C. Entropy and chaos in the Kac model. *Kinet. Relat. Models* **3**, 1, 85—122 (2010).
- [33] Carlen, E.A., Gabetta, E. and Toscani, G. : Propagation of smoothness and the rate of exponential convergence to equilibrium for a spatially homogeneous Maxwellian gas. *Comm. Math. Phys.* **199**, 3, 521—546 (1999).
- [34] Carlen, E.A. and Lu, M. : Fast and slow convergence to equilibrium Maxwellian molecules via Wild sums *J. Stat. Phys.* **112**, 1—2, 59—134 (2003).
- [35] Carleson, L. : Some analytic problems related to statistical mechanics. In *Euclidean Harmonic Analysis*, J.J. Benedetto, Éd. (Univ. of Maryland, 1979), *Lect. Notes Math.* 779, 5—45.
- [36] Cercignani, C. : On the Boltzmann equation with cutoff potentials. *Phys. Fluids* **10**, 2097 (1967).
- [37] Cercignani, C. : On the Boltzmann equation for rigid spheres. *Transport Theory Statist. Phys.* **2**, 3, 211—225 (1972).
- [38] Cercignani, C. : *The Boltzmann equation and its applications*. Springer-Verlag, New York, 1988.
- [39] Cercignani, C. : *Rarefied gas dynamics*. Cambridge University Press, Cambridge, 2000. From basic concepts to actual calculations.
- [40] Cercignani, C., Illner, R. and Pulvirenti, M. : *The Mathematical Theory of Dilute Gases*. Springer, New York, 1994.
- [41] Cover, T. M. and Thomas, J. A. : *Elements of information theory*. John Wiley & Sons Inc., New York, 1991. A Wiley-Interscience Publication.
- [42] Delcroix, J.-L. and Bers, A. : *Physique des plasmas (2 Vol.)*. InterEditions/CNRS Éditions, 1994.
- [43] Dembo, A. and Zeitouni, O. : *Large deviations techniques and applications*, 2nd ed. Springer-Verlag, New York, 1998.
- [44] Desvillettes, L. : Entropy dissipation rate and convergence in kinetic equations. *Comm. Math. Phys.* **123**, 4, 687—702 (1989).

- [45] Desvillettes, L. and Villani, C. : On the spatially homogeneous Landau equation for hard potentials. Part I : existence, uniqueness and smoothness. *Comm. Partial Differential Equations* **25**, 1-2, 179—259 (2000).
- [46] Desvillettes, L. and Villani, C. : On the trend to global equilibrium for spatially inhomogeneous kinetic systems : the Boltzmann equation. *Invent. Math.* **159**, 2, 245—316 (2005).
- [47] DiPerna, R. and Lions, P.-L. : On the Cauchy problem for the Boltzmann equation : Global existence and weak stability. *Ann. of Math. (2)* **130**, 312—366 (1989).
- [48] Dobrušin, R.L. : Vlasov equations. *Funktsional. Anal. i Prilozhen.* **13**, 2, 48—58, 96 (1979).
- [49] Dudley, R. M. : *Real analysis and probability*, vol. 74 of Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 2002. Revised edition of the 1989 original.
- [50] Filbet, F., Mouhot, C. and Pareschi, L. : Solving the Boltzmann equation in $N \log_2 N$. *SIAM J. Sci. Comput.* **28**, 3, 1029—1053 (2006).
- [51] Glassey, R.T. : *The Cauchy problem in kinetic theory*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1996.
- [52] Glassey, R. and Schaeffer, J. : Time decay for solutions to the linearized Vlasov equation. *Transport Theory Statist. Phys.* **23**, 4, 411—453 (1994).
- [53] Glassey, R. and Schaeffer, J. : On time decay rates in Landau damping. *Comm. Partial Differential Equations* **20**, 3—4, 647—676 (1995).
- [54] Golse, F. : From kinetic to macroscopic models. In Session “L’Etat de la Recherche” de la SMF (1998), “Kinetic equations and asymptotic theory”, F. Bouchut, F. Golse, M. Pulvirenti, coll. “Series in Appl. Math.”, Gauthier-Villars, 2000.
- [55] Golse, F. and Saint-Raymond, L. : The Navier—Stokes limit of the Boltzmann equation for bounded collision kernels. *Invent. Math.* **155**, 1, 81—161 (2004).
- [56] Grad, H. : On Boltzmann’s H -theorem. *J. Soc. Indust. Appl. Math.* **13**, 1, 259—277 (1965).
- [57] Grad, H. : Principles of the kinetic theory of gases. In Flügge’s *Handbuch des Physik*, vol. XII. Springer-Verlag, 1958, pp. 205—294.
- [58] Gualdani, M.P., Mischler, S. and Mouhot, C. : Factorization for non-symmetric operators and exponential H -theorem. Preprint, available online at <http://hal.archives-ouvertes.fr/ccsd-00495786>
- [59] Guo, Y. : The Landau equation in a periodic box. *Commun. Math. Phys.* **231**, 391—434 (2002).
- [60] Guo, Y. and Strain, R. M. : Exponential decay for soft potentials near Maxwellian. *Arch. Ration. Mech. Anal.* **187**, 2, 287—339 (2008).

- [61] Illner, R. and Pulvirenti, M. : Global validity of the Boltzmann equation for a two-dimensional rare gas in vacuum. *Comm. Math. Phys.* **105**, 2, 189—203 (1986). “Erratum and improved result”, *Comm. Math. Phys.* 121, 1, 143—146 (1989).
- [62] Illner, R. and Pulvirenti, M. : A derivation of the BBGKY-hierarchy for hard sphere particle systems. *Transport Theory Statist. Phys.* **16**, 7, 997—1012 (1987).
- [63] Isichenko, M. : Nonlinear Landau damping in collisionless plasma and inviscid fluid. *Phys. Rev. Lett.* **78**, 12, 2369—2372 (1997).
- [64] Jabin, P.-E. and Hauray, M. : N -particles approximation of the Vlasov equations with singular potential. *Arch. Ration. Mech. Anal.* **183**, 3, 489—524 (2007).
- [65] Janvresse, E. : Spectral gap for Kac’s model of Boltzmann equation. *Ann. Probab.* **29**, 1, 288—304 (2001).
- [66] Kac, M. : Foundations of kinetic theory. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954—1955*, vol. III (Berkeley and Los Angeles, 1956), University of California Press, pp. 171—197.
- [67] Landau, L.D. : Die kinetische Gleichung für den Fall Coulombscher Wechselwirkung. *Phys. Z. Sowjet* **10** (1936), 154. English translation : The transport equation in the case of Coulomb interactions, in *Collected papers of L.D. Landau*, edited and with an introduction by D. ter Haar, Pergamon Press, 1965, pp.163—170.
- [68] Landau, L.D. : On the vibration of the electronic plasma. *J. Phys. USSR* **10**, 25 (1946). English translation in *JETP* **16**, 574. Reprinted in *Collected papers of L.D. Landau*, edited with an introduction by D. ter Haar, Pergamon Press, 1965, pp. 445—460 ; and in *Men of Physics : L.D. Landau*, Vol. 2, Pergamon Press, D. ter Haar, ed. (1965).
- [69] Lanford, O.E. : Time evolution of large classical systems. In *Dynamical systems, theory and applications (Recontres, Battelle Res. Inst., Seattle, Wash., 1974)*. Springer, Berlin, 1975, pp. 1—111. *Lecture Notes in Phys.*, Vol. 38.
- [70] Lenard, A. : On Bogoliubov’s kinetic equation for a spatially homogeneous plasma. *Ann. Physics* **10**, 390—400 (1960).
- [71] Lifshitz, E. M. and Pitaevskiĭ, L. P. : *Course of theoretical physics [“Landau-Lifshits”]*. Vol. 10. Pergamon Press, Oxford, 1981. Translated from the Russian by J. B. Sykes and R. N. Franklin.
- [72] Lin, Z. and Zeng, C. : BGK waves and nonlinear Landau damping. Preprint, 2010.
- [73] Liverani, C. and Olla, S. : Toward the Fourier law for a weakly interacting anharmonic crystal. Preprint, 2010.
- [74] Loschmidt, J. : Über den Zustand des Wärmegleichgewichtes eines Systems von Körpern mit Rücksicht auf die Schwerkraft. *Wien. Ber.* **73**, 128 (1876).

- [75] Lynden-Bell, D. : The stability and vibrations of a gas of stars. *Mon. Not. R. astr. Soc.* **124**, 4, 279—296 (1962).
- [76] Lynden-Bell, D. : Statistical mechanics of violent relaxation in stellar systems. *Mon. Not. R. astr. Soc.* **136**, 101—121 (1967).
- [77] Malmberg, J. and Wharton, C. : Collisionless damping of electrostatic plasma waves. *Phys. Rev. Lett.* **13**, 6, 184—186 (1964).
- [78] Malmberg, J., Wharton, C., Gould, R. and O’Neil, T. : Plasma wave echo experiment. *Phys. Rev. Letters* **20**, 3, 95—97 (1968).
- [79] Maslen, D. : The eigenvalues of Kac’s master equation. *Math. Zeit.* **243**, 291—331 (2003).
- [80] Maxwell, J.C. : On the dynamical theory of gases. *Philos. Trans. Roy. Soc. London Ser. A* **157**, 49—88 (1867).
- [81] McKean, H.J. : Speed of approach to equilibrium for Kac’s caricature of a Maxwellian gas. *Arch. Rational Mech. Anal.* **21**, 343—367 (1966).
- [82] Mouhot, C. : Rate of convergence to equilibrium for the spatially homogeneous Boltzmann equation with hard potentials. *Comm. Math. Phys.* **261**, 629—672 (2006).
- [83] Mouhot, C. and Strain, R.M. : Spectral gap and coercivity estimates for linearized Boltzmann collision operators without angular cutoff. *J. Math. Pures Appl.* **87**, 5, 515—535 (2007).
- [84] Mouhot, C. and Villani, C. : Regularity theory for the spatially homogeneous Boltzmann equation with cut-off. *Arch. Rational Mech. Anal.* **173**, 2, 169—212 (2004).
- [85] Mouhot, C. and Villani, C. : On Landau damping. To appear in *Acta Math.* Available on line at arXiv :0904.2760
- [86] O’Neil, T. : Collisionless damping of nonlinear plasma oscillations. *Phys. Fluids* **8**, 12, 2255—2262 (1965).
- [87] Neunzert, H. : An introduction to the nonlinear Boltzmann—Vlasov equation. In *Kinetic theories and the Boltzmann equation*, C. Cercignani, Ed., vol. 1048 of *Lecture Notes in Math.*, Springer, Berlin, Heidelberg, 1984, pp. 60—110.
- [88] Penrose, O. : Electrostatic instability of a non-Maxwellian plasma. *Phys. Fluids* **3**, 258—265 (1960).
- [89] Poincaré, H. : Le mécanisme et l’expérience. *Revue de Métaphysique et de Morale* **I**, 534—537 (1893).
- [90] Pulvirenti, M. : From particle to transport equations. In Session “L’Etat de la Recherche” de la SMF (1998). “Kinetic equations and asymptotic theory”, F. Bouchut, F. Golse, M. Pulvirenti, coll. “Series in Appl. Math.”, Gauthier-Villars, 2000.
- [91] Robert, R. : Statistical mechanics and hydrodynamical turbulence. In *Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Zürich, 1994)* (Basel, 1995), Birkhäuser, pp. 1523—1531.

- [92] Ryutov, D. D. : Landau damping : half a century with the great discovery. *Plasma Phys. Control. Fusion* **41**, A1—A12 (1999).
- [93] Saint-Raymond, L. : *Hydrodynamic limits of the Boltzmann equation*. Lectures at SISSA, Trieste (2006), Springer Monograph, to appear.
- [94] Shannon, C.E. and Weaver, W. : *The mathematical theory of communication*. University of Illinois Press, Urbana, Illinois, 1949.
- [95] Spohn, H. : *Large scale dynamics of interacting particles*. Texts and Monographs in Physics. Springer-Verlag, Berlin, 1991.
- [96] Strain, R.M. : On the linearized Balescu—Lenard equation. *Comm. Partial Differential Equations* **32**, 1551—1586 (2007).
- [97] Sturm, K.-Th. : Entropic measure on multidimensional spaces. In *Stochastic analysis, random fields and applications VI*, Dalang, R., Dozzi, M., and Russo, F., Eds. Progress in Probability, Birkhäuser.
- [98] Sznitman, A.-S. : Equations de type de Boltzmann, spatialement homogènes. *Z. Wahrsch. Verw. Gebiete* **66**, 559—562 (1984).
- [99] Sznitman, A.-S. : *Topics in propagation of chaos*. In *École d’Été de Probabilités de Saint-Flour XIX* (1989), Lect. Notes in Math. 1464 (1991), Springer, Berlin, pp. 165—251.
- [100] Tanaka, H. : An inequality for a functional of probability distributions and its application to Kac’s one-dimensional model of a Maxwellian gas. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **27**, 47—52 (1973).
- [101] Toscani, G. and Villani, C. : Sharp entropy dissipation bounds and explicit rate of trend to equilibrium for the spatially homogeneous Boltzmann equation. *Comm. Math. Phys.* **203**, 3, 667—706 (1999).
- [102] Turkington, B. : Statistical equilibrium measures and coherent states in two-dimensional turbulence. *Comm. Pure Appl. Math.* **52**, 7, 781—809 (1999).
- [103] Villani, C. : *A review of mathematical topics in collisional kinetic theory*. In *Handbook of Mathematical Fluid Dynamics I*, 71—305, S. Friedlander et D. Serre, Édts, North-Holland, Amsterdam, 2002.
- [104] Villani, C. Cercignani’s conjecture is sometimes true and always almost true. *Commun. Math. Phys.* **234**, 455—490 (2003).
- [105] Villani, C. : Mathematics of granular materials. *J. Stat. Phys.* **124**, no. 2-4, 781—822 (2006).
- [106] Villani, C. *H-Theorem and beyond : Boltzmann’s entropy in today’s mathematics*. Conference proceedings “Boltzmann’s legacy”, Erwin-Schrödinger Institute (Vienna, July 2007).
- [107] Villani, C. : Hypocoercive diffusion operators. Text of my lecture International Congress of Mathematicians (Madrid, August 2006).
- [108] Villani, C. : Entropy production and convergence to equilibrium. Notes pour une série de cours à l’Institut Henri Poincaré, Paris (automne 2001). In *Entropy methods for the Boltzmann equation*, 1—70, Lecture Notes in Math., Vol. 1916, Springer, Berlin, 2008.

- [109] Villani, C. : *Hypocoercivity*. Mem. Amer. Math. Soc. **202**, no. 950 (2009).
- [110] Villani, C. : *Landau damping*. Lecture notes, CEMRACS 2010. Available at www.math.univ-lyon1.fr/~villani
- [111] Villani, C. : Is there any backward solution of the Boltzmann equation? Unpublished work, available at www.math.univ-lyon1.fr/~villani
- [112] Vlasov, A.A. : On the oscillation properties of an electron gas. Zh. Èksper. Teoret. Fiz. **8**, 291—318 (1938).
- [113] Zermelo, E. : Über einen Satz der Dynamik und die mechanische Wärmetheorie. Annalen der Physik **54**, 485—494 (1896).