H-Theorem and beyond: Boltzmann's entropy in today's mathematics

Cédric Villani

Abstract. Some of the objects introduced by Boltzmann, entropy in the first place, have proven to be of great inspiration in mathematics, and not only in problems related to physics. I will describe, in a slightly informal style, a few striking examples of the beauty and power of the notions cooked up by Boltzmann.

Introduction

In spite of his proudly unpolished writing and his distrust for some basic mathematical concepts (such as infinity or continuum), Ludwig Boltzmann has made major contributions to mathematics. In particular, the notion of **entropy**, and the famous H-Theorem, have been tremendous sources of inspiration to (a) understand our (physical) world in mathematical terms; (b) attack many (purely) mathematical problems. My purpose in this lecture is to develop and illustrate these claims. Before going on, I shall provide some quotations about the influence of Boltzmann on the science of his time. There is no need to recall how much Einstein, for instance, praised Boltzmann's work in statistical mechanics. But to better illustrate my point, I will quote some great *mathematicians* talking about Boltzmann:

All of us younger mathematicians stood by Boltzmann's side.

Arnold Sommerfeld (about a debate taking place in 1895)

Boltzmann's work on the principles of mechanics suggests the problem of developing mathematically the limiting processes (...) which lead from the atomistic view to the laws of motion of continua.

David Hilbert (1900)

Boltzmann summarized most (but not all) of his work in a two volume treatise Vorlesungen über Gastheorie. *This is one of the greatest books in the history of exact sciences and the reader is strongly advised to consult it.*

Mark Kac (1959)

This lecture is an extended version of a conference which I gave at various places (Vienna, Münich, Leipzig, Brisbane, Pisa) on the occasion of the hundredth anniversary of the death of Boltzmann. It is a pleasure to thank the organizers of these events for their excellent work, in particular Jakob Yngvason, Herbert Spohn, Manfred Salmhofer, Bevan Thompson, and Luigi Ambrosio. Thanks are also due to Wolfgang Wagner for his comments. The style will be somewhat informal and I will often cheat for the sake of pedagogy; more precise information and rigorous statements can be retrieved from the bibliography.

1 1872: Boltzmann's H-Theorem

The so-called Boltzmann equation models the dynamics of rarefied gases via a time-dependent density $f(x, v)$ of particles in phase space (x stands for position, v for velocity):

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

Here $v \cdot \nabla_x$ is the transport operator, while Q is the bilinear collision operator:

$$Q(f,f) = \int_{\mathbb{R}^3_{v_*} \times S^2} B \left[f(v')f(v'_*) - f(v)f(v_*) \right] dv_*\, d\sigma;$$

and $B = B(v - v_*, \sigma) \geq 0$ is the collision kernel. The notation v', v'_* denotes pre-collisional velocities, while v, v_* stand for post-collisional velocities (or the reverse, it does not matter in this case). I refer to my review text in the *Handbook of Mathematical Fluid Dynamics* [19] for a precise discussion, and a survey of the mathematical theory of this equation.

According to this model, under ad hoc boundary conditions, **the entropy S is nondecreasing** in time. That is, if one defines

$$S(f) = -H(f) := -\int_{\Omega \times \mathbb{R}^3_v} f(x,v) \log f(x,v)\, dv\, dx,$$

then the time-derivative of S is always nonnegative.

Boltzmann's theorem is more precise than just that: It states that a (strictly) positive amount of entropy is produced at time t, unless the density f is **locally Maxwellian** (hydrodynamic): that is, unless there exist fields ρ (scalar), u (vector) and T (scalar) such that

$$f(x,v) = M_{\rho,u,T}(v) = \frac{\rho(x)\, e^{-\frac{|v-u(x)|^2}{2T(x)}}}{(2\pi T(x))^{3/2}}.$$

In words, a hydrodynamic state is a density whose velocity dependence is Gaussian, with scalar covariance — at each position x. It depends on three local parameters: density ρ, velocity u, and temperature T (temperature measures the variance of the velocity distribution).

The time-derivative of the entropy is given by an explicit formula:

$$\frac{dS}{dt} = -\frac{dH}{dt} = \int_\Omega D(f(t,x,\cdot))\, dx,$$

where

$$D(f) = \frac{1}{4} \int_{v,v_*,\sigma} B\, [f(v')f(v'_*) - f(v)f(v_*)] \log \frac{f(v')f(v'_*)}{f(v)f(v_*)}\, dv\, dv_*\, d\sigma \qquad \geq 0.$$

So Boltzmann's theorem can be recast as

$$\big[D(f) = 0\big] \quad \Longleftrightarrow \quad \Big[f(v) = M_{\rho u T}(v) \ \text{ for some parameters } \rho, u, T \Big].$$

Let me now make a **disclaimer**. Although Boltzmann's H-Theorem is 135 years old, present-day mathematics is unable to prove it rigorously and in satisfactory generality. The obstacle is the same as for many of the famous basic equations of mathematical physics: we don't know whether solutions of the Boltzmann equations are smooth enough, except in certain particular cases (close-to-equilibrium theory, spatially homogeneous theory, close-to-vacuum theory). For the moment we have to live with this shortcoming.

2 Why is the H-Theorem beautiful?

Some obvious answers are:

(i) Starting from a model based on reversible mechanics and statistics, Boltzmann finds irreversibility (which triggered endless debates);

(ii) This is an exemplification of the Second Law of Thermodynamics (entropy can only increase in an isolated system), but it is a **theorem** — as opposed to a postulate.

Here are some more "mathematical" reasons:

(iii) Its proof is clever and beautiful, although not perfectly rigorous;

(iv) It provides a powerful a priori estimate on a complicated nonlinear equation;

(v) The H functional has a statistical (microscopic) meaning: it says how exceptional the distribution function is;

(vi) The H-Theorem gives some qualitative information about the evolution of the (macroscopic) distribution function.

All these ideas are still crucial in current mathematics, as I shall discuss in the sequel.

3 The H-Theorem as an a priori estimate on a complicated nonlinear equation

Let $f(t,\cdot)$ be a solution of the full Boltzmann equation. The H-Theorem implies the estimate

$$H(f(t)) + \int_0^t \int D(f(s,x,\cdot))\,dx\,ds \ \le H(f(0)), \qquad (1)$$

where H stands for the H functional, and D for the associated dissipation functional. There is in fact equality in (1) if the solution is well-behaved, but in practise it is easier to prove the inequality, and this is still sufficient for many purposes.

Inequality (1) is in fact **two** a priori estimates! Both the finiteness of the entropy and the finiteness of the (integrated) production of entropy are of great interest.

As a start, let us note that the finiteness of the entropy is a weak and general way to prevent **concentration** (“clustering”). For instance, the bound on H guarantees that the solution of the Boltzmann equation never develops Dirac masses. This fact is physically obvious, but not so trivial from the mathematical point of view.

The first important use of the entropy as an a priori estimate goes back to Arkeryd [2] in his study of the spatially homogeneous Boltzmann equation — exactly hundred years after Boltzmann's discovery!

Both the entropy estimate and the entropy production estimate are crucial in the famous 1989 DiPerna–Lions stability theorem [10]. To simplify things, this theorem states that *Entropy, entropy production and energy bounds guarantee that a limit of solutions of the Boltzmann equation is still a solution of the Boltzmann equation.*

It is not just for the sake of elegance that DiPerna and Lions used these bounds: save for estimates which derive from conservation laws, entropy bounds are still the *only* general estimates known to this day for the full Boltzmann equation!

Robust and physically significant, entropy and entropy production bounds have been used systematically in partial differential equations and probability theory, for hundreds of models and problems. Still today, this is one of the first estimates which one investigates when encountering a new model.

Let me illustrate this with a slightly unorthodox use of the entropy production estimate, on which I worked together with Alexandre, Desvillettes and Wennberg [1]. The regularizing properties of the evolution under the Boltzmann equation depend crucially on the microscopic properties of the interaction: long-range interactions are usually associated with regularization. At the level of the Boltzmann collision kernel, long range results in a divergence for small collision angles. As a typical example, consider a collision kernel $B = |v-v_*|^\gamma\, b(\cos\theta)$, where

$$b(\cos\theta)\sin\theta \simeq \theta^{-(1+\nu)}, \qquad 0<\nu<2$$

as $\theta \to 0$. (This corresponds typically to a radially symmetric force like r^{-s} in dimension 3, and then $\nu = 2/(s-1)$.)

The regularizing effect of this singularity can be seen on the entropy production. Indeed, we could prove an estimate of the form

$$\left\|(-\Delta_v)^{\nu/4}\sqrt{f}\right\|^2_{L^2_{\mathrm{loc}}} \le\ C\Big(\int f\,dv,\ \int f|v|^2\,dv,\ H(f)\Big)\ \Big[D(f) + \int f\,(1+|v|^2)\,dv\Big].$$

Thus, some regularity is controlled by the entropy production, together with natural physical estimates (mass, energy, entropy).

4 The statistical meaning of the H functional

As I said before, the basic statistical information contained in the H functional is about how exceptional the distribution function is. This interpretation is famous but still retains some of its mystery today. It appears not only in Boltzmann's work, but also in Shannon's theory of information, which in some sense announced the ocean of information in which we are currently living.

For those who believe that entropy has always been a crystal-clear concept, let me recall a famous quote by Shannon: "*I thought of calling it "information". But the word was overly used, so I decided to call it "uncertainty". When I discussed it with John von Neumann, he had a better idea: (...) "You should call it entropy, for two reasons. In first place your uncertainty has been used in statistical mechanics under that name, so it already has a name. In second place, and more important, no one knows what entropy really is, so in a debate you will always have the advantage.*" (Note the influence of mathematicians again, for good or for bad...)

Let me recall some basic concepts from information theory; for more information (if I may say) on the subject the reader can consult the classical treatise by Shannon [16], and the more recent one by Cover and Thomas [7].

In a nutshell, the Shannon–Boltzmann entropy $S = -H = -\int f \log f$ quantifies (in logarithmic scale) how much information there is in a "random" signal, or a language; think of f as the density of the distribution of the signal. For instance, a deterministic language means complete predictability, so no surprise and no information (think of a public statement by a politician); thus $S = -\infty$.

When there is a nontrivial reference probability measure, the proper formula for H is

$$H_\mu(\nu) = \int \rho \, \log \rho \, d\mu; \qquad \nu = \rho \, \mu.$$

Let me mention in passing another hero of modern information theory: the **Fisher information** $\displaystyle\int \frac{|\nabla f|^2}{f}$ quantifies how difficult it is to reconstruct the mean value of f from the observations. Also it plays an important role in many problems of "pure" mathematics; see for instance Voiculescu [24]. But this is a different story, so I won't develop on this fascinating subject.

From a physical point of view, the entropy functional measures the **volume** of **microstates** associated, to some degree of accuracy in macroscopic observables, to a given **macroscopic** configuration, or observable distribution function. All words are meaningful here, in particular the concept of entropy implies an observation, with some margin of error. (*Entropy* is an *anthropic* notion, as the joke goes.) The basic question related to entropy is "How exceptional is the observed configuration?"

To make this fuzzy discussion a bit more concrete, I shall recall a classical computation by Boltzmann. Take N identical particles, to be distributed over k boxes, and let $f_1, \ldots, f_k$ be some (rational) frequencies (i.e. numbers in $[0,1]$ adding up to 1). Let N_j be the number of particles in the box number j. Let further $\Omega_N(f)$ be the number of configurations such that $N_j/N = f_j$ for any j. ($\Omega_N(f)$ might be zero, so I shall implicitly assume that N is so chosen that this is not the case.)

After a little bit of combinatorics (which classically rely on Stirling's formula, but the latter can be dispended with [7]), one finds

$$\#\Omega_N(f_1, \ldots, f_k) \sim e^{-N \sum f_j \log f_j} \qquad \text{as } N \to \infty,$$

in the sense that $(1/N) \log \#\Omega_N$ converges to $-\sum f_j \log f_j$, which is nothing but the discrete version of $-\int f \log f$.

The famous **Sanov theorem**, part of the theory of large deviations [8], generalizes this computation, and puts the intuition of Boltzmann on rigorous mathematical grounds. Let $x_1, \ldots, x_n, \ldots$ be independent "microscopic variables" with law μ; define

$$\widehat{\mu}^N := \frac{1}{N} \sum_{i=1}^{N} \delta_{x_i}.$$

This is a random measure, usually called the empirical measure. It contains all the information which is "macroscopically observable" if we cannot distinguish between different particles.

How will the values of $\widehat{\mu}^N$ be distributed for large N?? The fuzzy answer is that $\widehat{\mu}^N$ *visits measures more or less often according to their entropy.* Formally:

$$\mathbb{P}[\widehat{\mu}^N \simeq \nu] \sim e^{-N H_\mu(\nu)}\, d(\nu)$$

A rigorous writing of Sanov's theorem involves three limiting processes, and there is no way one can do without them. Let $(\varphi_k)_{k\in\mathbb{N}}$ be a dense sequence of Lipschitz functions (think of them as "observables"); let ν be a measure (think of it as the "guessed empirical measure"); and let $\varepsilon > 0$ be some real number (think of it as an "observation error" measured on some choice of observables). Define

$$\Omega(N,\varepsilon,k) := \left\{ (x_1,\ldots,x_N); \quad \forall j \leq k, \quad \left| \frac{\varphi_j(x_1)+\ldots+\varphi_j(x_N)}{N} - \int \varphi_j \, d\nu \right| < \varepsilon \right\}$$

(This is the set of data compatible with the observations, given the admissible error.) Then

$$H_\mu(\nu) = \lim_{k\to\infty} \limsup_{\varepsilon\to 0} \limsup_{N\to\infty} \left(-\frac{1}{N} \log \mathbb{P}_{\mu^{\otimes N}}[\Omega(N,\varepsilon,k)] \right).$$

Behind the technical trickery, the reader should recognize the famous formula written on Boltzmann's grave, $S = k \log W$.

5 Boltzmann's ideas revisited by Voiculescu

In the past decade, Voiculescu made spectacular contributions to operator theory by adapting some ideas coming from statistical mechanics and statistics. [24] A few reminders about the theory of von Neumann algebras will convince the nonexpert reader that this has apparently nothing to do with the questions upon which Boltzmann has worked.

Initially motivated by quantum mechanics and group representation theory, the study of von Neumann algebras has evolved into a mathematical domain on its own, with many respectable internal problems. Let H be a separable Hilbert space, and let $\mathcal{B}(H)$ stand for the space of bounded operators on H, equipped with the operator norm. By definition, a von Neumann algebra $\mathcal{A}$ is a sub-algebra of $\mathcal{B}(H)$ which (a) contains the identity operator I; (b) is stable by passage to the adjoint, $A \to A^*$; (c) is closed for the weak topology, defined by the linear forms $A \to \langle A\xi, \eta\rangle$ (for any two vectors ξ, η in H).

The classification of von Neumann algebras is still an active topic with famous unsolved basic questions. By definition, a *type II_1 factor* is an infinite-dimensional von Neumann algebra $\mathcal{A}$ with trivial center, equipped with a **tracial state**, i.e. a linear form $\tau : \mathcal{A} \to \mathbb{C}$ such that (a) $\tau(A^*A) \geq 0$ for any $A \in \mathcal{A}$ (positivity property); (b) $\tau(I) = 1$ (unit mass property); (c) $\tau(AB) = \tau(BA)$ (traciality). Then $(\mathcal{A}, \tau)$ is called a **noncommutative probability space**.

The analogy between "classical" and noncommutative probability spaces can be pursued to some extent. In particular, one can define the **noncommutative distribution function** of an n-tuple $(A_1,\ldots,A_n)$ of self-adjoint elements in $(\mathcal{A},\tau)$: this is just the collection of all traces of all (noncommutative) polynomials of $A_1,\ldots,A_n$. (So this is a bunch of complex numbers, say $\tau(A_1)$, $\tau(A_1A_2)$, $\tau(A_1^2A_3A_2)$, etc.)

Some of these noncommutative probability spaces can be realized as "macroscopic" limits of large random matrices. Let indeed $X_1^{(N)},\ldots,X_n^{(N)}$ be an n-tuple of random $N \times N$ matrices. If the trace of any noncommutative polynomial constructed with these matrices has a definite limit, then one may obtain (via some nonessential abstract construction) elements $A_1,\ldots,A_n$ in a noncommutative probability space, in such a way that

$$\tau\Big(P(A_1,\ldots,A_n)\Big) = \lim_{N\to\infty} \frac{1}{N}\, \mathbb{E}\ \mathrm{tr}\ P(X_1^{(N)},\ldots,X_n^{(N)}),$$

where $\mathbb{E}$ stands for probabilistic expectation.

One of the most famous such cases is **Wigner's theorem**, which asserts that the limiting distribution of large Hermitian matrices with independent Gaussian entries is the so-called semi-circular law.

Voiculescu had the following idea: Let $A_1, \ldots, A_n$ be any n-tuple of self-adjoint operators, why not try to think of $\tau = \mathrm{law}\,(\mathrm{A}_1, \ldots, \mathrm{A}_n)$ as the observable limit of a family of large matrices? In a statistical physics perspective, the operators A_i would represent some kind of macroscopic system, and the large matrices would be the microscopic system. This led him to the following definitions:

$$\Omega(N, \varepsilon, k) := \Big\{ (X_1, \ldots, X_n),\ N \times N \text{ Hermitian; } \forall P \text{ polynomial of degree} \le k, \\ \left| \frac{1}{N} \operatorname{tr} P(X_1, \ldots, X_n) - \tau(P(A_1, \ldots, A_n)) \right| < \varepsilon \Big\};$$

$$\chi(\tau) := \lim_{k \to \infty} \limsup_{\varepsilon \to 0} \limsup_{N \to \infty} \left[\frac{1}{N^2} \log \mathrm{vol}(\Omega(N, \varepsilon, k)) - \frac{n}{2} \log N \right]$$

Once again, behind the abstract framework the reader should recognize Boltzmann's formula! By lack of competence, I shall not describe the powerful results obtained by Voiculescu thanks to this new tool; but instead reproduce the 2004 quotation of the National Academy of Sciences of USA:

Award in Mathematics: A prize awarded every four years for excellence in published mathematical research goes to Dan Virgil Voiculescu, professor, department of mathematics, University of California, Berkeley. Voiculescu was chosen "for the theory of free probability, in particular, using random matrices and a new concept of entropy to solve several hitherto intractable problems in von Neumann algebras."

What would be Boltzmann's reaction if he were to learn of such an unexpected outcome of his ideas?

6 H and hydrodynamic limit

The H-Theorem is one of the main conceptual tools which can make us believe in the **hydrodynamical limit**. Take for instance the Boltzmann equation, in the form

$$\tau \frac{\partial f}{\partial t} + v \cdot \nabla_x f = \frac{1}{\mathrm{Kn}} Q(f, f),$$

where τ (time-rescaling) and Kn (Knudsen number) are positive parameters. Assume that we are considering small fluctuations around a global equilibrium, and make the ansatz

$$f(t, x, v) = \frac{e^{-\frac{|v|^2}{2}}}{(2\pi)^{3/2}} \big(1 + \varepsilon g(t, x, v)\big).$$

As Kn becomes very small (physically speaking, this means that the mean free path is very short), the effect of collisions is enhanced, and the finiteness of the entropy production forces f to be very close to a **local Maxwellian**, that is a state of the form

$$f(x, v) \simeq M_{\rho\, u\, T}(x, v) = \rho(x) \, \frac{e^{-\frac{|v - u(x)|^2}{2T(x)}}}{(2\pi T(x))^{3/2}}.$$

Indeed, it is part of the H-Theorem that such distributions are the only ones for which the entropy production vanishes.

This approximation by a local Maxwellian imposes a tremendous reduction of the complexity of the description, called **hydrodynamic approximation**.

The mathematical study of the hydrodynamic approximation has a long history; in spite of its loose formulation, it is clear that Hilbert's sixth problem alludes to it (among other things). Many authors have established the validity of the hydrodynamic approximation under various sets of assumptions (with or without smoothness, close to or far from equilibrium, in a compressible or incompressible regime, etc.) Arkeryd, Asano, Bardos, Caflisch, Cercignani, DeMasi, Ellis, Esposito, Golse, Grad, Illner, Lebowitz, Levermore, Lions, Marra, Maslova, Masmoudi, Nishida, Pinsky, Pulvirenti, Romanovski, Saint-Raymond, Ukai, and other researchers contributed to this story.

One spectacular achievement in this class of problems is the **incompressible Navier–Stokes limit in the large**, recently accomplished by Golse and Saint-Raymond [13] (with previous input by Bardos, Golse, Levermore, Lions, Masmoudi). Stated very informally, the main result is as follows. For each $\varepsilon > 0$, let f_ε solve $\varepsilon\,\partial_t f_\varepsilon + v\cdot\nabla_x f_\varepsilon = \frac{1}{\varepsilon}Q(f_\varepsilon, f_\varepsilon)$ in $[0,T]\times\mathbb{R}^3\times\mathbb{R}^3$. Make certain assumptions on the collision kernel, and assume further that

$$H(f_\varepsilon|M) = \int f_\varepsilon \log\frac{f_\varepsilon}{M}\,dx\,dv = O(\varepsilon^2).$$

Let $u_\varepsilon = \frac{1}{\varepsilon}\int f_\varepsilon(x,v)\,v\,dv$. Then any weak cluster point $u = \lim_{k\to\infty}(u_{\varepsilon_k})$, where $\varepsilon_k \to 0$, solves the incompressible Navier–Stokes equation

$$\frac{\partial u}{\partial t} + (u\cdot\nabla)u + \nabla p = \nu\,\Delta u \qquad \nabla\cdot u = 0$$

for some viscosity coefficient $\nu > 0$ which depends on the form of Boltzmann's collision kernel.

From the physical point of view, the way to achieve incompressibility is via small perturbations of equilibrium; so the unknown u is not really a velocity, but rather a fluctuation of velocity (around rest). The time-rescaling is the way to go beyond the short timescale of *sound waves*.

What makes all the beauty of the Golse–Saint-Raymond theorem, and also makes its proof incredibly difficult, is the fact that there is **no assumption of smoothness or smallness on** u.

Not suprisingly, entropy and the H-Theorem play a key role in these results. They are used to

- define the notion of convergence: $H(f|M) = O(\varepsilon^2)$, where $H(f|M)$ stands for a natural notion of "relative" H function of f with respect to the equilibrium M;
- get the compactness in the limit $\varepsilon \to 0$;
- prove the hydrodynamic behavior in the limit;
- control the conservation laws asymptotically in the limit. Indeed, the DiPerna–Lions solutions, used in this result, are so weak that there is no reason why they should satisfy the local conservation laws which are at the basis of the hydrodynamic behavior... To circumvent this major difficulty, one shows that the possible departure from hydrodynamic behavior can be controlled by entropic estimates; here is one such inequality due to Golse and Levermore [12]:

$$\left|\int \frac{Q(f,f)}{1+(f/M)}\,|v|^2\,dv\right| \le C\,\left(H(f|M) + D(f) + 1\right).$$

There seems to be no physical intuition for this last use of the entropy estimates; we can appreciate on this occasion the versatility of this tool.

7 The H Theorem as a source of qualitative macroscopic information

Let us now go back to the Boltzmann equation

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

with fixed physical scales, say in a bounded physical domain.

As already said, when the H function does not decrease strictly, f is **locally Maxwellian**:

$$f(x,v) = \rho(x)\,\frac{e^{-\frac{|v-u(x)|^2}{2T(x)}}}{(2\pi T(x))^{3/2}},$$

where ρ, u, T may depend on x.

On the other hand, when the H function is minimal, f is **globally** Maxwellian: not only does it take a Maxwellian form, but also the density ρ and the temperature T are constant all over the domain, while the velocity is everywhere equal to 0. (This rule admits some exceptions in case of exceptional geometry, but is true for, say, a generic bounded domain in $\mathbb{R}^2$ or $\mathbb{R}^3$, with specular reflection on the boundary.)

These elements suggest that the H-Theorem is a great tool to study relaxation to equilibrium.

An ambitious program about entropy production was started in the early nineties by Carlen–Carvalho and Desvillettes, with some critical impulse by Cercignani (and inspiration from earlier work of McKean). The goal of this program was to *study the large-time behavior of the Boltzmann equation in the large, through the behavior of the H functional.* One hopes to prove in particular that the solution $f(t,\cdot)$ approaches a global Maxwellian equilibrium as $t \to \infty$, *keeping a control on time scales!*

A noticeable achievement of this program is the following result due to Desvillettes and myself [9]. It is a conditional result, in the sense that it applies under regularity conditions which in general are still an open problem for the Boltzmann equation:

Let $f(t,x,v)$ be a solution of the Boltzmann equation, with appropriate boundary conditions. Assume that

(i) f is very regular (uniformly in time): all moments ($\int f|v|^k\,dv\,dx$) are finite and all derivatives (of any order) of f are bounded;

(ii) f is strictly positive: $f(t,x,v) \geq Ke^{-A|v|^q}$.

Then $f(t,\cdot)$ approaches the unique equilibrium as $t \to \infty$, and the convergence is at least like $O(t^{-\infty})$ (the distance to equilibrium goes to 0 faster than any inverse power of time).

At present, this theorem applies only in certain particular cases (spatial homogeneity; close-to-equilibrium) because of the limitations of the regularity theory of the Boltzmann equation. What it achieves, in full generality, is a *conversion of regularity bounds into decay bounds.*

The proof of this theorem uses differential inequalities of first and second order, coupled via many inequalities including:

- precised **entropy production inequalities** (who would have guessed...). For instance, under some strong a priori estimates on f (smoothness, moments, positivity), one can show that

$$D(f) \geq K_{f,\varepsilon}[H(f) - H(M^f_{\rho u T})]^{1+\varepsilon}.$$

(This is a modified version of the so-called "Cercignani conjecture" [6, 18, 20].)

- the instability of hydrodynamical description: for instance we show that

$$\frac{d^2}{dt^2}\left(\int (f - M^f_{\rho\, u\, T})^2\right) \geq K \int_\Omega (|\nabla_x T|^2 + |\mathrm{dev}(u)|^2)\,dx \;-\; C_\varepsilon(f)\,\|f - M^f_{\rho\, u\, T}\|^{1-\varepsilon}_{L^2}\, H(f|M)^{\frac{1-\varepsilon}{2}},$$

where $M^f_{\rho\, u\, T}$ stands for the local Maxwellian with same density ρ, velocity u and temperature T as f; and dev is the deviatoric part of u, i.e. the traceless symmetric part of the Reynolds tensor ∇u.

Such inequalities mean that if f becomes very close to a hydrodynamic state, then the time-evolution of the (squared) distance between f and the set of hydrodynamical states will evolve in a convex manner, and therefore f will depart from hydrodynamic behavior.

Another key feature of the proof is the fact that entropy sees both kinetic and hydrodynamic effects. Indeed, it can be separated into a "purely kinetic" and a "purely hydrodynamic" contributions: $H(f) - H(M) = [H(f) - H(M^f_{\rho\, u\, T})] \;+\; \mathcal{H}(\rho,u,T)$, where $\mathcal{H}(\rho,u,T) = \displaystyle\int_\Omega \rho \log \frac{\rho}{T^{3/2}}\,dx$.

A careful study of the proof led Desvillettes and myself to conjecture the existence of strong time-oscillations in the entropy production. These oscillations were spectacularly confirmed in numerical simulations by Francis Filbet (see figures below), and are currently being studied, both numerically and theoretically [11].

More information about the history and achievements of the subject can be found in the proceedings of the 2003 International Congress of Mathematical Physics [21], or in a course which I taught at the Institut Henri Poincaré [22].

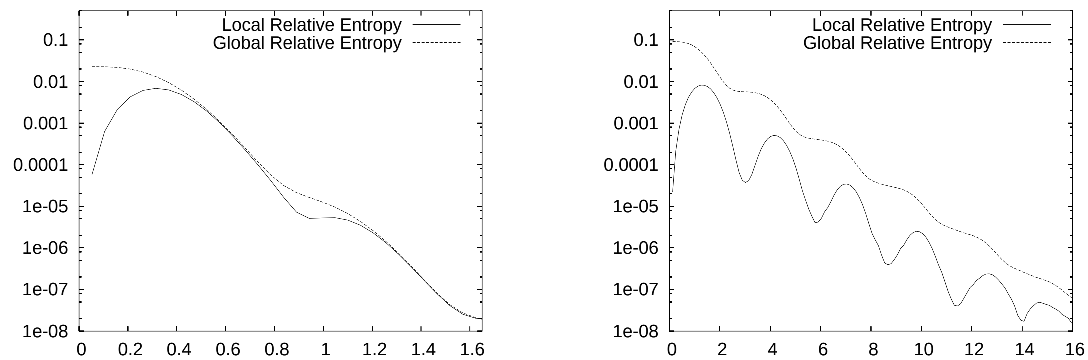

Figure 1. Time-decay of the H-function, in logarithmic scale, for the Boltzmann equation in one dimension of space and two dimensions of velocity, with periodic boundary conditions. The curve above is the H-function, the curve below is the purely kinetic part of the H-function; when the two curves are far from each other the gas is almost hydrodynamical, when they are close to each other it is almost homogeneous.

8 More on the qualitative behavior of the entropy

The importance and flexibility of Boltzmann's H functional as a way to describe the qualitative behavior of a system goes much beyond the field of partial differential equations. In the last two sections of this text I shall illustrate this by two examples: (a) the central limit theorem in probability theory; (b) Ricci curvature bounds in Riemannian geometry.

9 The central limit theorem

Let $X_1, X_2, \ldots, X_n, \ldots$ be identically distributed, independent real random variables; assume that $\mathbb{E}X_j^2 < \infty, \quad \mathbb{E}X_j = 0$. Then

$$\frac{X_1 + \ldots + X_N}{\sqrt{N}} \xrightarrow[N\to\infty]{} \text{Gaussian random variable.}$$

This is the central limit theorem which we learn in basic probability courses.

A few years ago, Ball, Barthe and Naor [3] interpreted this as an *irreversible loss of information* along the sequence of random variables; namely,

$$\text{Entropy}\left(\frac{X_1 + \ldots + X_N}{\sqrt{N}}\right) \quad \text{increases with } N.$$

(Weaker versions of this theorem, involving only powers of 2, had been proven before by Barron; and Carlen–Soffer.)

This "entropic" proof of the central limit theorem is very different from the usual proof based on Fourier transform; it is also much more complicated. Still it gives an information-theoretical interpretation of the central limit theorem which Boltzmann certainly would have appreciated very much.

10 H functional and Ricci curvature

My last example will lead me into differential geometry. The Ricci curvature is (together with the sectional and scalar curvatures) one of the three most popular notions of curvature. If M is a Riemannian manifold, then the Ricci curvature Ric_x at x is a quadratic form on T_xM.

A quick (albeit incomprehensible) definition of the Ricci curvature is by contraction of the Riemann curvature tensor: $(\mathrm{Ric})_{ij} = (\mathrm{Riem})^k_{kij}$. Intuitively, the Ricci curvature measures the rate of separation of geodesics in a given direction, **in the sense of volume** (Jacobian). As usual, positive curvature indicates a tendency for geodesics to converge, while negative curvature indicates a tendency to diverge faster than in Euclidean space.

A more hand-on approach to Ricci curvature bounds is in terms of distortion coefficients. For instance, it is equivalent to say that the Ricci curvature of a Riemannian manifold M is nonnegative; or that its distortion coefficients are never less than 1, i.e. one always overestimates the surface of observed light sources.

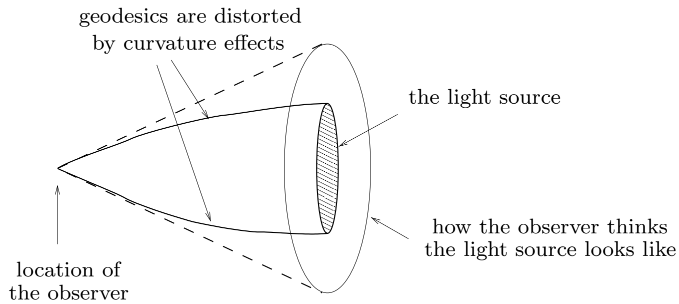

Figure 2. Because of positive curvature effects, the observer overestimates the surface of the light source.

Lower bounds on Ricci curvature are of constant use in Riemannian geometry: they appear as privileged assumptions for isoperimetric inequalities; heat kernel estimates; Sobolev inequalities; diameter control; spectral gap inequalities; upper bounds on the volume growth; compactness criteria for families of manifolds; etc.

The following theorem was proven independently by Lott and me [14]; and by Sturm [17]:

A limit of manifolds with nonnegative Ricci curvature, is also of nonnegative curvature.

This theorem is interesting because the notion of limit used is a very weak one, namely the measured Gromov–Hausdorff topology; and there is no reason why the Ricci tensor would pass to the limit in the process. The limit might even occur with a reduction of dimension (collapsing); in which case one has to use a slightly more intrinsic notion of Ricci curvature allowing for a change of reference measure.

Since I mention this theorem in this lecture, the reader has probably guessed that its proof uses Boltzmann's entropy! And indeed the proof of the theorem does rely on entropy, in relation with the **optimal transport of probability measures**, along a direction of research developed by various authors (Cordero-Erausquin, Lott, McCann, Otto, von Renesse, Schmuckenschläger, Sturm, and myself).

To explain this connection I shall describe what I call the *lazy gas experiment*.

Take a perfect gas in which particles do not interact, and ask him to move from a certain prescribed density field at time $t = 0$, to another prescribed density field at time $t = 1$. Since the gas is lazy, he will find a way to do so by spending a minimal amount of work (least action path). Measure the entropy of the gas at each time, and check that it always lie *above* the line joining the final and initial entropies. If such is the case, then we know that we live in a nonnegatively curved space.

Of course these heuristics do not explain why the entropy is precisely the relevant functional to measure the concentration of the gas; this prediction was made by Otto and myself in 2000 [15], after a study of relations between optimal transport, logarithmic Sobolev inequalities and the entropy functional. In the reference book [23] I explain all this and much much more; still I remain marvelled by this new role of Boltzmann's ubiquitous entropy.

References

[1] Alexandre, R., Desvillettes, L., Villani, C., and Wennberg, B. Entropy dissipation and long-range interactions. *Arch. Ration. Mech. Anal. 152*, 4 (2000), 327–355.

[2] Arkeryd, L. On the Boltzmann equation. *Arch. Rational Mech. Anal. 45* (1972), 1–34.

[3] Artstein, S., Ball, K. M., Barthe, F., and Naor, A. Solution of Shannon's problem on the monotonicity of entropy. *J. Amer. Math. Soc. 17*, 4 (2004), 975–982.

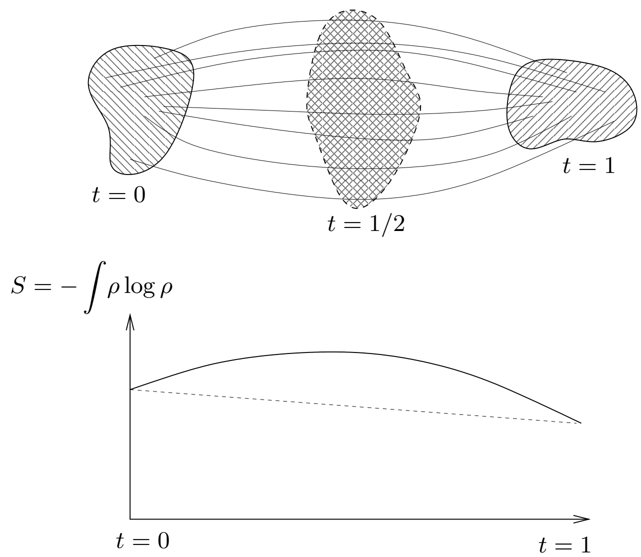

Figure 3. The lazy gas experiment: To go from initial state to final state, the lazy gas uses a path of least action. In a nonnegatively curved world, the trajectories of particles first diverge, then converge, so that at intermediate times the gas can afford to have a lower density (higher entropy).

[4] Boltzmann, L. Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen. *Sitzungsberichte der Akademie der Wissenschaften 66* (1872), 275–370. Translation : Further studies on the thermal equilibrium of gas molecules, in *Kinetic Theory 2*, 88–174, Ed. S.G. Brush, Pergamon, Oxford (1966).

[5] Boltzmann, L. *Lectures on gas theory.* University of California Press, Berkeley, 1964. Translated by Stephen G. Brush. Reprint of the 1896–1898 Edition by Dover Publications, 1995.

[6] Cercignani, C. H-Theorem and trend to equilibrium in the kinetic theory of gases. *Arch. Mech. 34* (1982), 231–241.

[7] Cover, T. M., and Thomas, J. A. *Elements of information theory.* John Wiley & Sons Inc., New York, 1991.

[8] Dembo, A., and Zeitouni, O. *Large deviations techniques and applications*, second ed. Springer-Verlag, New York, 1998.

[9] Desvillettes, L., and Villani, C. On the trend to global equilibrium for spatially inhomogeneous kinetic systems: the Boltzmann equation. *Invent. Math. 159*, 2 (2005), 245–316.

[10] DiPerna, R. J., and Lions, P.-L. On the Cauchy problem for the Boltzmann equation: Global existence and weak stability. *Ann. of Math. (2) 130* (1989), 312–366.

[11] Filbet, F., Mouhot, C., and Pareschi, L. Solving the Boltzmann equation in $N \log_2 N$. *SIAM J. Sci. Compput. 28*, 2 (2006), 1029–1053.

[12] Golse, F., and Levermore, C. D. Stokes-Fourier and acoustic limits for the Boltzmann equation: convergence proofs. *Comm. Pure Appl. Math. 55*, 3 (2002), 336–393.

[13] Golse, F., and Saint-Raymond, L. The Navier–Stokes limit of the Boltzmann equation for bounded collision kernels. *Invent. Math. 155*, 1 (2004), 81–161. za

[14] Lott, J., and Villani, C. Ricci curvature for metric-measure spaces via optimal transport. To appear in *Ann. of Math. (2).*

[15] Otto, F., and Villani, C. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal. 173*, 2 (2000), 361–400.

[16] Shannon, C. E., and Weaver, W. *The Mathematical Theory of Communication.* The University of Illinois Press, Urbana, Ill., 1949.

[17] Sturm, K.-T. On the geometry of metric measure spaces. I, II. *Acta Math. 196*, 1 (2006), 65–131, 133–177.

[18] 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 (1999), 667–706.

[19] VILLANI, C. A review of mathematical topics in collisional kinetic theory. In *Handbook of mathematical fluid dynamics, Vol. I.* North-Holland, Amsterdam, 2002, pp. 71–305.

[20] VILLANI, C. Cercignani's conjecture is sometimes true and always almost true. *Comm. Math. Phys. 234*, 3 (2003), 455–490.

[21] VILLANI, C. Entropy production and convergence to equilibrium for the Boltzmann equation. In *XIVth International Congress on Mathematical Physics*. World Sci. Publ., Hackensack, NJ, 2005, pp. 130–144.

[22] VILLANI, C. Entropy dissipation and convergence to equilibrium. Notes from a series of lectures in Institut Henri Poincaré, Paris (2001); updated in 2007. In *Entropy Methods for the Boltzmann Equation*, *Lect. Notes in Math. 1916*, Springer, Berlin, 2007, pp. 1–70.

[23] VILLANI, C. Optimal transport, old and new. Notes for the 2005 Saint-Flour summer school; available online at `www.umpa.ens-lyon.fr/~cvillani`.

[24] VOICULESCU, D. Lectures on free probability theory. In *Lectures on probability theory and statistics (Saint-Flour, 1998)*, vol. 1738 of *Lecture Notes in Math.* Springer, Berlin, 2000, pp. 279–349.

ENS Lyon (UMR CNRS 5669) & Institut Universitaire de France
E-mail: cedric.villani@umpa.ens-lyon.fr