ENTROPY AND *H* THEOREM

THE MATHEMATICAL LEGACY OF LUDWIG BOLTZMANN

Golm, July 10, 2009

Cédric Villani
ENS Lyon & Institut Henri Poincaré

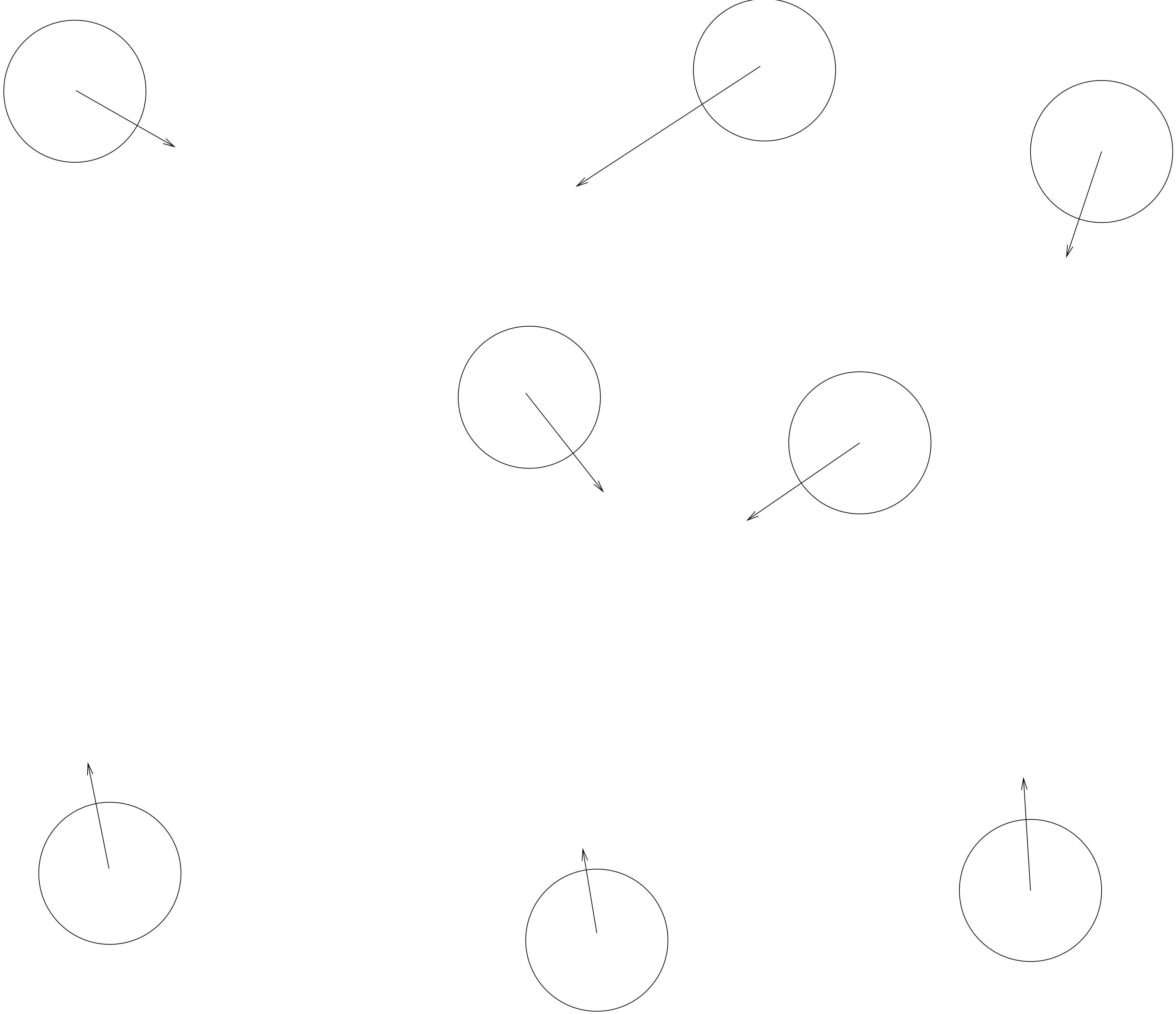
Cutting-edge physics at the end of nineteenth century

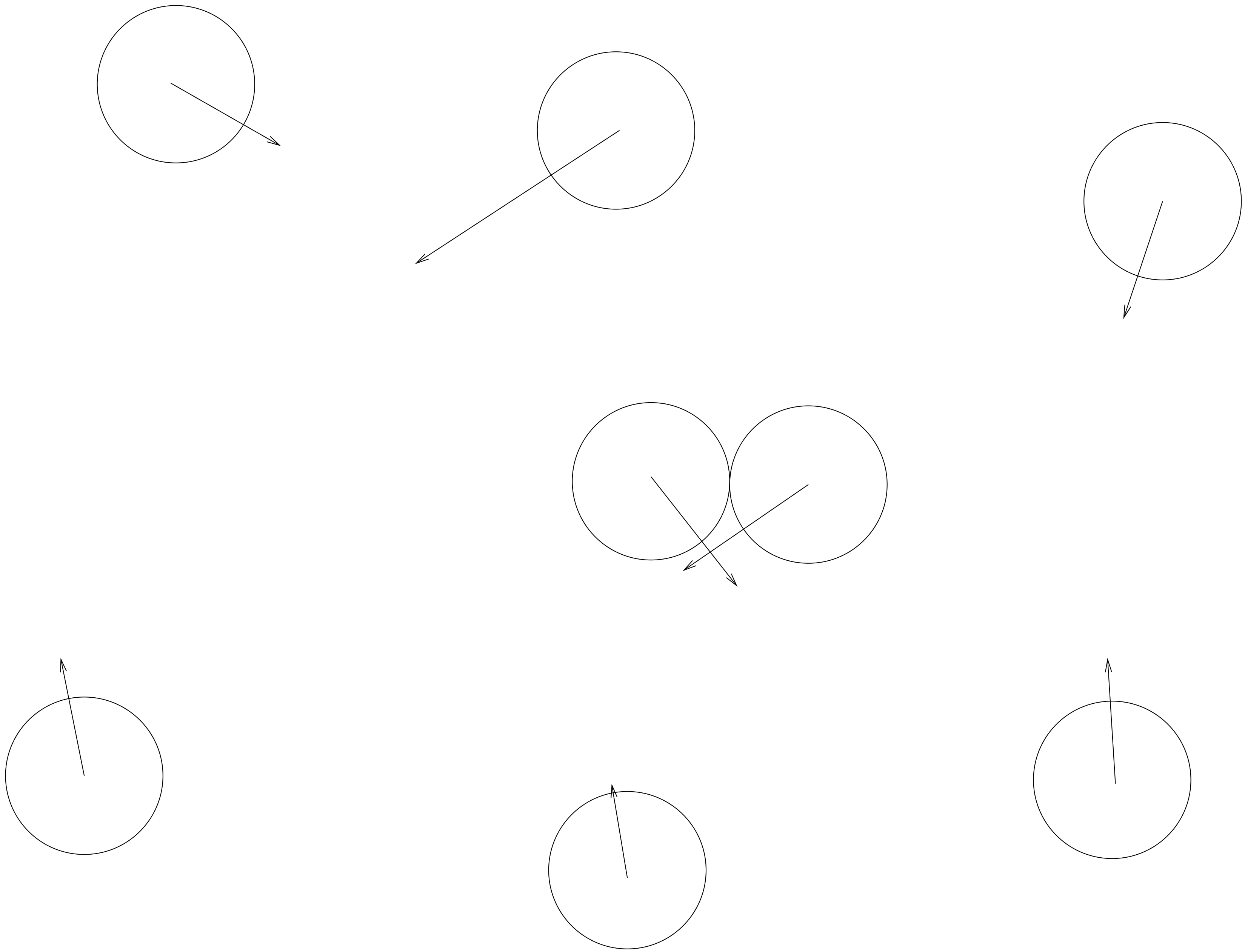
Long-time behavior of a (dilute) classical gas

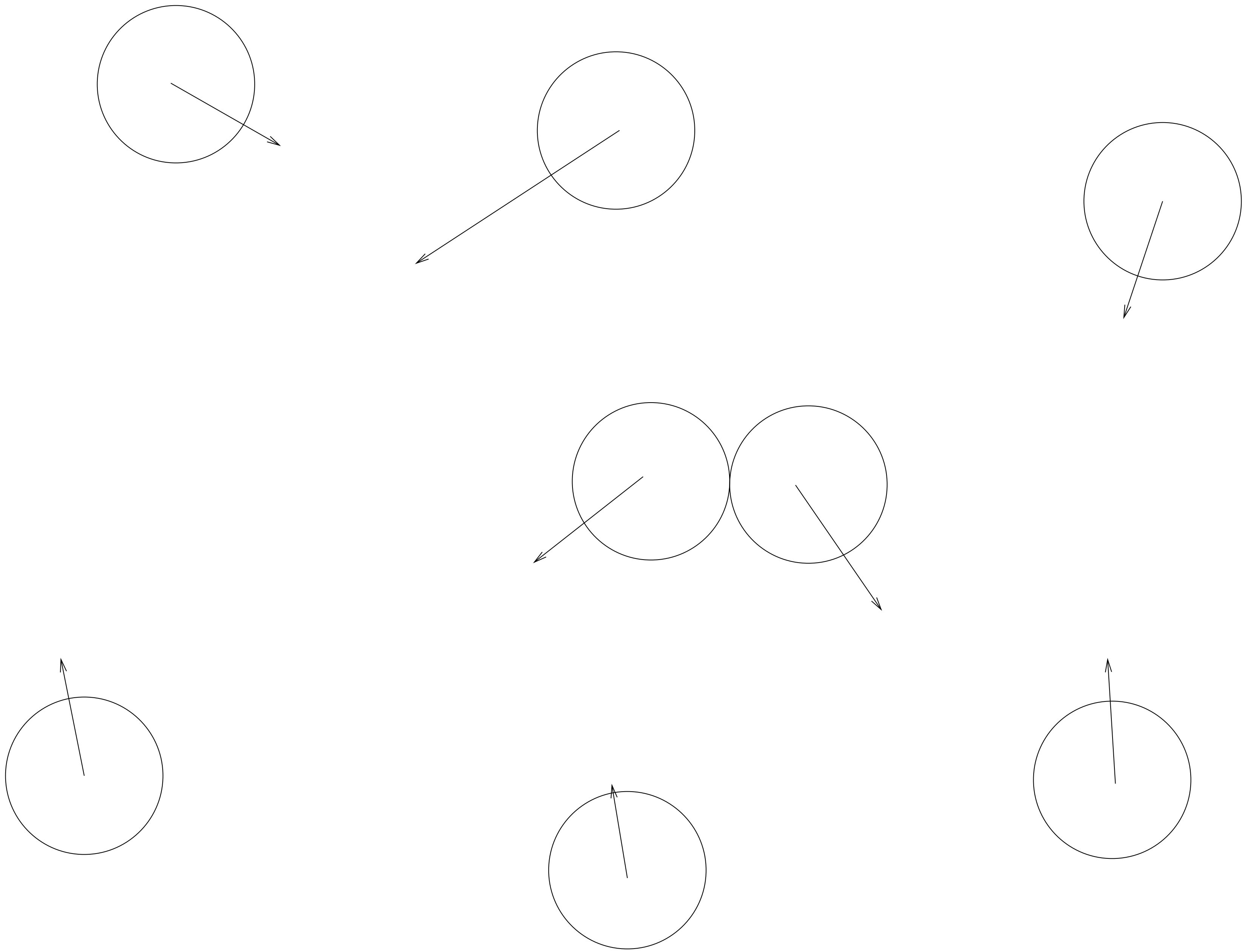
Take many (say 10^{20}) small hard balls,

bouncing against each other, in a box

Let the gas evolve according to Newton's equations







Prediction by Maxwell and Boltzmann



The distribution function is asymptotically Gaussian

$$f(t,x,v) \simeq a \exp\left(-\frac{|v|^2}{2T}\right) \qquad \text{as } t \to \infty$$

Based on two major conceptual advances

- the Boltzmann equation models the dynamics of rarefied gases via the position-velocity density $f(x,v)$:

$$\frac{\partial f}{\partial t} + \sum_{i=1}^{3} v_i \frac{\partial f}{\partial x_i} = Q(f,f)$$

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

(+ boundary conditions)

- the increase of ENTROPY (H Theorem)

1872: Boltzmann's H Theorem

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

Boltzmann identifies S with the entropy of the gas

and proves that S can only increase in time

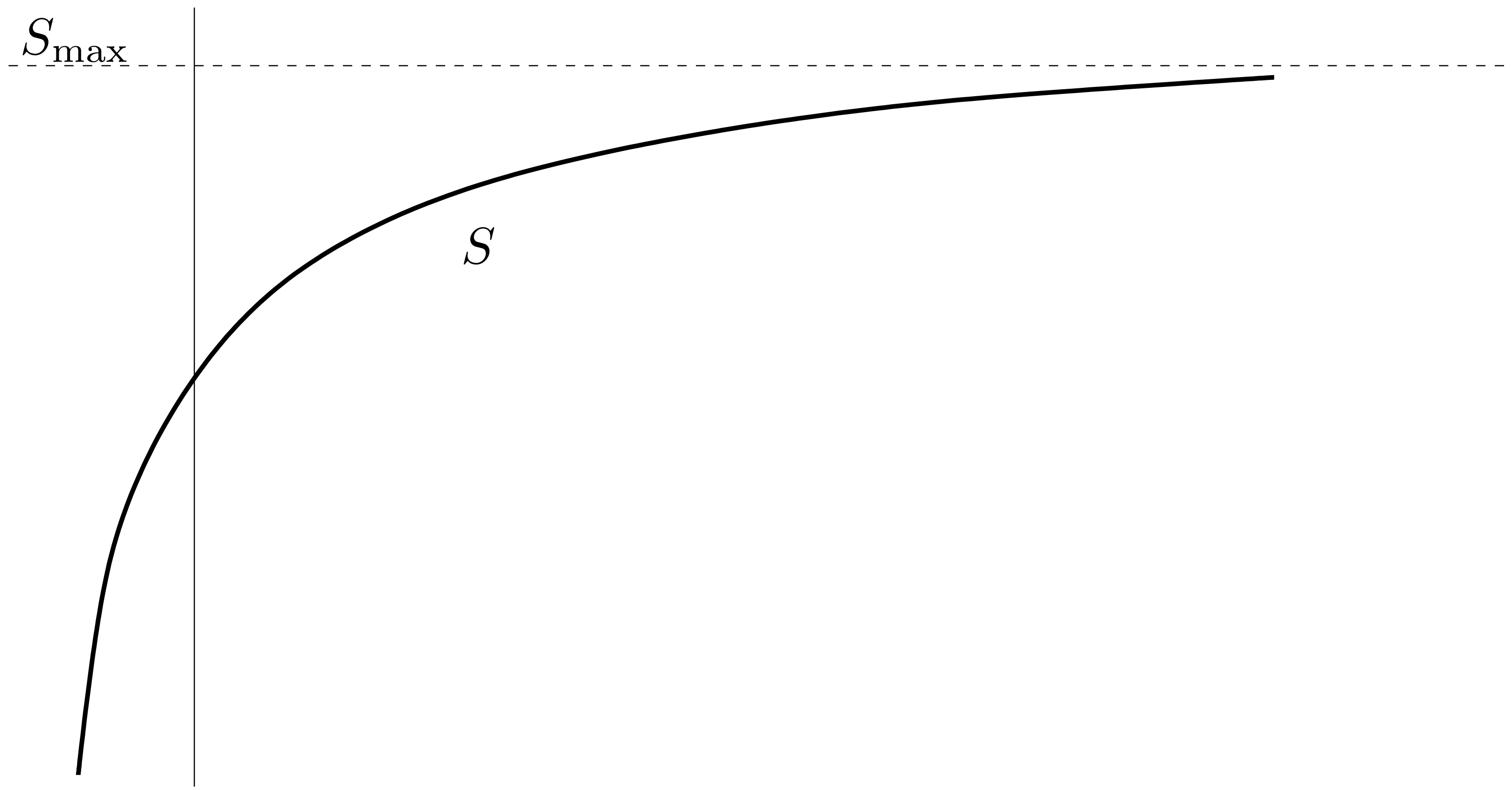
(strictly unless the gas is in a hydrodynamical state)

— an instance of the Second Law of Thermodynamics

The state of maximum entropy given the conservation of total mass and energy is a Gaussian distribution

... and this Gaussian distribution is the only one which prevents entropy from growing further

Plausible time-behavior of the entropy



If this is true, then the distribution becomes Gaussian!

Why is the *H* Theorem beautiful?

- Starting from a model based on reversible mechanics + statistics, Boltzmann finds irreversibility

- This is a theorem — as opposed to a postulate

More "mathematical" reasons:

- Beautiful proof, although not perfectly rigorous

- A priori estimate on a complicated nonlinear equation

- The *H* functional has a statistical (microscopic) meaning: how exceptional is the distribution function

- Gives some qualitative information about the evolution of the (macroscopic) distribution function

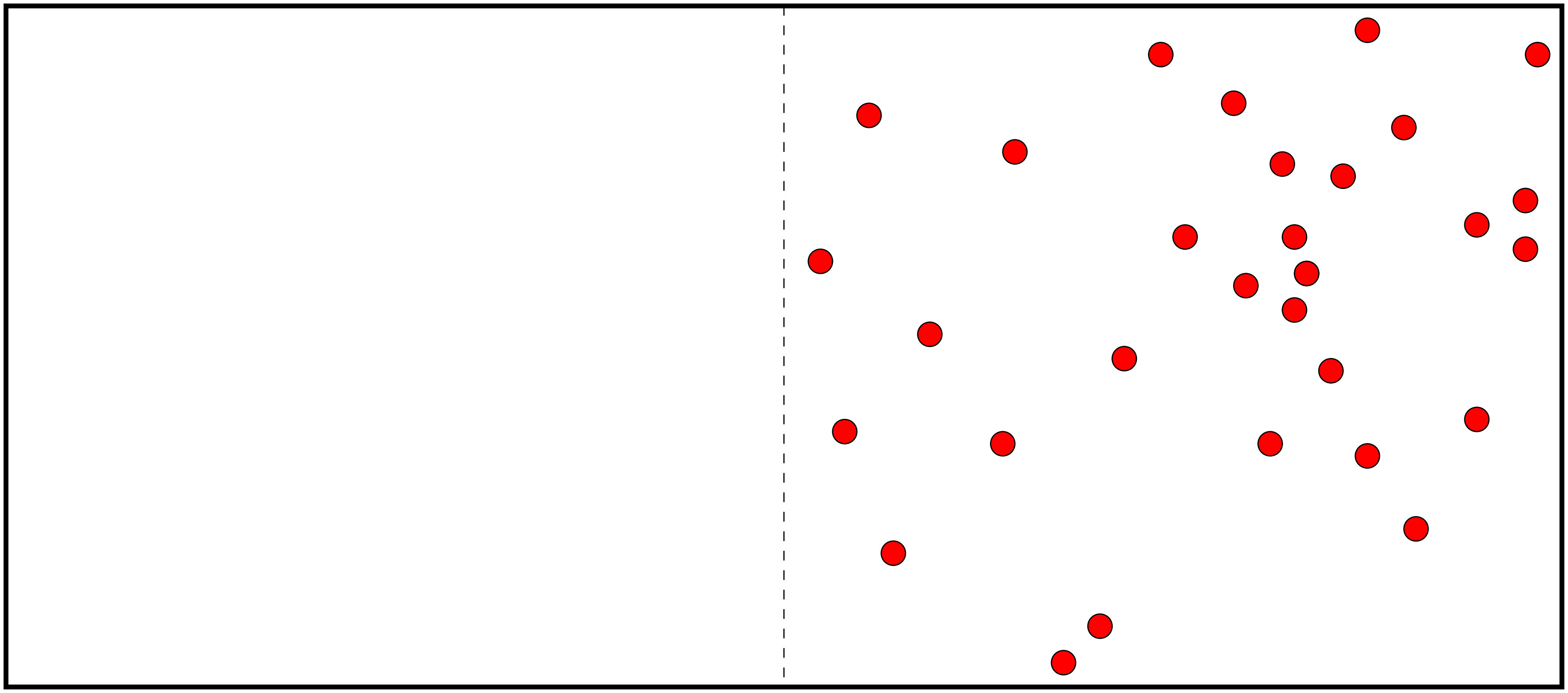
These ideas are still crucial in current mathematics

Not all did appreciate this theorem!

Ostwald, Loschmidt, Zermelo, Poincaré... questioned Boltzmann's arguments

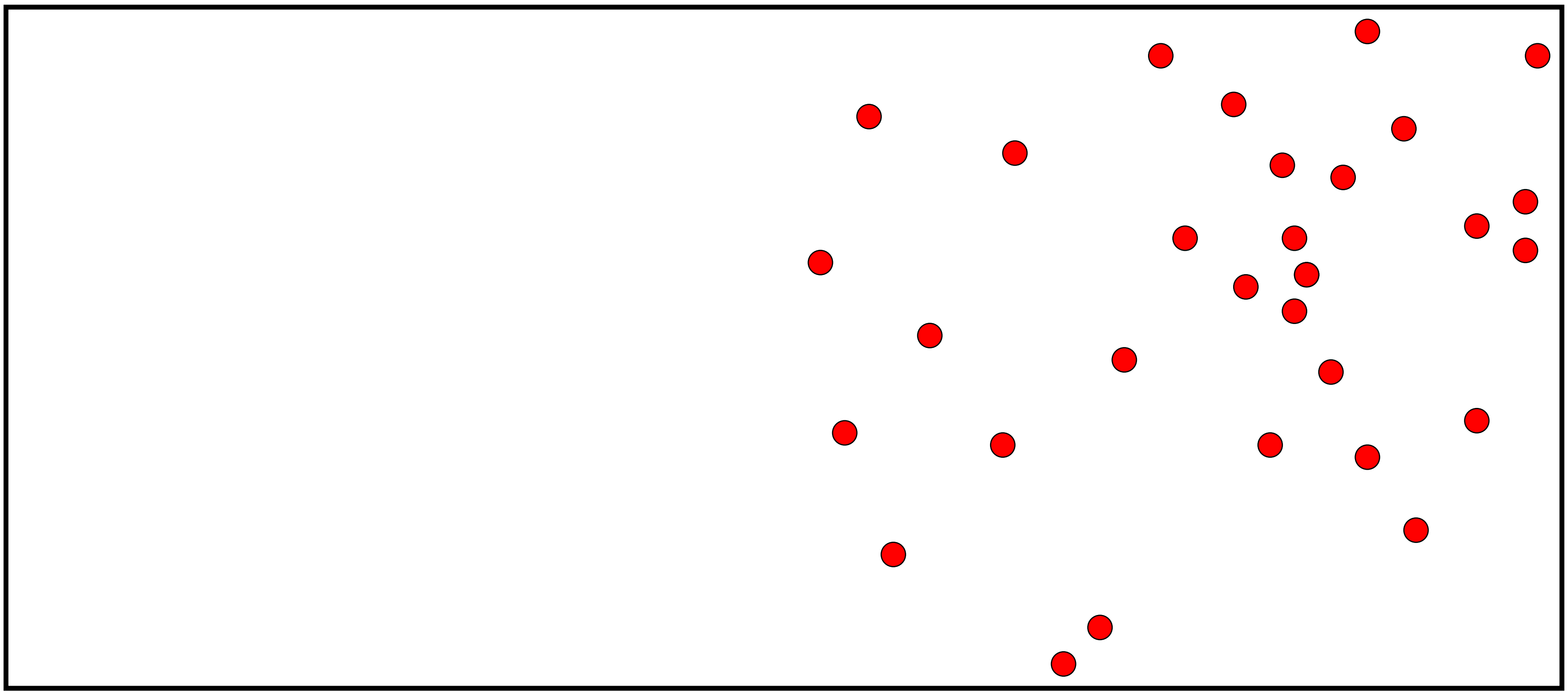
Not all did appreciate this theorem!

Ostwald, Loschmidt, Zermelo, Poincaré... questioned Boltzmann's arguments



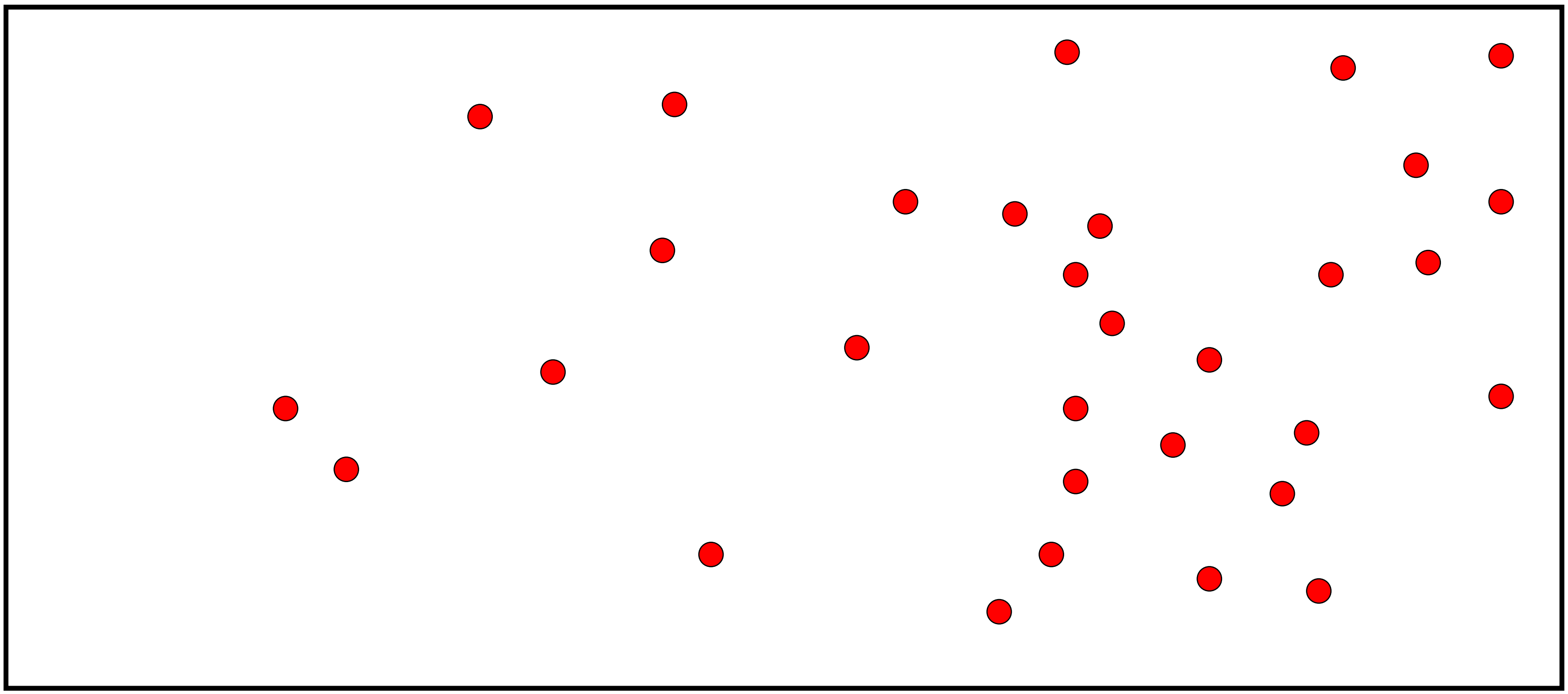
Not all did appreciate this theorem!

Ostwald, Loschmidt, Zermelo, Poincaré... questioned Boltzmann's arguments



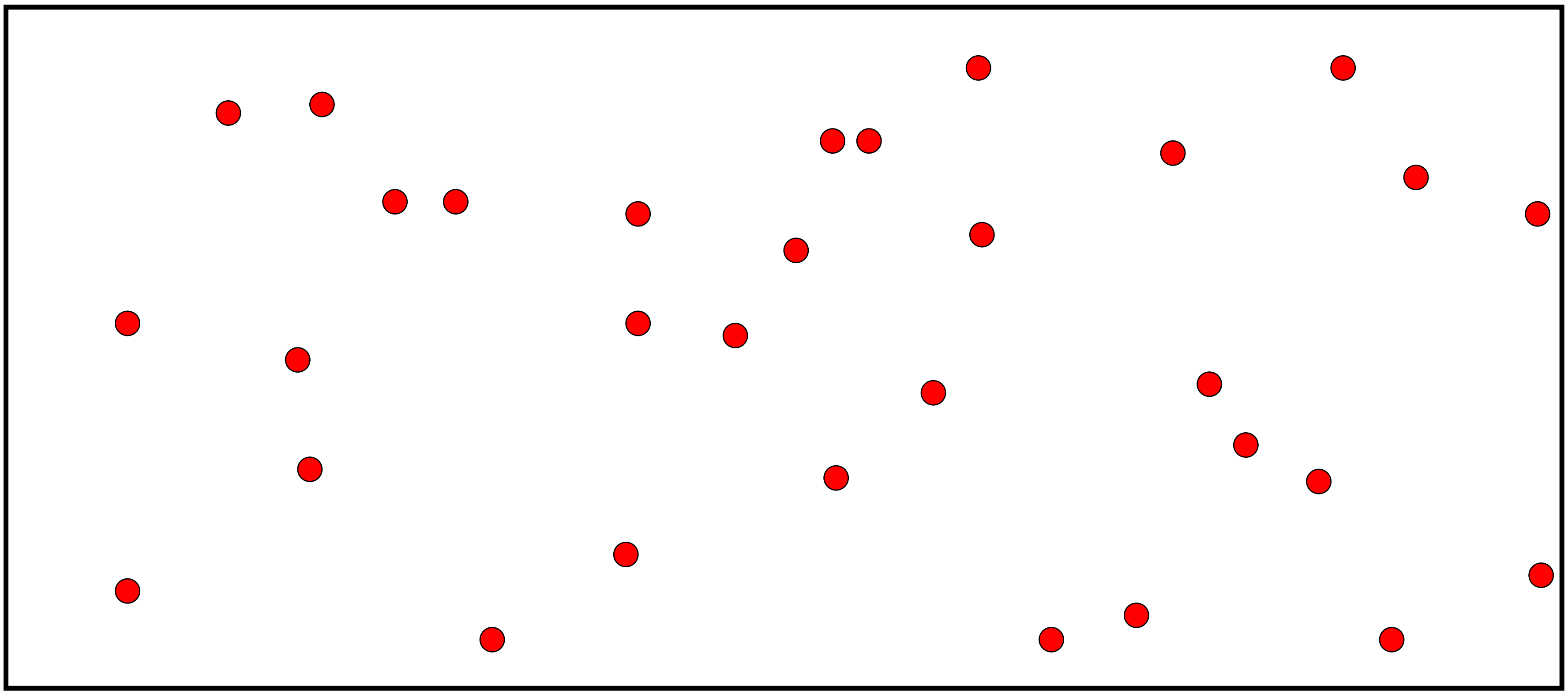
Not all did appreciate this theorem!

Ostwald, Loschmidt, Zermelo, Poincaré... questioned Boltzmann's arguments



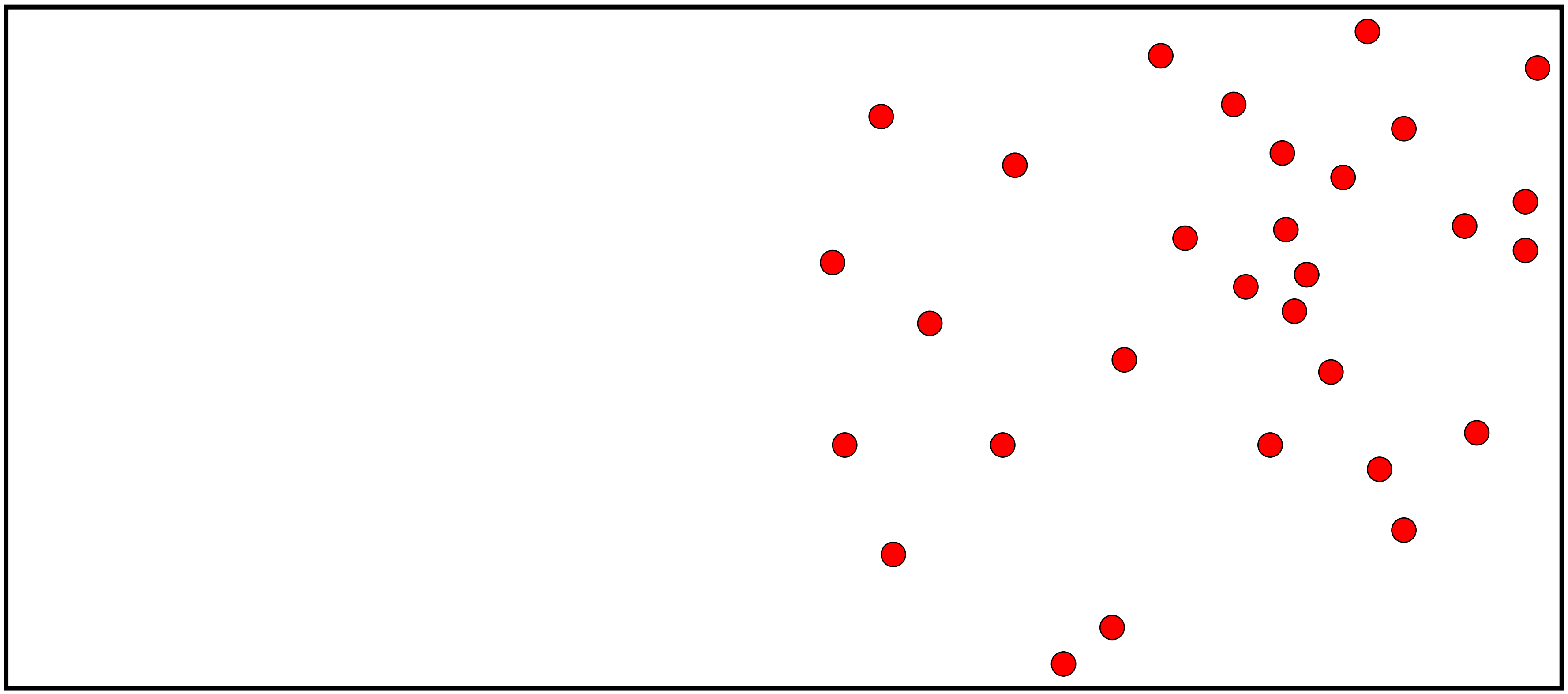
Not all did appreciate this theorem!

Ostwald, Loschmidt, Zermelo, Poincaré... questioned Boltzmann's arguments



Not all did appreciate this theorem!

Ostwald, Loschmidt, Zermelo, Poincaré... questioned Boltzmann's arguments



Mathematicians chose their side

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

Arnold Sommerfeld (about a 1895 debate)

Boltzmann's work on the principles of mechanics suggest 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)

Three remarkable features of the H functional/Theorem

1) A priori estimate on a complicated nonlinear equation

2) Statistical (microscopic) meaning: how exceptional is the distribution function

3) Qualitative information about the evolution of the (macroscopic) distribution function

The H Theorem as an a priori estimate

Consider a solution of the Boltzmann equation

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

If H is finite at initial time, it will remain so at later times

Moreover, the amount of produced entropy is a priori controlled

$$H(f(t)) + \int_0^t \int EP(f(s, x))\, dx\, ds \ \leq H(f(0))$$

These are two a priori estimates!

Importance of the entropy a priori estimates

Finiteness of the entropy is a weak and general way to prevent concentration ("clustering").

First important use : Arkeryd (1972) for the spatially homogeneous Boltzmann equation.

Both estimates are crucial in the DiPerna–Lions stability theorem (1989): Entropy, entropy production and energy bounds guarantee that a limit of solutions of the BE is a solution of the BE

Nowadays entropy and entropy production estimates (robust and physically significant) are being used systematically for hundreds of problems in probability and partial differential equations

Example (Alexandre, Desvillettes, V, Wennberg, 2000)

Regularizing effects of long-range interaction can be seen on the entropy production

collision kernel $B = |v - v_*|^\gamma \, b(\cos\theta)$,

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

(force like $r^{-s} \Longrightarrow \nu = 2/(s-1)$)

$$\left\| (-\Delta_v)^{\nu/4} \sqrt{f} \right\|^2_{L^2_{\mathrm{loc}}}$$

$$\leq \; C\Big(\int f\, dv, \int f|v|^2\, dv, H(f)\Big) \left[EP(f) + \int f(1+|v|^2)\, dv \right]$$

Three remarkable features (continued)

1) A priori estimate on a complicated nonlinear equation

2) Statistical (microscopic) meaning: how exceptional is the distribution function

3) Qualitative information about the evolution of the (macroscopic) distribution function

The content of the H functional

Mysterious and famous, also appears in Shannon's theory of information

In Shannon's own words:

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."

Information theory

The Shannon–Boltzmann entropy $S = -\int f \log f$ quantifies how much information there is in a "random" signal, or a language.

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

Microscopic meaning of the entropy functional

Measures the volume of microstates associated,

to some degree of accuracy in macroscopic observables,

to a given macroscopic configuration (observable distribution function)

$\Longrightarrow$ How exceptional is the observed configuration?

Boltzmann's formula

$$S = k \log W$$


S = k. log W
LVDWIG
BOLTZMANN
1844 1906
DR PHIL PAUL
BOLTZMAN

$\longrightarrow$ How to go from $S = k \log W$ to $S = -\int f \log f$?

Famous computation by Boltzmann

N particles in k boxes

$f_1, \ldots, f_k$ some (rational) frequencies; $\sum f_j = 1$

N_j = number of particles in box $\#j$

$\Omega_N(f)$ = number of configurations such that $N_j/N = f_j$

Then as $N \to \infty$

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

$$\frac{1}{N} \log \#\Omega_N(f_1, \ldots, f_k) \simeq -\sum f_j \log f_j$$

Sanov's theorem

A mathematical translation of Boltzmann's intuition

$x_1, \ldots, x_n, \ldots$ ("microscopic variables") independent with law μ;

$\widehat{\mu}^N := \frac{1}{N} \sum_{i=1}^{N} \delta_{x_i}$ (random measure, "macroscopically observable")

What measure will we observe??

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

Rigorous writing: $H_\mu(\nu) = \lim_{k\to\infty} \limsup_{\varepsilon\to 0} \limsup_{N\to\infty}$

$$- \frac{1}{N} \log \mathbb{P}_{\mu^{\otimes N}} \left[\left\{ \forall j \leq k, \quad \left| \frac{\varphi_j(x_1) + \ldots + \varphi_j(x_N)}{N} - \int \varphi_j \, d\nu \right| < \varepsilon \right\} \right]$$

Voiculescu's adaptation of Boltzmann's idea

Recall: Von Neumann algebras

Initial motivations: Quantum mechanics, group representation theory

H a Hilbert space

$\mathcal{B}(H)$: bounded operators on H, with operator norm

Von Neumann algebra $\mathcal{A}$: a sub-algebra of $\mathcal{B}(H)$,

- containing I
- stable by $A \to A^*$
- closed for the weak topology (w.r.t. $A \to \langle A\xi, \eta\rangle$)

The classification of VN algebras is still an active topic with famous unsolved basic problems

States

Type II_1 factor := infinite-dimensional VN algebra with trivial center and a tracial state = linear form τ s.t.

$$\tau(A^*A) \geq 0 \quad \text{(positivity)}$$

$$\tau(I) = 1 \quad \text{(unit mass)}$$

$$\tau(AB) = \tau(BA)$$

$(\mathcal{A}, \tau)$: noncommutative probability space

Noncommutative distribution functions

Let $A_1, \ldots, A_n$ be self-adjoint in $(\mathcal{A}, \tau)$.

$\text{law}(A_1, \ldots, A_n)$: the collection of all traces of all (noncommutative) polynomials of $A_1, \ldots, A_n$

Noncommutative probability spaces as "macroscopic" limits

$X_1^{(N)}, \ldots, X_n^{(N)}$ random $N \times N$ matrices

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

may define a noncommutative probability space.

Voiculescu's entropy

Think of $\tau = \mathrm{law}(A_1, \ldots, A_n)$ as the observable limit of a family of "microscopic systems" = large matrices

$$\Omega(N, \varepsilon, k) :=$$

$$\Big\{(X_1, \ldots, X_n),\ N \times N \text{ Hermitian; } \forall P \text{ polynomial of degree} \leq k,$$

$$\left| \frac{1}{N} \operatorname{tr} P(X_1, \ldots, X_n) - \tau(P(A_1, \ldots, A_n)) \right| < \varepsilon \Big\}$$

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

Quotation (2004):

National Academy of Sciences – USA: Award in Mathematics: A prize of $5,000 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."

Three remarkable features (continued)

1) A priori estimate on a complicated nonlinear equation

2) Statistical (microscopic) meaning: how exceptional is the distribution function

3) Qualitative information about the evolution of the (macroscopic) distribution function

The H Theorem is the main “explanation” for the hydrodynamic approximation of the Boltzmann equation

In a regime where there are many collisions per unit of time, finiteness of the entropy production forces f to be very close to a local Maxwellian:

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

(only states for which entropy production vanishes)

This imposes a tremendous reduction of the complexity

The theoretical justification of this approximation has been studied by many authors (part of Hilbert’s sixth problem)

Back to the large-time behavior

A theoretical program started in the early nineties by Carlen–Carvalho and Desvillettes, with some impulse by Cercignani

Problem: Study the large-time behavior of the Boltzmann equation in the large, through the behavior of the entropy production

$\longrightarrow$ prove that $f(t,\cdot)$ approaches the global Maxwellian (equilibrium) as $t \to \infty$?
(**keeping a control on time scales!!**)

What do we want to prove?

Prove that a "nice" solution of the Boltzmann equation approaches Gaussian equilibrium: 2 pages (basically Boltzmann's argument)

Get quantitative estimates like "After a time ....., the distribution is close to Gaussian, up to an error of 1 %": 90 pages of proof (1989–2004)

Note: General current trend in mathematics: go back to constructive arguments

Conditional convergence theorem (Desvillettes – V)

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) are bounded;

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

Then $f(t) \longrightarrow f_\infty$ at least like $O(t^{-\infty})$ as $t \to \infty$

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

- precised entropy production inequalities (information theoretical input)

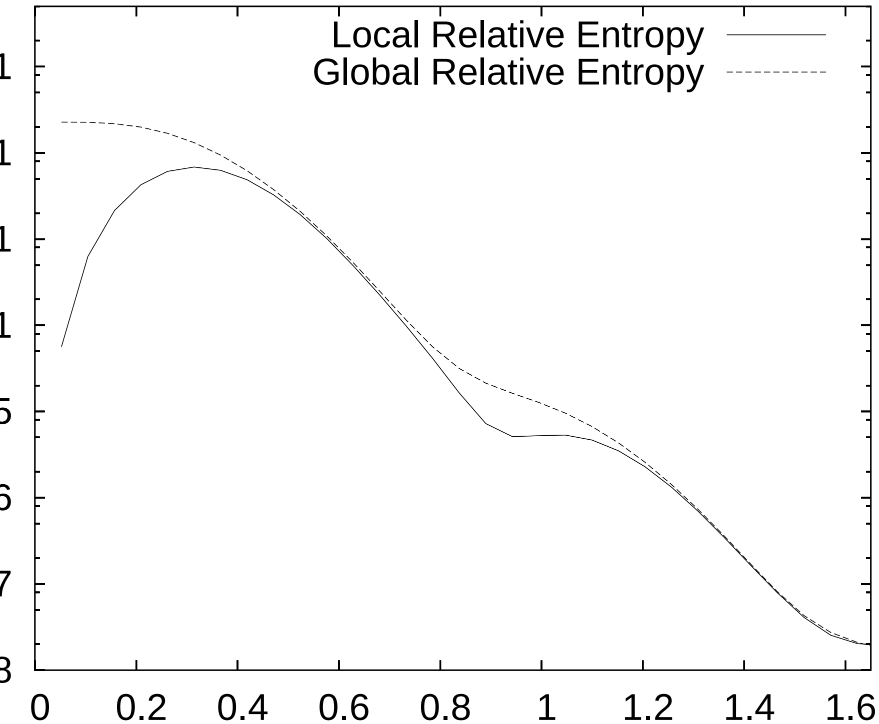
- Instability of hydrodynamical description (fluid mechanics input)

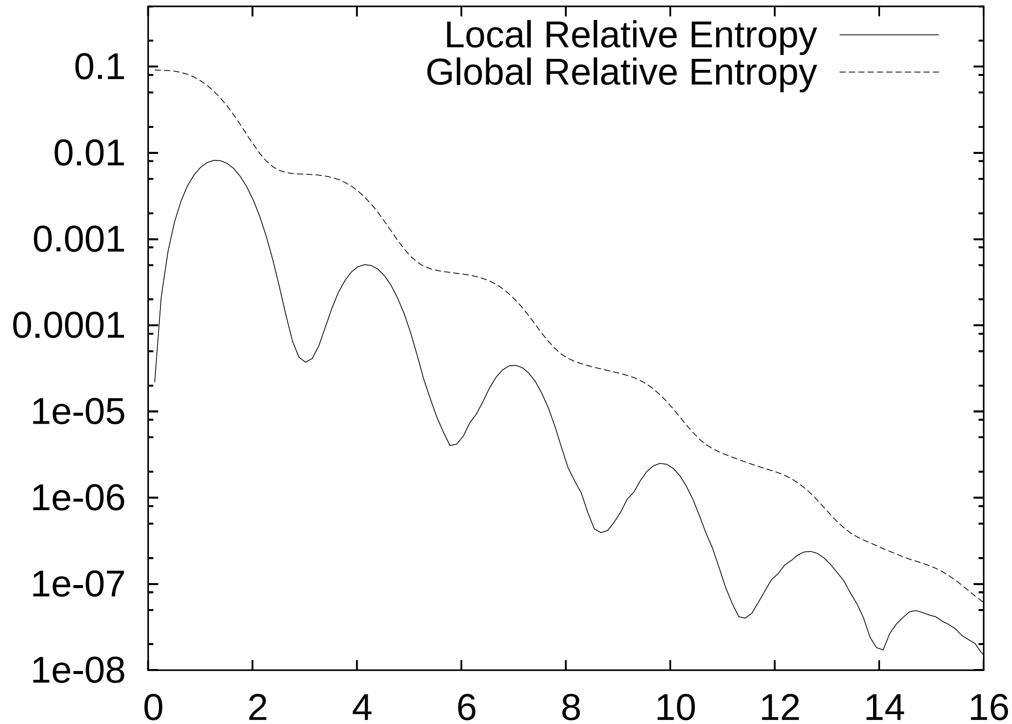
- decomposition of entropy in kinetic and hydrodynamic parts

- functional inequalities coming from various fields (information theory, quantum field theory, elasticity theory, etc.)

It led to the discovery of oscillations in the entropy production

Numerical simulations by Filbet





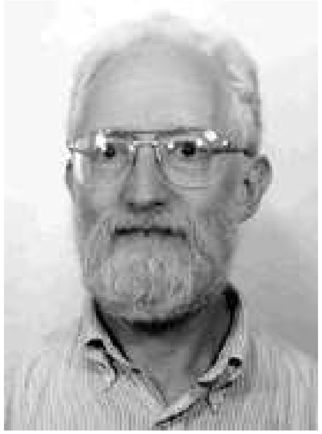
These oscillations of the entropy production slow down the convergence to equilibrium (fluid mechanics effect)

and are related to current research on the convergence for nonsymmetric degenerate operators (“**hypocoercivity**”)

A final point to the Boltzmann controversy??

NOT YET

(a) The Boltzmann equation has still not been rigorously established, except for a rare cloud in the whole space (Lanford 1973, Illner–Pulvirenti 1986)



(b) We **DON'T KNOW** how to prove the *H* Theorem!!

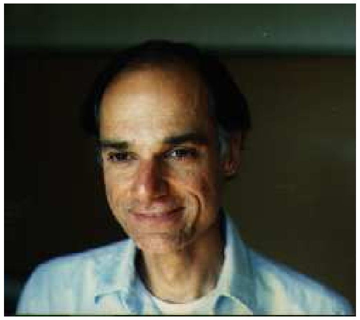
"Slight analytical difficulty": the existence of smooth solutions, known only in particular cases.

$1,000,000 problem

Take a solution of the incompressible Navier–Stokes equation, assume it is smooth initially, does it remain smooth??

For Boltzmann equation

Same problem!



In 1989 DiPerna and P.-L. Lions proved the existence and stability of solutions of the Boltzmann equation

... but nothing about the smoothness!

The proof of approach to equilibrium needs it!!

Some major difficulties may still be untouched

Meanwhile, the entropy story goes on...

Two further unexpected examples:

- Central limit theorem
- Ricci curvature bounds

The central limit theorem

$X_1, X_2, \ldots, X_n, \ldots$ identically distributed, independent real random variables;

$$\mathbb{E}X_j^2 < \infty, \quad \mathbb{E}X_j = 0$$

Then

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

Ball–Barthe–Naor (2004): Irreversible loss of information

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

(some earlier results: Linnik, Barron, Carlen–Soffer)

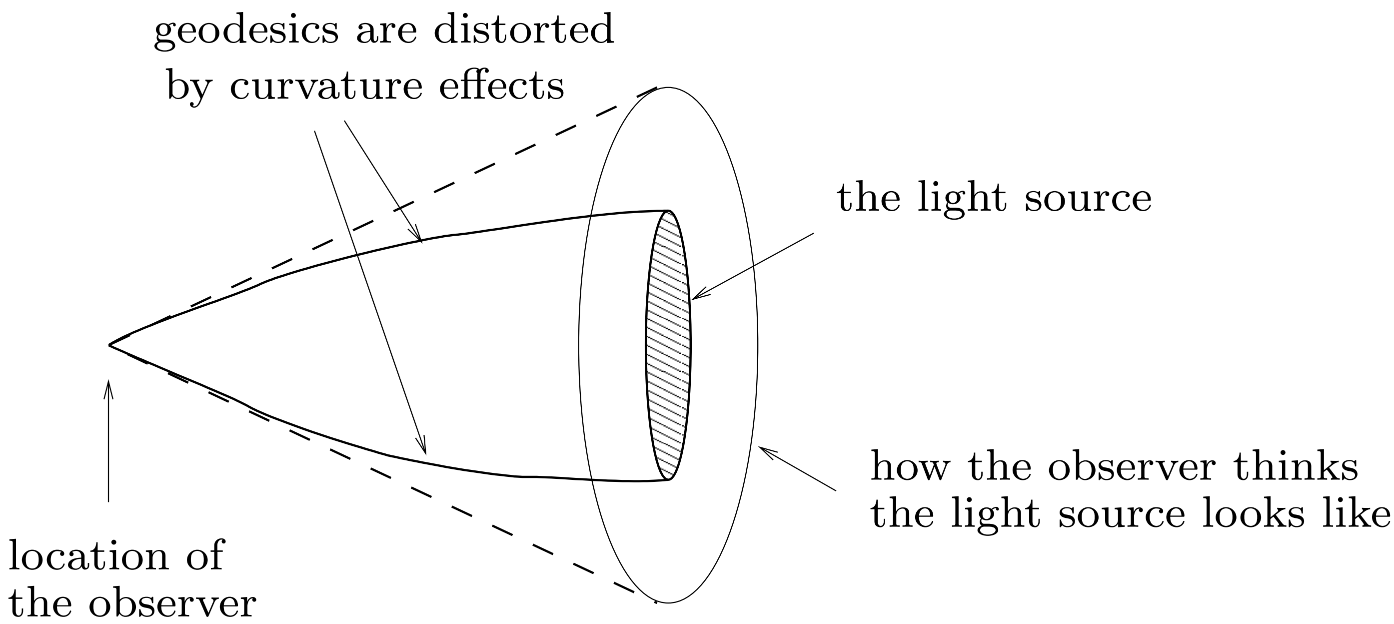
Ricci curvature

One of the three most popular notions of curvature. At each x, Ric is a quadratic form on T_xM.

$(\mathrm{Ric})_{ij} = (\mathrm{Riem})^k_{kij}$

It measures the rate of separation of geodesics in a given direction, in the sense of volume (Jacobian)

Distortion



Because of positive curvature effects, the observer overestimates the surface of the light source; in a negatively curved world this would be the contrary.

$$[\,\text{Distortion coefficients always} \geq 1\,] \iff [\,\text{Ric} \geq 0\,]$$

Lower bounds on Ricci curvature are of constant use

- Isoperimetric inequalities
- Heat kernel estimates
- Sobolev inequalities
- Diameter control
- Spectral gap inequalities
- Volume growth
- Compactness of families of manifolds
- etc.

The connection between optimal transport of probability measures and Ricci curvature was recently studied

(McCann, Cordero-Erausquin, Otto, Schmuckenschläger, Sturm, von Renesse, Lott, V)

A theorem obtained by these tools (Lott–V; Sturm):

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

Limit in which sense? Measured Gromov–Hausdorff topology (very weak: roughly speaking, ensures convergence of distance and volume)

A key ingredient: Boltzmann's entropy!

The lazy gas experiment

Describes the link between Ricci and optimal transport