OPTIMAL TRANSPORTATION, DISSIPATIVE PDE'S AND FUNCTIONAL INEQUALITIES

C. VILLANI

Recent research has shown the emergence of an intricate pattern of tight links between certain classes of optimal transportation problems, certain classes of evolution PDE's and certain classes of functional inequalities. It is my purpose in these notes to convey an idea of these links through (hopefully) pedagogical examples taken from recent works by various authors. During this process, we shall encounter such diverse areas as fluid mechanics, granular material physics, mean-field limits in statistical mechanics, and optimal Sobolev inequalities.

I have written two other texts dealing with mass transportation techniques, which may complement the present set of notes. One [41] is a set of lectures notes for a graduate course taught in Georgia Tech, Atlanta; the other one [40] is a short contribution to the proceedings of a summer school in the Azores, organized by Maria Carvalho; I have tried to avoid repetition. With respect to both abovementioned sources, the present notes aim at giving a more impressionist picture, with priority given to the diversity of applications rather than to the systematic nature of the exposition. The plan here is the opposite of the one that you would expect in a course: it starts with applications and ends up with theoretical background. There is a lot of overlapping with the proceedings of the Azores summer school, however the latter was mainly focusing on the problem of trend to equilibrium for dissipative equations. Most of the material in sections III and IV is absent from the Atlanta lecture notes.

I chose not to start by giving precise definitions of mass transportation; in fact, each lecture will present a slightly different view on mass transportation.

It is a pleasure to thank the CIME organization for their beautiful work on the occasion of the summer school in Martina Franca, in which these lectures have been given. I also thank Yann Brenier for his suggestions during the preparation of these lectures.

Contents

Let me start by stating several problems which seem loosely related, and do not seem to have anything to do with mass transportation.

Problem # 1 (rate of convergence as $t \to \infty$): Consider the diffusive, nonlinear partial differential equation where the unknown $(f_t)_{t\geq 0}$ is a time-dependent probability density on $\mathbb{R}^d$,

$$\frac{\partial f}{\partial t} = \sigma \Delta f + \nabla \cdot (f \nabla V) + \nabla \cdot (f \nabla (f * W)), \qquad t \geq 0,\ x \in \mathbb{R}^d \tag{1}$$

(here ∇ stands for the gradient operator in $\mathbb{R}^d$, while $\nabla \cdot$ is its adjoint, the divergence operator; moreover, V and W are smooth potentials). Does this equation admit a stationary state ? If the answer is yes, do solutions of (1) converge towards this stationary state ?

Problem # 2 (rate of convergence as $N \to \infty$): Consider a bunch of N particles in $\mathbb{R}^d$, with respective positions X_t^i $(1 \leq i \leq N)$ at time $t \geq 0$, solutions of the stochastic differential equation

$$dX_t^i = dB_t^i - \nabla V(X_t^i)\, dt - \frac{1}{N} \sum_{j=1}^N \nabla W(X_t^i - X_t^j)\, dt, \qquad 1 \leq i \leq N,$$

starting from some chaotic initial configuration $X_0 = (X_0^1, \ldots, X_0^N)$ with law $\mathcal{L}(X_0) = f_0^{\otimes N}\, dx$ on $(\mathbb{R}^d)^N$. Under some assumptions on V and W it is known that for each time $t > 0$ the density of particles in $\mathbb{R}^d$ associated with this system converges, as $N \to \infty$, towards the unique solution of (1). This means for instance, that for any bounded continuous function φ,

$$\frac{1}{N} \sum_{i=1}^N \varphi(X_t^i) \xrightarrow[N \to \infty]{} \int_{\mathbb{R}^d} f_t\, \varphi\, dx,$$

in the sense of convergence in law for random variables (note that the left-hand side is random, while the right-hand side is not). Can one estimate the speed of convergence as $N \to \infty$?

Problem # 3 (optimal constants): What is the optimal constant, and how do minimizers look like, in the Gagliardo-Nirenberg interpolation inequality

$$\|w\|_{L^p(\mathbb{R}^n)} \leq C \|\nabla w\|_{L^2(\mathbb{R}^n)}^{\theta} \|w\|_{L^q(\mathbb{R}^n)}^{1-\theta} \tag{2}$$

(with some compatibility conditions on p, q) ? What about the Young inequality

$$\|f * g\|_{L^r(\mathbb{R}^n)} \leq C \|f\|_{L^p(\mathbb{R}^n)} \|g\|_{L^q(\mathbb{R}^n)} \qquad ?$$

All three problems have been studied by various authors and solved in certain cases. Some of them are quite old, like the study of optimal Young inequalities, which goes back to the seventies. Among recent works, let me mention Carrillo, McCann and Villani [12] for problem 1; Malrieu [24] for problem 2; Dolbeault and del Pino [19] for the Gagliardo-Nirenberg inequality in problem 3; Barthe [4] for the Young inequality in the same problem. It turns out that in all these cases, either optimal mass transportation was explicit in the solution, or it has been found to provide much more transparent proofs than those which have been first suggested. My goal here is not to explain *why* mass transportation is efficient in such problems, but to give an idea of *how* it can be used in these various contexts.

II. A study of fast trend to equilibrium

Let us start with the study of equation (1). We shall only consider a particular case of it :

$$\frac{\partial f}{\partial t} = \Delta f + \nabla \cdot (f \nabla (f * W)), \tag{3}$$

where W is a strictly convex, symmetric ($W(-z) = W(z)$) potential, growing superquadratically at infinity. Note that equation (3) has two conservation laws, the total mass of particles and the center of mass, $\int x f(x)\, dx$. The particular case we have in mind is $W(z) = |z|^3/3$, which appears (in dimension 1) in a kinetic modelling of granular material undergoing diffusion in velocity space [5]; the case without diffusion was studied by McNamara and Young [28], Benedetto et al. [6]. On physical grounds, one could also include a term like $\theta \nabla \cdot (fx)$ ($\theta > 0$) in the right-hand side of (3); this would in fact simplify the analysis below. Also, as we shall see the particular form of W does not really matter for us; what matters is its strict convexity.

Of course equation (3) is a diffusion equation, with a nonlinear transport term. However, this classification overlooks the fact that it has a particular structure: it can be rewritten as

$$\frac{\partial f}{\partial t} = \nabla \cdot \Big(f [\nabla \log f + \nabla (f * W)] \Big) \tag{4}$$

$$= \nabla \cdot \left(f \nabla \frac{\delta E}{\delta f} \right), \tag{5}$$

where

$$E(f) = \int f \log f + \frac{1}{2} \int_{\mathbb{R}^d \times \mathbb{R}^d} f(x) f(y)\, W(x-y)\, dx\, dy,$$

and $\delta E/\delta f$ stands for the gradient of E with respect to the standard L^2 Hilbert structure. In the sequel, we shall call E the "free energy" (meaning that it is some combination of an internal energy, or entropy term, and of an interaction energy).

As a consequence of (5), we immediately see that, if we set rigorous justification aside,

$$\frac{d}{dt} E(f_t) = \int \frac{\delta E}{\delta f} \nabla \cdot \left(f \nabla \frac{\delta E}{\delta f} \right) = - \int f \left| \nabla \frac{\delta E}{\delta f} \right|^2 \leq 0.$$

It is therefore natural to expect that $f(t, \cdot) = f_t$ will converge towards the minimizer f_∞ of the free energy E. Here we implicitly assume the existence and uniqueness of this minimizer (in the class of probability measures with fixed center of mass), which was proven by McCann [27].

Using a standard convexity method, which happens to work in this context only for the particular case $d = 1$ and for a quadratic or cubic interaction, Benedetto et al [5] proved convergence of f_t towards f_∞ as $t \to \infty$. No rate of convergence was available.

Then mass transportation was input in the problem, via some ideas going back to Otto, and led to spectacular improvements. In a recent work, Carrillo, McCann and Villani [12] prove that the convergence is exponential :

$$\| f_t - f_\infty \|_{L^1} \leq C e^{-\lambda t}, \tag{6}$$

and show how to recover explicit estimates for C and λ in terms of $\int f_0(x)|x|^2\,dx$ only. Moreover, the proof does not need the cubic form of the potential, neither the restriction to dimension 1. It is also a "fully nonlinear" argument, in the sense that it does not use any linearization procedure (which would usually result in the destruction of any hope of estimating C).

Let us give a precise statement, slightly more general.

Theorem 1. *Let $f = (f_t)_{t\geq 0}$ be a solution of*

$$\frac{\partial f}{\partial t} = \Delta f + \nabla\cdot(f\nabla(f*W)),$$

where W is a C^2 symmetric interaction potential satisfying $D^2W(z)\geq K|z|^\gamma$ for some $\gamma>0$, and $|\nabla W(z)|\leq C(1+|z|)^\beta$ for some $\beta\geq 0$ (example : $W(z)=|z|^3/3$). Let

$$E(f) = \int f\log f + \frac{1}{2}\int W(x-y)\,f(x)\,f(y)\,dx\,dy,$$

and let f_∞ be the unique minimizer of F with the same center of mass as f. Then, for any $t_0>0$ there exist constants $C_0,\lambda_0>0$, explicitly computable and depending on f only via an upper bound for $\int f_0(x)|x|^2\,dx$, such that

$$t\geq t_0 \Longrightarrow \|f_t - f_\infty\|_{L^1}\leq C_0 e^{-\lambda_0 t}.$$

How does the argument work ? The first main idea (quite natural) is to focus on the rate of dissipation of energy,

$$\begin{aligned} D(f) &= \int f\left|\nabla\frac{\delta E}{\delta f}\right|^2 \\ &= \int f\left|\nabla(\log f + W*f)\right|^2, \end{aligned}$$

and to look for a **functional inequality** of the form

$$D(f)\geq \text{const.}[E(f)-E(f_\infty)]. \tag{7}$$

Of course, if this inequality holds true, by combining it with the identity $(d/dt)[E(f_t) - E(f_\infty)] = -D(f_t)$ we shall obtain exponential convergence, as desired. An important feature here is that we have temporarily left the world of PDE's to enter that of functional inequalities, and this will allow a lot of flexibility in the treatment.

To better understand the nature of (7), let us imagine for a few moments that there is a linear drift term $\nabla(f\nabla V)$ instead of the nonlinear term $\nabla(f\nabla(W*f))$; and accordingly, replace the interaction part of the free energy by $\int fV$. Then, assuming that V has been normalized in such a way that $\int e^{-V}=1$, inequality (7) turns into

$$\int f|\nabla(\log f)+\nabla V|^2 \geq \text{const.}\left(\int f\log f + \int fV\right). \tag{8}$$

This inequality is well-known, and called a **logarithmic Sobolev inequality** for the reference measure e^{-V}. The terminology logarithmic Sobolev inequality may sound strange, but it is justified by the fact that (8) can be rewritten in the equivalent formulation

$$\int h^2 \log(h^2)\, d\mu \leq \text{const.} \int |\nabla h|^2 \, d\mu + \left(\int h^2 \, d\mu \right) \log \left(\int h^2 \, d\mu \right),$$

with $d\mu = e^{-V}\, dx$, and this last inequality asserts the embedding $H^1(d\mu) \subset L^2 \log L(d\mu)$, a weak, limit case of Sobolev embedding. This embedding is actually **not always** true; it depends on μ. A famous result by Bakry and Emery [3] states that this is indeed the case if V is uniformly convex.

Logarithmic Sobolev inequalities have become a field of mathematics by themselves, after the famous works by Gross in the mid-seventies. The interesting reader can find many references in the surveys [20] (already 10 years old) and [1] (in french). This may seem strange, because they would seem to be weaker than usual Sobolev inequalities. One reason why they are so important to many researchers, is that the constants involved do not depend on the dimension of space, contrarily to standard Sobolev inequalities; thus logarithmic Sobolev inequalities are a substitute to Sobolev inequalities in infinite dimension. As far as we are concerned, they have another interest: they are well-adapted to the study of the asymptotic behavior of diffusion equations, because they behave well at the neighborhood of the minimizer of E (both sides of (8) vanish simultaneously). In view of this, it is natural to look upon (7) as a "nonlinear" generalization of (8).

There are many known proofs of (8), but the great majority of them fail to yield generalizations like (7). However, at the moment we know of two possible ways towards (7). One is the adaptation of the so-called Bakry-Emery strategy, and has been obtained by a reinterpretation of a famous argument from [3]. It is based on computing the *second derivative* of $E(f_t)$ with respect to time, and showing that the resulting functional, call it $DD(f)$, can be compared directly to $D(f)$; then use the evolution equation to convert this result of comparison between DD and D, into a result of comparison between D and E. The other strategy relies on mass transportation, and more precisely the displacement interpolation technique introduced in McCann [27] and further investigated in Otto [32], Otto and Villani [33]. We insist that even if mass transportation is absent from the first argument, it can be seen as underlying it, as explained in [33] or [12].

In the sequel we only explain about the second strategy; we note that, at least in some cases, its implementation can be significantly simplified by a technique due to Cordero-Erausquin [15, 16]. In this point of view, the key property behind (7) is **displacement convexity**. Let us explain this simple, but quite interesting concept.

Let f_0 and f_1 be two probability densities on $\mathbb{R}^d$. To be more intrinsic we should state everything in terms of probability *measures*, but in the present context it is possible to deal with functions. By a theorem of Brenier [9] and McCann [26] and maybe earlier authors, there exists a unique gradient of convex function (unique on the support of f_0), $\nabla\varphi$, which **transports** f_0 onto f_1, in the sense that the image measure of $f_0(x)\, dx$ by $\nabla\varphi$ is $f_1(x)\, dx$. In other words, for all bounded continuous function h on $\mathbb{R}^d$,

$$\int h(x) f_1(x)\, dx = \int h(\nabla\varphi(x)) f_0(x)\, dx.$$

We shall use the notation

$$\nabla\varphi\# f_0 = f_1.$$

Moreover, this map $T = \nabla\varphi$ is obtained as the unique solution of **Monge's optimal mass transportation problem**

$$\inf_{T\# f_0 = f_1} \int f_0(x)|x - T(x)|^2\, dx.$$

Note that this construction naturally introduces a notion of distance between f_0 and f_1. This notion coincides with the Monge-Kantorovich (or Wasserstein) distance of order 2,

$$\begin{aligned} W_2(f_0, f_1) &= \sqrt{\inf_{T\# f_0 = f_1} \int f_0(x)|x - T(x)|^2\, dx} \\ &= \sqrt{\int f_0(x)|x - \nabla\varphi(x)|^2\, dx}. \end{aligned}$$

As noticed by McCann [27], this procedure makes it possible to define a natural "interpolant" $(f_s)_{0\le t\le 1}$ between f_0 and f_1, by

$$f_s = [(1-s)\mathrm{id} + s\nabla\varphi]\# f_0.$$

By definition, a functional E is displacement convex if, whenever f_0 and f_1 are two probability densities, the function $E(f_s)$ is a convex function of the parameter s. To compare this definition with that of convexity in the usual sense, note that the latter can be rephrased as: *for all f_0 and f_1, the function $E((1-s)f_0 + sf_1)$ is a convex function of $s \in [0, 1]$.*

Just as for usual convexity, one can refine the concept of displacement convexity into that of λ-uniform displacement convexity ($\lambda \ge 0$), meaning that

$$\frac{d^2}{ds^2} E(f_s) \ge \lambda W_2(f_0, f_1)^2.$$

Of course 0-uniform displacement convexity is just displacement convexity.

At this point it is useful to give some examples. Researchers working on these questions have focused on three model functionals :

- $E(f) = \int fV$ is λ-displacement convex if $D^2V \ge \lambda\mathrm{Id}$. Here D^2 stands for the Hessian operator on $C^2(\mathbb{R}^d)$.
- $E(f) = \frac{1}{2}\int f(x)f(y)W(x-y)\, dx\, dy$ is λ-uniformly displacement convex if $D^2W \ge \lambda\mathrm{Id}$, and if one restricts to some set of probability measures with fixed center of mass ($\int f_0(x)x\, dx = \int f_1(x)x\, dx$ in the definition). Note that in our context, W is not in general uniformly convex, so that we cannot go beyond displacement convexity.
- $E(f) = \int U(f(x))\, dx$ is displacement convex if $r \mapsto r^d U(r^{-d})$ is convex nonincreasing on $\mathbb{R}_+$. This is in particular the case if $U(r) = r\log r$, as in the present situation.

Now comes the core of the argument towards (7). It consists in using displacement convexity to establish the stronger functional inequality, holding for all probability densities f_0 and f_1 with common center of mass,

$$E(f_0) - E(f_1) \leq \sqrt{D(f_0)} W_2(f_0, f_1) - \frac{K}{2} W(f_0, f_1)^2, \tag{9}$$

for some constant $K > 0$, depending on f_0 and f_1 only via an upper bound on $E(f_0)$ and $E(f_1)$.

How should one think of inequality (9) ? If we see things from a PDE point of view, the left-hand side measures some discrepancy between f_0 and f_1 in some sense which looks slightly stronger than L^1 (something like $L \log L$). On the other hand, the right-hand side involves both a weak control of the distance between f_0 and f_1 (the W_2 terms) and a control of the smoothness of f_0 (the $D(f_0)$ term, which involves gradients of f_0). It can therefore be thought of as an **interpolation inequality**.

The interest of (9) for our problem is clear: by Young's inequality, it implies

$$E(f_0) - E(f_1) \leq \frac{1}{2K} D(f_0),$$

which is precisely what we are looking at if we replace f_1 by the minimizer f_∞ of E. The constant K will only depend on an upper bound on the entropy of $E(f_0)$; we will see later how this assumption can be dispended of in the final results.

On the other hand, as we shall see in a moment, (9) can be proved rather easily by displacement convexity, based on the Taylor formula

$$E(f_1) = E(f_0) + \left.\frac{d}{ds}\right|_{s=0} E(f_s) + \int_0^1 (1-s) \frac{d^2}{ds^2} E(f_s)\, ds, \tag{10}$$

and the estimates

$$\frac{d^2}{ds^2} E(f_s) \geq K W_2(f_0, f_1)^2, \tag{11}$$

$$\left| \left.\frac{d}{ds}\right|_{s=0} E(f_s) \right| \leq \sqrt{D(f_0)} W_2(f_0, f_1). \tag{12}$$

Clearly, the combination of (10), (11) and (12) solves our problem. We mention that the very same strategy was used in Otto and Villani [33] to give a new proof of the Bakry-Emery theorem.

Let us first show how to prove (11) in the case when W is uniformly convex, i.e. there exists some $\lambda > 0$ such that $D^2 W \geq \lambda \mathrm{Id}$. First let us show what was claimed above, namely that in this situation E is λ-uniformly displacement convex when the center of mass is fixed. Let f_0 and f_1 be two probability densities, and $(f_s)_{0 \leq s \leq 1}$ the associated displacement interpolant. By the change of variable formula, f_s satisfies the Monge-Ampère type equation

$$f_0(x) = f_s\big((1-s)x + s\nabla\varphi(x)\big) \det\big((1-s)\mathrm{Id} + sD^2\varphi\big), \tag{13}$$

On the other hand, also by changing variables, we have

$$\int f_s \log f_s = \int f_s((1-s)x + s\nabla\varphi(x)) \log f_s((1-s)x + s\nabla\varphi(x)) \ \det\big((1-s)\mathrm{Id} + sD^2\varphi\big),$$

and combining this with (13), one finds

$$\int f_s \log f_s = \int f_0(x) \log \frac{f_0(x)}{\det\big((1-s)\mathrm{Id} + sD^2\varphi\big)}, \tag{14}$$

$$= \int f_0 \log f_0 - \int f_0 \log\det\big((1-s)\mathrm{Id} + sD^2\varphi\big). \tag{15}$$

A nontrivial result by McCann [27] shows that this procedure is rigorous. Then, one sees that the right-hand side is a convex function of s.

Next, by the definition of image measure,

$$\frac{1}{2}\int f_s(x) f_s(y) W(x-y)\, dx\, dy$$

$$= \frac{1}{2}\int f_0(x) f_0(y) W\big([(1-s)x + s\nabla\varphi(x)] + [(1-s)y + s\nabla\varphi(y)]\big)\, dx\, dy, \tag{16}$$

and from this we see that

$$\frac{d^2}{ds^2}\frac{1}{2}\int f_s(x) f_s(y) W(x-y)\, dx\, dy =$$
$$\frac{1}{2}\int f_0(x) f_0(y) \Big\langle D^2W\big((1-s)(x-y) + s(\nabla\varphi(x) - \nabla\varphi(y))\big)\cdot[\xi(x)-\xi(y)], [\xi(x)-\xi(y)]\Big\rangle\, dx\, dy,$$

where we use the shorthand $\xi(x) = x - \nabla\varphi(x)$. From our assumption on D^2W, this is bounded below by

$$\frac{\lambda}{2}\int f_0(x) f_0(y) |\xi(x) - \xi(y)|^2\, dx\, dy,$$

which turns out to be precisely

$$2\frac{\lambda}{2}\int f_0(x)|\xi(x)|^2\, dx = \lambda W_2(f_0, f_1)^2,$$

in view of the identity

$$\int f_0(x)\xi(x)\, dx = \int f_0(x) x\, dx - \int f_0(x)\nabla\varphi(x)\, dx$$
$$= \int f_0(x) x\, dx - \int f_1(y) y\, dy = 0.$$

To summarize, at this point we have shown (11) under the restrictive assumption that W be uniformly convex. What if W behaves like the cubic potential ? For instance, let us assume $D^2W(z) \geq \psi(|z|)$, where ψ is continuous and vanishes at $z = 0$. Then we can treat this case just as the previous one, with the help of the following unexpected lemma (certainly suboptimal):

$$\int f_0(x)f_0(y)\psi(|x-y|)|\xi(x)-\xi(y)|^2\,dx\,dy \geq K(f_0)\int f_0(x)f_0(y)|\xi(x)-\xi(y)|^2\,dx\,dy,$$

$$K(f_0) = \frac{1}{8}\inf_{z_1,z_2\in\mathbb{R}^d}\int f_0(x)\inf[\psi(|x-z_1|),\psi(|x-z_2|)]\,dx.$$

Note that $K(f_0)$ is strictly positive as soon as f_0 is a probability density; and can be estimated from below in terms of an upper bound on $E(f_0)$.

Now, let us sketch the proof of (12). From (15) and (16), it is not difficult to compute explicitly

$$\frac{d}{ds}\bigg|_{s=0} E(f_s) = -\int f_0(x)[\Delta\varphi(x)-d] + \int f_0(x)f_0(y)\nabla W(x-y)\cdot[\nabla\varphi(x)-x],$$

and after a few manipulations this can be rewritten as

$$\int f_0(x)\nabla(\log f_0 + f_0 * W)(x)\cdot[\nabla\varphi(x)-x]\,dx.$$

Then, by Cauchy-Schwarz inequality,

$$\left|\frac{d}{ds}\bigg|_{s=0} E(f_s)\right| \leq \sqrt{\int f_0|\nabla(\log f_0 + f_0 * W)|^2}\sqrt{\int f_0(x)|\nabla\varphi(x)-x|^2\,dx},$$

which is precisely (12). This ends our study of (7).

Let us conclude this section by explaining how one deduces (6) from (7). This will be the occasion to enter back into the PDE world... We shall just sketch the main steps, without entering their proof.

First, as a consequence of (5) and of the displacement convexity of E, one can prove [34] the estimate

$$E(f_t) \leq \frac{W_2(f_0,f_\infty)^2}{4t} + E(f_\infty). \tag{17}$$

Note that one can interpretate this inequality as a parabolic regularization inequality: it states that the size of f_t in $L\log L$ (measured by $E(f_t)$) becomes finite like $O(1/t)$ as t becomes positive.

Since $E(f_{t_0})$ is finite for any $t_0 > 0$, and since $E(f_t)$ is uniformly bounded by $E(f_{t_0})$ for $t \geq t_0$, it is then possible, for $t \geq t_0$, to apply (7) and deduce that

$$\frac{d}{dt}[E(f_t)-E(f_\infty)] \leq -\text{const.}[E(f_t)-E(f_\infty)].$$

From this one of course concludes that

$$E(f_t) - E(f_\infty) = O(e^{-\mu t})$$

for some $\mu > 0$.

Next, another functional inequality due to Carrillo, McCann and Villani [12] states that

$$W_2(f, f_\infty) \le \sqrt{\text{const.}[E(f) - E(f_\infty)]}.$$

In the case when the interaction energy is replaced by the simpler energy $\int fV$, with a uniformly convex V, then there is also a similar inequality, proven by Otto and Villani [33], which is a generalization of a well-known inequality by Talagrand [36]. At this point we are able to conclude that

$$W_2(f_t, f_\infty) = O(e^{-\frac{\mu}{2}t}). \tag{18}$$

The game now is to interpolate this information with adequate smoothness bounds, to gain convergence of $\|f_t - f_\infty\|_{L^1}$. For this, one can first prove that

$$D(f_t) \le \frac{W_2(f_t, f_\infty)^2}{t^2},$$

exactly in the same spirit as (17). From this follows a uniform bound on $D(f_t)$ for $t \ge t_0 > 0$. Combining this with the easy inequality

$$\int f \left| \nabla \log \frac{f}{f_\infty} \right|^2 \le C[D(f) + E(f)],$$

one finds that

$$\sup_{t \ge t_0} \int f_t \left| \nabla \log \frac{f_t}{f_\infty} \right|^2 < +\infty. \tag{19}$$

Estimate (19) is precisely what will play the role of smoothness bound for f_t. From the convexity of $-\log f_\infty$, one can prove an interpolation inequality similar to (9),

$$\int f \log \frac{f}{f_\infty} \le C W_2(f, f_\infty) \sqrt{\int f_t \left| \nabla \log \frac{f_t}{f_\infty} \right|^2}.$$

Combining this inequality with (18) and (19), we obtain

$$\int f_t \log \frac{f_t}{f_\infty} = O(e^{-\frac{\mu}{2}t}).$$

To conclude the argument it suffices to recall the classical Csiszár-Kullback-Pinsker inequality, in the form

$$\|f - f_\infty\|_{L^1} \le \sqrt{2 \int f \log \frac{f}{f_\infty}}.$$

In the previous section, I have explained how to use some of the mass transportation formalism in order to prove fast trend to equilibrium. In this section, I will present another example in which mass transportation enables one to prove that convergence to equilibrium has to be slow, in some sense. Moreover, we shall see that mass transportation here has a remarkable role of helping physical intuition. All of the material in this section is taken from a recent work by Caglioti and the author.

Let us start with the general definition of **Monge-Kantorovich distances**. Let $d(x,y) = |x-y|$ be the Euclidean distance (or any other continuous distance) on $\mathbb{R}^d$, and let $p \geq 1$ be a real number. Whenever μ and ν are two probability measures on $\mathbb{R}^d$, one can define their Monge-Kantorovich, or Wasserstein, distance of order p,

$$W_p(\mu,\nu) = \inf \left\{ \left[\int d(x,y)^p \, d\pi(x,y) \right]^{1/p} ; \quad \pi \in \Pi(\mu,\nu) \right\},$$

where $\Pi(\mu,\nu)$ stands for the set of all probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν. More explicitly, $\pi \in \Pi(\mu,\nu)$ if and only if, whenever φ, ψ are continuous bounded functions on $\mathbb{R}^d$,

$$\int [\varphi(x) + \psi(y)] \, d\pi(x,y) = \int \varphi \, d\mu + \int \psi \, d\nu.$$

The Monge-Kantorovich distances have been used in various contexts in statistical mechanics, going back at least to the works of Dobrushin, and Tanaka, in the seventies. Tanaka [37, 29, 38] noticed that W_2 is a *nonexpansive* distance along solutions of the spatially homogeneous Boltzmnan equation for Maxwell molecules (see [41] for a review). Even if this is a very special case of Boltzmann equation, his remark led to interesting progress and a better understanding of trend to equilibrium in this context.

Here I shall explain about another use of the Wasserstein distances in kinetic theory: for a very simple model of granular media, closely related to the example of the previous section.

First, some very sketchy background about granular media. This field has become extremely popular over the last decades, and there are dozens of models for it. Some people use kinetic models (in phase space, i.e. position and velocity) based on **inelastic** collisions, i.e. allowing deperdition of energy. In particular, inelastic Boltzmann or inelastic Enskog equations are used for this. From a mathematical point of view, these are extremely complicated models, about which essentially nothing is known (see the references in [14] and in [39]). Among basic PDE's from classical fluid mechanics, the standard Boltzmann equation is usually considered as a terribly difficult one, but the inelastic one seems to be even ways ahead in this respect.

In particular, the asymptotic behavior of solutions to the inelastic Boltzmann equation constitutes a formidable problem. There is no H theorem in this setting, and natural asymptotic states would have zero temperature, i.e. all the particles at a given position should travel at the same speed. In particular, the corresponding density of particles would be very singular. It is unknown whether concentration phenomena may occur in finite time, and it is debated whether there is a relevant hydrodynamic scaling. A key element of this debate seems to be the existence of "**homogeneous cooling states**", i.e. particular spatially homogeneous, self-similar solutions converging to a dirac mass as $t \to \infty$, taking the form

$$f_S(t,v) = \frac{1}{\alpha^N(t)} F\left(\frac{v - v_0}{\alpha(t)}\right).$$

It is believed by some authors that such states do exist and "attract" (locally in x) all solutions. Thus they would play the same role in this context, than local thermodynamical equilibria (Maxwellian distributions) do in the classical setting.

From a mathematical point of view, what can be said about homogeneous cooling states ? We begin with some bad news: Bobylev, Carrillo and Gamba [8] have shown, for a simplified model ("inelastic Maxwell molecules") that there does not in general exist homogeneous cooling states. Only in some restricted, weaker meaning can the concept of universal asymptotic profile be salvaged.

In this lecture we consider a *considerably* simplified model, introduced by McNamara and Young [28]. It is one-dimensional and spatially homogeneous; so the unknown is a time-dependent probability measure on $\mathbb{R}$, to be thought of as a velocity space. It reads

$$\frac{\partial f}{\partial t} = \frac{\partial}{\partial x}\left(f \frac{\partial}{\partial x}(f * W)\right), \qquad W(z) = \frac{|z|^3}{3}. \tag{20}$$

In the sequel, we shall often write $f_t(x)$ for the solution of (20), even if this solution should be thought of as a measure rather than as a density.

In fact, we already encountered in the previous lecture a diffusive variant of this model. The degree of simplification leading from the inelastic Boltzmann equation to (20) can probably be compared to that which leads from the compressible Navier-Stokes equation to Burgers' equation.

Let us consider the problem of asymptotic behavior for (20). Without loss of generality, we assume that the mean velocity is 0,

$$\int f_t(x)\, x\, dx = 0.$$

By elementary means it is easy to show

$$\frac{d}{dt} \int f_t(x)|x|^2\, dx \le -C \left(\int f_t(x)|x|^2\, dx\right)^{3/2},$$

and from this one deduces

$$\sqrt{\int f_t(x)|x|^2\, dx} = O\left(\frac{1}{t}\right)$$

as $t \to \infty$. It turns out that this rate is optimal. Note that

$$\sqrt{\int f_t(x)|x|^2\, dx} = W_2(f_t\, dx, \delta_0).$$

In this oversimplified case, is there a self-similar solution ? After determining the scaling invariance of the equation, it is natural to set

$$\tau = \log t, \qquad g(\tau, x) = t f(t, tx) = e^{\tau} f(e^{\tau}, e^{\tau} x). \tag{21}$$

A few lines of computation lead to an equation on the new density g,

$$\frac{\partial g}{\partial \tau} = \frac{\partial}{\partial x}\left(g \frac{\partial}{\partial x}\left(g * W - \frac{x^2}{2}\right)\right). \tag{22}$$

So, again we recognize the same familiar structure as in the previous lecture,

$$\frac{\partial g}{\partial \tau} = \frac{\partial}{\partial x}\left[g \frac{\partial}{\partial x} \frac{\delta E}{\delta g}\right],$$

where now

$$E(g) = \frac{1}{2}\int g(x)g(y)\frac{|x-y|^3}{3}\,dx\,dy - \int g(x)\frac{|x|^2}{2}\,dx. \tag{23}$$

This functional E admits a whole family of critical points, but one unique minimizer, which is a singular measure,

$$g_\infty = \frac{1}{2}\left(\delta_{-\frac{1}{2}} + \delta_{\frac{1}{2}}\right).$$

A nontrivial theorem by Benedetto, Caglioti and Pulvirenti [6] states that if the initial datum g_0 is absolutely continuous with respect to Lebesgue measure, then g_τ converges (in weak measure sense) towards g_∞ as $\tau \to \infty$. In particular, for all $p \geq 1$ one has

$$W_p(g_\tau, g_\infty) \xrightarrow[\tau\to\infty]{} 0.$$

Note that in this context the use of Wasserstein distance is very natural, since the equilibrium is singular. It would *not* be possible to use L^1 distance, as in the previous lecture.

To g_∞ is associated a self-similar solution of (20), via (21). It reads

$$S_t = \frac{1}{2}\left(\delta_{-\frac{1}{2t}} + \delta_{\frac{1}{2t}}\right). \tag{24}$$

Moreover,

$$W_p(f_t, S_t) = \frac{W_p(g_\tau, g_\infty)}{t},$$

which shows that

$$W_p(f_t, S_t) = o\left(\frac{1}{t}\right). \tag{25}$$

Since on the other hand, $W(f_t, \delta_0) = O(1/t)$, we see that (24) indeed plays the role of a homogeneous cooling state.

Now can one get an estimate of how better is the approximation obtained by replacing δ_0 by S_t ? This amounts to study the *rate of convergence for the rescaled problem.* So a natural question is whether there is *exponential convergence to equilibrium* for solutions of (22). And in view of our study of the previous lecture, we could try a functional approach of this problem, by looking for a functional inequality of the form

$$\int g \left| \frac{\partial}{\partial x} \left(g * W - \frac{x^2}{2} \right) \right|^2 dx \geq \text{const.}[E(g) - E(g_\infty)].$$

It turns out that such an inequality is **false**, and this has to do with the negative sign in (23). If one thinks in terms of usual convexity, then the minus sign seems no worse than the positive sign, but if one thinks in terms of *displacement convexity*, then we see that the impact of this change is dramatic. Indeed, the functional $f \mapsto \int f(x)|x|^2\, dx$ is uniformly displacement convex, while its negative is uniformly displacement concave.

In fact, one can prove that **there is no exponential convergence** at the level of (22). A recent result by Caglioti and the author establishes the estimate

$$\int_0^\tau W_p(g_s, g_\infty)\, ds \geq K \log \tau, \tag{26}$$

for some constant $K > 0$ depending on the initial datum. This inequality holds true as soon as g_0 (or f_0) is distinct from a convex combination of two symmetric dirac masses. As a consequence,

$$\int_0^T W_p(f_t, S_t)\, dt \geq K \log\log T,$$

which shows that "morally" $W_p(f_t, S_t)$ should not decrease to 0 faster than $1/(t \log t)$. This means that the improvement which one obtains when replacing δ_0 by S_t is at most logarithmic in time, hence very, very poor.

Again, here below is a precise theorem.

Theorem 2. *Let g_0 be a probability measure on $\mathbb{R}$, which is not a symmetric convex combination of two delta masses, and let g_τ be the corresponding solution to* (22)*. Then, for all $p \in [1, +\infty)$,*

$$\int_0^{+\infty} W_p(g_\tau, g_\infty)\, d\tau = +\infty. \tag{27}$$

More precisely, there exists some constant $K > 0$, depending on g_0, such that, as $\tau \to \infty$,

$$\int_0^\tau W_p(g_s, g_\infty)\, ds \geq K \log \tau. \tag{28}$$

Corollary 3. *Let f_0 be a probability measure on $\mathbb{R}$, which is not a symmetric convex combination of two delta masses, and let f_t be the corresponding solution to* (20)*. Then, for all $p \in [1, +\infty)$,*

$$\int_0^{+\infty} W_p(f_t, S_t)\, dt = +\infty. \tag{29}$$

More precisely, there exists some constant $K > 0$, depending on f_0, such that, as $T \to \infty$,

$$\int_0^T W_p(f_t, S_t)\, dt \geq K \log\log T. \tag{30}$$

In the rest of this lecture, I will try to convey an idea of how the proof works. One striking feature of it is that the introduction of the Wasserstein distance leads almost by itself to the solution. The key properties of the problem are (a) the presence of **vacuum** and **singularities** in the asymptotic state (one of these properties would suffice), (b) the structure of **nonlinear transport equation** in (22) (here we shall take advantage of the nonlinearity). The basic idea, stated informally, is that, due to the presence of vacuum and singularities, it requires a lot of work from the solution to approach the asymptotic state, which can be expressed by the strength of some velocity field; but the velocity field which drives the particles is coupled to the density, and approaches zero when the density approaches the asymptotic state. All in all, the convergence cannot be fast.

In the argument, let us assume for simplicity that f_0, and therefore g_0, is absolutely continuous with respect to Lebesgue measure (a property which is preserved by the flow). Rewrite equation (22) as

$$\frac{\partial g}{\partial \tau} + \frac{\partial}{\partial x}\big(g\,\xi[g]\big) = 0, \tag{31}$$

$$\xi[g](x) = x - \int g(y)(x-y)|x-y|\,dy. \tag{32}$$

Note that $\xi[g]$ lies in C^1 without any regularity assumption on g. Thus we can appeal to the usual theory of characteristics. If we introduce the vector fields T_τ solutions of

$$\frac{d}{d\tau}T_\tau(x) = \xi[g_\tau]\circ T_\tau(x),$$

we know from (31) that

$$g_\tau = T_\tau \# g_0. \tag{33}$$

Next, a remark which will simplify the analysis is that the ordering of particles is preserved. This is a general property of one-dimensional transport equations: two characteristic curves cannot cross because they are integral curves of a vector field. In particular, the **median** of the particles is preserved : if m stands for a median of the probability density g_0 at $\tau = 0$, then $m_\tau = T_\tau(m)$ is a median of g_τ. For simplicity, let us assume that there is just one median; then the density g_0 is strictly positive around its median m, in the sense that the interval $[m-\varepsilon, m+\varepsilon]$ carries a positive mass, for any $\varepsilon > 0$.

A fundamental quantity in a transport equation is the divergence of the velocity field. It provides a kind of measure on how fast the trajectories of the particles diverge. In dimension 1, this is just the derivative of the velocity field, and it is an easy exercice to deduce from (32) that $\xi[g]$ achieves its maximum on the set of medians of g (just differentiate (32) twice). In particular, $\xi[g_\tau]$ achieves its maximum precisely at m_τ.

We are now ready to explain the core of the proof of (26). To get a better intuition of what is going on, the reader is encouraged to draw pictures of particle trajectories. In order to converge towards the equilibrium state g_∞ as $\tau \to \infty$, it is necessary that particles which were initially infinitesimally close to each other, but located on different sides of the median, become asymptotically very far apart from each other (close to $-1/2$

or to $+1/2$). This is possible only if the time-integral of the divergence of the velocity field diverges as $\tau \to \infty$. More precisely, if $a_- < m < a_+$ and $a_\pm(\tau) = T_t(a_\pm)$, then

$$\begin{aligned}\frac{d}{d\tau}[a_+(\tau) - a_-(\tau)] &= \xi[g_\tau](a_+(\tau)) - \xi[g_\tau](a_-(\tau)) \\ &\leq \left\| \frac{d\xi[g_\tau]}{dx} \right\|_\infty [a_+(\tau) - a_-(\tau)],\end{aligned}$$

so that

$$a_+(\tau) - a_-(\tau) \leq [a_+ - a_-] \exp\left(\int_0^\tau \left\| \frac{d\xi[g_s]}{dx} \right\|_\infty ds \right). \tag{34}$$

If we let $\tau \to \infty$, then from $a_-(\tau) \to -1/2$ and $a_+(\tau) \to +1/2$, we will deduce that

$$\exp\left(\int_0^\infty \left\| \frac{d\xi[g_s]}{dx} \right\|_\infty ds \right) \geq \frac{1}{[a_+ - a_-]},$$

and then by letting $a_+ - a_-$ go to 0, we find

$$\int_0^\infty \left\| \frac{d\xi[g_s]}{dx} \right\|_\infty ds = +\infty, \tag{35}$$

as announced.

But condition (35) will be extremely hard to achieve, because $\xi = 0$ at equilibrium. In fact, one has the continuity estimate, in Wasserstein distance,

$$\left\| \frac{d\xi_\tau}{dx} \right\|_\infty \leq 2W_1(g_\tau, g_\infty), \tag{36}$$

which implies that the divergence of ξ has to be all the smaller that one is close to equilibrium.

The proof of (36) is almost self-evident: on one hand, by direct computation,

$$\frac{d\xi_\tau}{dx} = 2\int_{x \leq m_\tau} g_\tau(x)\left(x + \frac{1}{2}\right) dx - 2\int_{x \geq m_\tau} g_\tau(x)\left(x - \frac{1}{2}\right) dx;$$

on the other hand,

$$W_1(g_\tau, g_\infty) = \int_{x \leq m_\tau} g(x)\left|x + \frac{1}{2}\right| dx + \int_{x \geq m_\tau} g(x)\left|x - \frac{1}{2}\right| dx.$$

The combination of (35) and (36) implies

$$\int_0^\infty W_1(g_\tau, g_\infty)\, ds = +\infty,$$

which is already an indication of slow decay to equilibrium. Under the assumption that g_0 is bounded below close to its median, it is possible to push the reasoning further in order to obtain the more precise result mentioned earlier. For this one just has to take advantage once more of the definition of the Monge-Kantorovich distance, and to establish that

$$W_1(g_\tau, g_\infty) \geq \frac{1}{2}\left(\int_{a_-(\tau)}^{a_+(\tau)} g_\tau\right)[1 - (a_+(\tau) - a_-(\tau))]$$
$$= \frac{1}{2}\left(\int_{a_-}^{a_+} g\right)[1 - (a_+(\tau) - a_-(\tau))].$$

Combining this with (34) and choosing $a_+ - a_-$ to be of the order of $W_1(g_\tau, g_\infty)$, one obtains a Gronwall-type inequality on this Wasserstein distance, which implies (26).

Note that here we have only taken advantage of the presence of vacuum in the support of the asymptotic state, not of the presence of singularities. Taking advantage of these singularities is slightly more complicated, but leads to (26) under more general assumptions, since one does not need to assume that the median is unique. The spirit is quite the same as before: if the support of the solution is made of more than just two points, this means that part of the particles will have distinct trajectories converging towards a single point, which is possible only if the time-integral of the divergence of the velocity field diverges to $-\infty$. Then this enters in conflict with the fact that the velocity field vanishes at equilibrium, and one can more or less repeat the same reasoning as above.

In this lecture, I will explain about the use of some estimates from the theory of **concentration of measure** in some mean-field limit problems. The ideas below have been developed by a number of authors, in particular the group of Bakry and Ledoux in Toulouse. Since I will consider a stochastic microscopic model, I shall use a tiny bit of probabilistic formalism. The standard notations P and E will denote respectively the probability of an event, and the expectation of a measurable function. Whenever μ is a probability measure, F a measurable function and A a measurable set, I shall also use the convenient shorthand $\mu[F \in A] = \mu[\{x;\, F(x) \in A\}]$.

Recall the simple mean-field limit problem which was mentioned in the introduction: a random system of N particles, with respective positions $X_t^i \in \mathbb{R}^d$ ($1 \le i \le N$), obeying the system of stochastic differential equations

$$dX_t^i = dB_t^i - \nabla V(X_t^i)\, dt - \frac{1}{N}\sum_{j=1}^{N} \nabla W(X_t^i - X_t^j)\, dt. \tag{37}$$

The initial positions of the N particles are assumed to be random, independent and identically distributed. Moreover, the $(B_t^i)_{t\ge 0}^{1\le i\le N}$ are N independent Brownian motions in $\mathbb{R}^d$.

Under various assumptions on V, W, one can establish that the empirical measure at time t, which is a random measure, has a deterministic limit as $N \to \infty$:

$$\hat{\mu}_t^N = \frac{1}{N}\sum_{j=1}^{N} \delta_{X_t^i} \xrightarrow[N\to\infty]{} f_t(x)\, dx.$$

Here the convergence as $N \to \infty$ is in the sense of convergence in law of random variables, and for the weak topology of measures. Note that the left-hand side is a random measure, while the right-hand side is deterministic. The convergence can be restated as follows: let φ be a bounded smooth test-function, then, as $N \to \infty$,

$$E\left|\frac{1}{N}\sum_{j=1}^{N} \varphi(X_t^i) - \int_{\mathbb{R}^d} f_t(x)\,\varphi(x)\, dx\right| \xrightarrow[N\to\infty]{} 0.$$

The use of averages ("observables") like $(1/N)\sum \varphi(X_t^i)$ is quite natural. Moreover, the limit density f_t is characterized by its being a solution of the simple McKean-Vlasov equation

$$\frac{\partial f}{\partial t} = \frac{1}{2}\Delta f + \nabla\cdot(f\,\nabla V) + \nabla\cdot(f\,\nabla(f * W)). \tag{38}$$

As shown in Sznitman [35], this property of the empirical measure becoming deterministic in the limit is equivalent to the requirement that the law of the N particles be **chaotic** as $N \to \infty$, which means, roughly speaking, that k particles among N (k fixed, $N \to \infty$) look like independent, identically distributed random variables.

The problem addressed here is to find a quantitative version of this asymptotic behavior of the empirical measure. Let φ be a smooth test-function, can one find a bound on how much the average $(1/N)\sum\varphi(X_t^i)$ deviates from its asymptotic value $\int f_t(x)\varphi(x)\,dx$? More precisely, given some $\varepsilon > 0$, can one estimate

$$P\left(\left|\frac{1}{N}\sum_{i=1}^{N}\varphi(X_t^i) - \int_{\mathbb{R}^d} f_t(x)\varphi(x)\,dx\right| \geq \varepsilon\right) \qquad ?$$

The motivations for this question do not only come from theoretical purposes, but also from numerical simulations and the will to justify the use of particle methods.

This problem is quite reminiscent of Sanov's theorem about large deviations for empirical measures. However, here the deviations are taken with respect to the limit value, so it has more to do with a law of large numbers, than with a large deviations principle. One possible way of attacking this question is the so-called theory of **concentration of measure**.

Let us give a very sketchy and elementary background about this theory, which is often traced back to works by Lévy, Gromov (linked with isoperimetric inequalities), and Milman. Roughly speaking, the main paradigm is as follows: let μ be some nice (to be precised later) reference measure on some measure space; for instance a gaussian measure on $\mathbb{R}^n$. Let A be some set of positive measure, say $\mu[A] \geq 1/2$. Then, in some sense, most points in our space are "not too far" from A. A typical example is that if μ is a gaussian measure on $\mathbb{R}^n$, and

$$A_t \equiv \{x \in \mathbb{R}^n;\ \mathrm{dist}(x, A) \leq t\},$$

then $\mu[A_t] \to 1$ as $t \to \infty$, very quickly; more precisely, there exists $C, c > 0$ such that

$$\mu[A_t] \geq 1 - Ce^{-ct^2}.$$

The best possible rate c is obtained by the study of gaussian isoperimetry. It is important to note that the constants c and C will depend on a lower bound on the covariance matrix of the Gaussian, but not on the dimension n.

An equivalent way of formulating this principle consists in stating that "any Lipschitz function is not far from being constant". More precisely, if μ still denotes some gaussian measure, it is possible to show that whenever φ is a Lipschitz function on $\mathbb{R}^n$ with Lipschitz constant $\|\varphi\|_{\mathrm{Lip}}$, then for all $r > 0$,

$$\mu\big[|\varphi - E_\mu\varphi| \geq r\big] \leq C\exp\left(-\frac{cr^2}{\|\varphi\|_{\mathrm{Lip}}^2}\right). \tag{39}$$

Such principles are particularly useful in large dimension. It is easy to understand why: let φ be a Lipschitz function on $\mathbb{R}^d$, with Lipschitz constant $\|\varphi\|_{\mathrm{Lip}}$, and let

$$\Phi(x^1, \ldots, x^N) = \frac{1}{N}\sum_{i=1}^{N}\varphi(x^i).$$

Then, an application of Cauchy-Schwarz inequality gives

$$\|\Phi\|_{\mathrm{Lip}} \leq \frac{\|\varphi\|_{\mathrm{Lip}}}{\sqrt{N}}. \tag{40}$$

Combining this with (39), we see that when N is large, Φ is sharply concentrated (with respect to μ) around its mean value.

During the last decade, the theory of concentration of measure was spectacularly developed by Talagrand in the framework of product spaces. On this occasion he introduced powerful and elementary induction methods to prove concentration inequalities in extremely general settings. Alternative, "functional" approaches have been developed by various authors, in particular Ledoux and Bobkov. One of their main contribution was to show how certain classes of functional inequalities enabled one to recover concentration inequalities, via more "global" and intrinsic methods (not depending so much on the space being product). All this is very well explained in the review paper by Ledoux [22] (in french) or in his lecture notes [23].

Typical classes of useful inequalities in this context are the logarithmic Sobolev inequalities, or the Poincaré inequalities. Let us recall their definitions.

Definition. *A probability measure μ satisfies a logarithmic Sobolev inequality with constant $\lambda > 0$ if for all $f \in L^1(d\mu)$ such that $f \geq 0$, $\int f\, d\mu = 1$, one has*

$$\int f \log f\, d\mu \leq \frac{1}{2\lambda} \int \frac{|\nabla f|^2}{f}\, d\mu.$$

It satisfies a Poincaré inequality with constant λ if for all $f \in L^2(d\mu)$,

$$\int f^2\, d\mu - \left(\int f\, d\mu\right)^2 \leq \frac{1}{\lambda} \int |\nabla f|^2\, d\mu.$$

It is known that if a probability measure satisfies a logarithmic Sobolev inequality with constant λ, then it also satisfies a Poincaré inequality with the same constant. Probability measures on $\mathbb{R}^n$ that admit a smooth, positive density behaving like $e^{-|x|^\alpha}$ as $|x| \to \infty$ satisfy a logarithmic Sobolev inequality if and only if $\alpha \geq 2$, and a Poincaré inequality if and only if $\alpha \geq 1$.

From the works of Ledoux, Bobkov, Götze and others it is known that if μ satisfies a logarithmic Sobolev inequality, then it satisfies concentration inequalities similar to those which are satisfied by a gaussian measure. Instead of explaining their methods, and in particular Ledoux's adaptation of the so-called "Herbst argument", I shall rather take a more indirect, but maybe more intuitive route towards this result. It will make the Wasserstein distance play a prominent role.

The following theorem is one of the main results of Otto and Villani [33] (the original version of the result contained a minor additional assumption, which was later removed by Bobkov et al. [7] with a very different proof).

Theorem 4. *Assume that μ has a smooth density and satisfies a logarithmic Sobolev inequality with constant $\lambda > 0$. Then it also satisfies the following inequality: for all $f \in L^1(d\mu)$, $f \geq 0$, $\int f\, d\mu = 1$,*

$$W_2(f\mu, \mu) \leq \sqrt{\frac{2}{\lambda} \int f \log f\, d\mu}. \tag{41}$$

In the particular case when μ is a gaussian measure, inequality (41) is due to Talagrand; accordingly we shall call (41) a Talagrand inequality.

Now, here is an idea of the original proof of theorem 4. Let V be defined by $d\mu(x) = e^{-V(x)}\,dx$. Introduce the auxiliary PDE

$$\frac{\partial f}{\partial s} = \Delta f - \nabla V \cdot \nabla f,$$

with initial datum $f(0,\cdot) = f$, and denote by $(f_s)_{s\geq 0}$ its solution. One can check that $f_s \longrightarrow 1$ as $s \to \infty$. In particular,

$$W_2(f\mu, f_0\mu) = 0; \qquad \int f_0 \log f_0\, d\mu = \int f \log f\, d\mu, \tag{42}$$

$$W_2(f\mu, f_s\mu) \xrightarrow[s\to\infty]{} W_2(f\mu,\mu); \qquad \int f_s \log f_s\, d\mu \xrightarrow[s\to\infty]{} 0. \tag{43}$$

It is shown in [33] that

$$\frac{d}{ds} W_2(f_s\mu, \mu) \leq \sqrt{\int \frac{|\nabla f|^2}{f}\, d\mu}. \tag{44}$$

The function $t \mapsto W_2(f_s\mu,\mu)$ may not be differentiable, but then one can interpretate the derivative in (44) in distribution sense. On the other hand, it is easily checked that for each $s > 0$,

$$\frac{d}{ds}\sqrt{\int f_s \log f_s\, d\mu} = -\frac{\displaystyle\int \frac{|\nabla f_s|^2}{f_s}\, d\mu}{2\sqrt{\displaystyle\int f_s \log f_s\, d\mu}},$$

and by logarithmic Sobolev inequality we deduce that

$$\frac{d}{ds}\sqrt{\int f_s \log f_s\, d\mu} \leq -\sqrt{\frac{\lambda}{2}}\sqrt{\int \frac{|\nabla f_s|^2}{f_s}\, d\mu}. \tag{45}$$

Finally, the combination of (44), (45) (both of them integrated from $s = 0$ to $+\infty$), (42) and (43) implies inequality (41).

Below are two important consequences of inequality (41), in the form of **concentration inequalities**. Both of them are valid in a very general framework (Polish spaces...).

Consequence 4.1 (Marton; Talagrand)**.** *For all set A with positive measure, and for all $t \geq 0$,*

$$\mu[A_t] \geq 1 - \exp\left\{-\frac{\lambda}{2}\left(t - \sqrt{\frac{2}{\lambda}\log\frac{1}{\mu[A]}}\right)^2\right\}. \tag{46}$$

Consequence 4.2 (Bobkov and Götze)**.** *For all Lipschitz function* φ*, satisfying* $\|\varphi\|_{\mathrm{Lip}} \leq 1$ *and* $\int \varphi \, d\mu = 0$*, one has*

$$\forall t > 0 \qquad \int e^{t\varphi} \, d\mu \leq e^{\frac{t^2}{2\lambda}}. \tag{47}$$

Or equivalently, for all Lipschitz function φ*,*

$$\int e^{t\varphi} \, d\mu \leq \exp\left(t \int \varphi \, d\mu \right) \exp\left(\frac{t^2}{2\lambda} \|\varphi\|_{\mathrm{Lip}}^2 \right). \tag{48}$$

As a further consequence, for all Lipschitz function φ*,*

$$\mu\left[\varphi - \int \varphi \, d\mu \geq \varepsilon \right] \leq \exp\left(-\frac{\lambda \varepsilon^2}{2\|\varphi\|_{\mathrm{Lip}}^2} \right), \tag{49}$$

and of course, by symmetry,

$$\mu\left[\left| \varphi - \int \varphi \, d\mu \right| \geq \varepsilon \right] \leq 2\exp\left(-\frac{\lambda \varepsilon^2}{2\|\varphi\|_{\mathrm{Lip}}^2} \right), \tag{50}$$

Proof of consequence 4.1. For any measurable set B with $\mu[B] > 0$, it is clear that $1_B \mu/\mu[B]$ is a probability measure. So whenever A and B are measurable subsets with positive measure, then

$$W_2\left(\frac{1_A}{\mu[A]}\mu, \frac{1_B}{\mu[B]}\mu \right) \leq W_2\left(\frac{1_A}{\mu[A]}\mu, \mu \right) + W_2\left(\mu, \frac{1_B}{\mu[B]}\mu \right).$$

Applying inequality (41) to both terms in the right-hand side, one finds

$$W_2\left(\frac{1_A}{\mu[A]}\mu, \frac{1_B}{\mu[B]}\mu \right) \leq \sqrt{\frac{2}{\lambda} \log \frac{1}{\mu[A]}} + \sqrt{\frac{2}{\lambda} \log \frac{1}{\mu[B]}}.$$

On the other hand, one can give the following interpretation of $W_2((1_A/\mu[A])\mu, (1_B/\mu[B]))^2$. It is the minimal work necessary to transport all of the mass of the probability measure $(1_A/\mu[A])\mu$, onto the probability measure $(1_B/\mu[B])\mu$, taking into account that moving one unit of mass by a distance d costs d^2. Therefore, this is bounded by $d(A,B)^2$, with $d(A,B) = \inf\{d(x,y);\ x \in A,\ y \in B\}$. So in the end,

$$d(A,B) \leq \sqrt{\frac{2}{\lambda} \log \frac{1}{\mu[A]}} + \sqrt{\frac{2}{\lambda} \log \frac{1}{\mu[B]}}. \tag{51}$$

If we now choose $B = A_t^c$, we have $d(A,B) \geq t$ and $\mu[B] = 1 - \mu[A_t^c]$, and (51) transforms into (46) after appropriate rewriting. □

Proof of consequence 4.2. Let μ satisfy (41), and let f be a nonnegative L^1 function, such that $f\mu$ is a probability measure. First of all, by Hölder inequality,

$$W_1(f\mu, \mu) \leq W_2(f\mu, \mu).$$

A general duality principle for mass transportation, called the **Kantorovich duality**, and which in the case of W_1 is often called the **Kantorovich-Rubinstein theorem**, states that

$$W_1(f\mu, \mu) = \sup_{\|\varphi\|_{\mathrm{Lip}} \le 1} \left(\int \varphi \, f d\mu - \int \varphi \, d\mu \right).$$

Thus, if φ satisfies the assumptions of the theorem ($\int \varphi \, d\mu = 0$, $\|\varphi\|_{\mathrm{Lip}} \le 1$), one has

$$\begin{aligned} \int \varphi f \, d\mu \le W_2(f\mu, \mu) &\le \sqrt{\frac{2}{\lambda} \int f \log f \, d\mu} \\ &\le \frac{t}{2\lambda} + \frac{1}{t} \int f \log f \, d\mu. \end{aligned}$$

In particular,

$$\sup_f \left(\int \varphi f \, d\mu - \frac{1}{t} \int f \log f \, d\mu \right) \le \frac{t}{2\lambda}, \tag{52}$$

where the supremum is taken over all $L^1(d\mu)$ functions f such that $f \, d\mu$ is a probability measure. From standard Legendre duality, the supremum in (52) turns out to be

$$\frac{1}{t} \log \left(\int e^{t\varphi} \, d\mu \right).$$

This concludes the proof of (47). Then the equivalence between (47) and (48) is immediate. Finally, (49) follows by the choice $t = \lambda\varepsilon / \|\varphi\|_{\mathrm{Lip}}^2$. □

Combining consequence 4.2 with (40), one obtains

Corollary 5. *Let μ satisfy a logarithmic Sobolev inequality with constant λ on $(\mathbb{R}^d)^N$. Then, whenever φ is a Lipschitz function with Lipschitz constant bounded by 1,*

$$\mu \left[\left| \frac{1}{N} \sum_{i=1}^N \varphi(x^i) - E_\mu \left(\frac{1}{N} \sum_{i=1}^N \varphi(x^i) \right) \right| \ge \varepsilon \right] \le 2 \exp\left(-\frac{\lambda N \varepsilon^2}{2} \right). \tag{53}$$

In the end of this lecture, I describe how Malrieu [24, 25] has used the concentration inequalities above in the study of the mean-field model (37). The following result is taken from [24].

Theorem 6. *Consider the system (37), with initial positions distributed according to the chaotic distribution $f_0^{\otimes N}$. Assume that V is smooth and uniformly convex with constant $\beta > 0$, and that W is smooth, convex, even and has polynomial growth at infinity. Let $(f_t)_{t \ge 0}$ be the solution to the McKean-Vlasov equation (38). Further assume that $\int f_0(x)|x|^p \, dx < +\infty$ for some p large enough, that $\int f_0 \log f_0 < +\infty$, and that f_0 satisfies a logarithmic Sobolev inequality. Then, there exist various constants $C > 0$ such that* *(i) propagation of chaos holds uniformly in time: there exists a symmetric system of "particles" $(Y_t^1, \dots, Y_t^N)$, identically distributed and independent, with law $f_t(x)\,dx$, such that*

$$\sup_{t\geq 0} E|X_t^i - Y_t^i|^2 \leq \frac{C}{N};$$

(ii) For any Lipschitz function φ with Lipschitz constant bounded by 1,

$$P\left(\left| \frac{1}{N} \sum_{i=1}^N \varphi(X_t^i) - \int_{\mathbb{R}^d} \varphi(x) f_t(x)\,dx \right| \geq \sqrt{\frac{C}{N}} + r \right) \leq 2 \exp\left(-\frac{Nr^2}{2D_t} \right),$$

where D_t is uniformly bounded as $t \to \infty$;

(iii) let $\mu_t^{(1,N)}$ be the law of X_t^1. Then, it converges exponentially fast to equilibrium as $t \to \infty$, uniformly in N:

$$W_2(\mu_t^{(1,N)}, \mu_\infty) \leq Ce^{-\lambda t}.$$

Remark: These results take important advantage of the uniform convexity of V. They do not hold in the case where, for instance, $V = 0$, even if W is uniformly convex. The problem comes from the fact that the quantity $\sum X_t^i$ is invariant under the action of the drift associated with the potential W, and undergoes only diffusion. As shown by Malrieu [25], one can fully remedy this problem by projecting the whole system onto the hyperplane $(\sum X^i = 0)$, which amounts to getting rid of this "approximate conservation law".

Sketch of proof of Theorem 6. The first part is conceptually easy, but requires some skill in the computations. The idea to introduce the system Y_t^i, as a system of N independent particles, each of which has law f_t, was popularized by Sznitman [35]. It is very likely that X_t and Y_t are quite close, because, by Itô's formula, Y_t solves

$$dY_t^i = dB_t^i - \nabla V(Y_t^i) - \nabla(f_t * W)(Y_t^i).$$

With the help of this formula, a long calculation shows that, if

$$\alpha(t) = E[(X_t^1 - Y_t^1)^2],$$

then

$$\alpha'(t) \leq -2\beta\alpha(t) + \frac{C}{\sqrt{N}}\alpha(t)^{1/2}.$$

Moreover, one is allowed to choose $Y_0^i = X_0^i$, so that $\alpha(0) = 0$. Then, Gronwall's lemma implies a bound like

$$\alpha(t)^{1/2} \leq \frac{C}{\sqrt{N}}[1 - e^{-\beta t}]$$

(here, C stands for various positive constants).

In many problems of mean-field limit this first part is very important because it shows that the understanding of the simpler system (Y_t) is enough to get a good knowledge of the original system (X_t). Surprisingly, the sequel of the proof will take extremely little advantage of this fact.

If φ is a smooth Lipschitz test-function, from this first part we deduce immediately that

$$|E\varphi(X_t^1) - E\varphi(Y_t^1)| \le \sqrt{\frac{C}{N}}.$$

Since the particles are exchangeable, this also shows that

$$\left| E\left(\frac{1}{N}\sum_{i=1}^N \varphi(X_t^i)\right) - E\left(\frac{1}{N}\sum_{i=1}^N \varphi(Y_t^i)\right)\right| \le \sqrt{\frac{C}{N}}.$$

As a consequence,

$$P\left(\left|\frac{1}{N}\sum \varphi(X_t^i) - \int f_t(x)\varphi(x)\,dx\right| \ge \frac{C}{\sqrt{N}} + r\right)$$
$$= P\left(\left|\frac{1}{N}\sum \varphi(X_t^i) - E\left(\frac{1}{N}\sum_{i=1}^N \varphi(Y_t^i)\right)\right| \ge \frac{C}{\sqrt{N}} + r\right)$$
$$\le P\left(\left|\frac{1}{N}\sum \varphi(X_t^i) - E\left(\frac{1}{N}\sum_{i=1}^N \varphi(X_t^i)\right)\right| \ge r\right)$$

Now, to prove (ii) it is sufficient to show that the law μ of X_t satisfies a concentration inequality like (53); hence it is sufficient to prove that it satisfies a logarithmic Sobolev inequality. As a consequence of Itô's formula, this probability measure solves the diffusion equation

$$\frac{\partial \mu}{\partial t} = \frac{1}{2}\Delta\mu + \nabla\cdot(\nabla \mathbf{V}\mu), \tag{54}$$

where Δ and ∇ are differential operators acting on $L^1(\mathbb{R}^{dN})$, and

$$\mathbf{V}(x^1,\dots x^N) = \sum_{i=1}^N V(x^i) + \frac{1}{2N}\sum_{i,j} W(x^i - x^j).$$

One easily checks that $\mathbf{V}$ is uniformly convex with constant β. A nontrivial result by Bakry [2] (heavily relying on the linearity of the microscopic equation (54)) implies that μ_t satisfies a logarithmic inequality with constant $\lambda_t = [e^{-2\beta t}\lambda_0^{-1} + \beta^{-1}(1-e^{-2\beta t})]^{-1}$, if μ_0 itself satisfies a logarithmic Sobolev inequality with constant λ_0. This ends the proof of (ii).

As for the proof of (iii), it is based on ideas quite similar to those of lecture II about the speed of approach to equilibrium. One of the key facts here is that both the function $\int f \log f \, d\mu$ and the Wasserstein distance behave very well as the dimension N increases. This allows one to provide an argument of trend to equilibrium which works "uniformly well with respect to the dimension." For more information, the reader can consult [24, 25]. □

These developments suggest many interesting problems. For instance, can the convexity assumptions on the interaction potential be relaxed ? Can phase transition phenomena occur if the potential is plainly non-convex ?

In this last lecture I shall present Otto's construction, which provides a re-interpretation of the optimal mass transportation problem for quadratic cost, in a point of view which is very appealing both from the point of view of fluid mechanics, and from the geometrical point of view. This re-interpretation is formal, and at first sight it is not clear what one gains by it; but it leads to an important harvest of results, some of which are explained in the end. If one keeps Otto's interpretation in mind, then some of the strange formulas encountered before in lectures II and IV, such as (11), (17) or (41), appear more natural. As a word of caution, let me insist that this interpretation is definitely not a universal explanation to all the interesting phenomena about mass transportation.

To begin with, let me present the **Benamou-Brenier formulation** of optimal transportation with quadratic cost.

Let ρ_0 and ρ_1 be two probability density on $\mathbb{R}^n$, say with compact support. Think of them as describing two different states of a certain bunch of particles. Assume now that at time $t = 0$, this bunch of particles is in the state described by ρ_0, and you have the possibility to make them move around by imposing any time-dependent velocity field you wish in $\mathbb{R}^n$. This has an energetical cost, which coincides with the total kinetic energy of the particles. Your goal is to have, at time $t = 1$, the particles in state ρ_1, and furthermore you are looking for the solution which will require the least amount of work. Stated informally, your goal is to minimize the **action**

$$A = \int_0^1 \left(\sum_x |\dot{x}(t)|^2 \right) dt,$$

with x varying in the set of all particles, and x at time 0 is distributed according to ρ_0, at time 1 according to ρ_1.

This can be rephrased in a purely Eulerian way: let ρ_t be the density of the bunch of particles at time t, and v_t the associated velocity field (defined only on the support of ρ_t). Then the preceding problem becomes

$$\inf_{\rho,v} \left\{ \int_0^1 \int \rho_t(x)|v_t(x)|^2 \, dx \, dt; \qquad \frac{\partial \rho_t}{\partial t} + \nabla \cdot (\rho_t v_t) = 0 \right\}, \tag{55}$$

where the infimum is taken over all time-dependent probability densities $(\rho_t)_{0\le t\le 1}$ which agree with ρ_0 and ρ_1 at respective times $t = 0$ and $t = 1$, and overall time-dependent velocity fields $(v_t)_{0\le t\le 1}$ which convect ρ_t, as expressed by the continuity equation in the right-hand side of (55).

This minimization problem is not very precisely defined as far as functional spaces are concerned. If one wants to give a mathematically rigorous definition, it is best to change unknowns (ρ, v) for $(\rho, P) = (\rho, \rho v)$ and to rewrite (55) as

$$\inf_{\rho,P} \int_0^1 \int \frac{|P_t(x)|^2}{\rho_t(x)} \, dx \, dt; \qquad \frac{\partial \rho_t}{\partial t} + \nabla \cdot P_t = 0. \tag{56}$$

Then this minimization problem makes sense under minimal regularity assumptions on (ρ, P) (say continuous with respect to time, ρ taking values in probability space and P in vector-valued distributions). This comes from the joint convexity of $|P|^2/\rho$ as a function of P and ρ, and from the linearity of constraints.

Theorem 7 (Benamou-Brenier)**.** *Assume that ρ_0, ρ_1 are probability densities on $\mathbb{R}^n$ with compact support, and let I be the value of the infimum in* (56)*. Then*

$$I = \min_{T\#\rho_0=\rho_1} \int |x - T(x)|^2 \, \rho_0(x) \, dx = W_2(\rho_0, \rho_1)^2. \tag{57}$$

Note that in this statement, we have again identified measures with their densities. The second inequality in (57) is part of the Brenier-McCann theorem alluded to in lecture II, so the new point is only that this infimum also coincides with I.

Sketch of proof. Here is a formal proof of (57), in which regularity issues are discarded. Let us first check that

$$I \geq W_2(\rho_0, \rho_1)^2.$$

Let (ρ, v) be admissible in (56), and introduce the characteristics associated with v. These are the integral curves $(T_t x)_{0 \leq t \leq 1}$ defined by

$$\frac{d}{dt} T_t x = v_t(T_t x).$$

From the theory of linear transport equations it is known that $\rho_t = T_t\#\rho_0$; in particular $\rho_1 = T_1\#\rho_0$. Then,

$$\begin{aligned} \int \rho_t(x)|v_t(x)|^2 \, dx &= \int \rho_0(x)|v_t(T_t x)|^2 \, dx \\ &= \int \rho_0(x) \left| \frac{d}{dt} T_t x \right|^2 \, dx. \end{aligned}$$

Therefore, integrating in t and using Jensen's inequality, we find

$$\begin{aligned} \int_0^1 \int \rho_t(x)|v_t(x)|^2 \, dx \, dt &\geq \int \rho_0(x) \left| \int_0^1 \left(\frac{d}{dt} T_t x \right) \, dt \right|^2 \, dx \\ &= \int \rho_0(x)|T_1(x) - T_0(x)|^2 \, dx \\ &= \int \rho_0(x)|T_1(x) - x|^2 \, dx. \end{aligned}$$

Since $T_1\#\rho_0 = \rho_1$, this quantity is bounded below by the right-hand side of (57).

In a second step, let us check the reverse inequality

$$I \leq W_2(\rho_0, \rho_1)^2.$$

Let $T = \nabla\varphi$ be given by the Brenier-McCann theorem: it satisfies $T\#\rho_0 = \rho_1$ and

$$W_2(\rho_0, \rho_1)^2 = \int \rho_0(x)|x - T(x)|^2 \, dx.$$

From this T it is easy to construct an admissible (ρ, v): just set

$$T_t(x) = (1-t)x + tT(x), \qquad v_t = \left(\frac{d}{dt} T_t \right) \circ T_t^{-1},$$

$$\rho_t = T_t \# \rho_0.$$

Then one easily checks that for all $t \in [0,1]$,

$$\int \rho_t(x)|v_t(x)|^2\, dx = \int \rho_0(x)|T(x) - x|^2\, dx = W_2(\rho_0, \rho_1)^2.$$

In particular,

$$\int_0^1 \left(\int \rho_t |v_t|^2 \right) dt = W_2(\rho_0, \rho_1)^2,$$

which implies the conclusion. □

It was noticed by Otto that theorem 7 leads to the re-interpretation of the Wasserstein distance as the geodesic distance for some "Riemannian" structure on the set of probability densities:

$$W_2(\rho_0, \rho_1)^2 = \inf \left\{ \int_0^1 \left\| \frac{\partial \rho}{\partial t} \right\|_{\rho(t)}^2 dt; \qquad \rho(0) = \rho_0,\ \rho(1) = \rho_1, \right\}$$

where the "metric" on the "tangent space" at ρ would be defined by

$$\left\| \frac{\partial \rho}{\partial s} \right\|_\rho^2 = \inf_v \left\{ \int \rho |v|^2; \quad \frac{\partial \rho}{\partial s} + \nabla \cdot (\rho v) = 0 \right\}.$$

Here is the interpretation: let $\partial\rho/\partial s$ be an infinitesimal variation of the probability density ρ. If one thinks of ρ as the density of a bunch of particles, this infinitesimal variation can be associated to many possible velocity fields for the particles: namely, all those fields v's which satisfy the continuity equation $\nabla \cdot (\rho v) = -\partial\rho/\partial s$. To each of these velocity fields is associated a kinetic energy, which is the total kinetic energy of the particles. Now, to this infinitesimal variation associate the minimum of the kinetic energy, for all possible choices of the velocity field.

Assume that ρ is smooth and positive, then an optimal v should be characterized by its being a *gradient*. Indeed, let v be optimal, and let w be any divergence-free vector field, then $v_\varepsilon = v + \varepsilon(w/\rho)$ is another admissible vector field, so the associated kinetic energy is no less than the kinetic energy associated to v. Thus,

$$\int \rho|v|^2 \le \int \rho|v_\varepsilon|^2 = \int \rho|v|^2 + 2\varepsilon \int v \cdot w + \varepsilon^2 \int \frac{|w|^2}{\rho}.$$

Simplifying and letting $\varepsilon \to 0$, we see that $\int v \cdot w = 0$. Since w was arbitrary in the space of divergence-free vector fields, v should be a gradient.

Note that, at least formally, the elliptic PDE

$$-\nabla \cdot (\rho \nabla \varphi) = \frac{\partial \rho}{\partial s} \tag{58}$$

should be uniquely solvable (up to an additive constant) for φ.

Remark: If we wished to discuss the whole thing in a bounded subset of $\mathbb{R}^n$, rather than in the whole of $\mathbb{R}^n$, then we should choose a convex set containing the supports of ρ_0 and ρ_1, and impose Neumann boundary condition for (58).

To sum up, in Otto's formalism the Wasserstein distance is the geodesic distance associated to the following metric: for each probability density ρ, whenever $\partial\rho/\partial s$ and $\partial\rho/\partial t$ are any two infinitesimal variations of ρ, define

$$\left\langle \frac{\partial\rho}{\partial s}, \frac{\partial\rho}{\partial t} \right\rangle_\rho = \int \rho \nabla\varphi \cdot \nabla\psi, \tag{59}$$

where φ, ψ are the solutions of

$$-\nabla\cdot(\rho\,\nabla\varphi) = \frac{\partial\rho}{\partial s}, \qquad -\nabla\cdot(\rho\,\nabla\psi) = \frac{\partial\rho}{\partial t}.$$

The construction above cannot apparently be put on solid ground, in the sense that it seems hopeless to properly define a Riemannian structure along the lines suggested above. However, this formalism has two main advantages: first, in some situations it backs up intuition in a powerful way; second, it suggests very efficient rules of formal computations. Indeed, as soon as one has defined a Riemannian structure, then associated to it are rules of **differential calculus**: in particular, one can define a **gradient** and a **Hessian** operators. Recall the general definition of a gradient operator: it is defined by the identity

$$\left\langle \mathrm{grad}F(\rho), \frac{\partial\rho}{\partial s} \right\rangle_\rho = DF(\rho)\cdot\frac{\partial\rho}{\partial s},$$

where F is any ("smooth") function on the manifold into consideration, and DF the differential of F. With the above definitions, it can be checked that the gradient of a function F on the set of probability densities is defined by

$$\mathrm{grad}F(\rho) = -\nabla\cdot\left(\rho\,\nabla\frac{\delta F}{\delta\rho}\right), \tag{60}$$

where $\delta F/\delta\rho$ denotes, as in lecture II, the gradient of the functional F with respect to the L^2 Euclidean structure.

As for Hessian operators, it is not very interesting to give a general formula, but one can often reconstruct them from the formula

$$\left\langle \mathrm{Hess}F(\rho)\cdot\frac{\partial\rho}{\partial s}, \frac{\partial\rho}{\partial s} \right\rangle = \left.\frac{d^2}{ds^2}\right|_{s=0} F(\rho_s),$$

with $(\rho_s)_{s\geq 0}$ standing for the geodesic curve issued from ρ with velocity $\partial\rho/\partial s$. This geodesic can be described as follows: let φ be associated to $\partial\rho/\partial s$ via (58), then

$$\rho_s = [(1-s)\mathrm{id} + s\varphi]\#\rho.$$

As a consequence of (60), equations with the particular structure (5) can be seen as **gradient flows** of the energy F with respect to Otto's differential structure. This is very interesting if one looks at the long-time or the short-time behavior of equations such as (5). Indeed, there are strong connections between the behavior of a gradient flow,

$$\frac{\partial f}{\partial t} = -\mathrm{grad}F(\rho)$$

and the convexity properties of F. In particular, here we understand that the natural notion of convexity, for the asymptotic properties of (5), is the convexity in the sense of Otto's differential calculus, i.e. convexity along geodesics, which coincides with the **displacement convexity** discussed in section II. We also understand that formula (9) is just the natural Taylor formula in this context, to quantify how convex the energy is.

Here is an example of what can be deduced from these considerations.

Formal theorem to be checked on each case (Otto and Villani)**.** *Let F be a λ-uniformly displacement convex functional. Then it admits a unique minimizer ρ_∞, and one has the functional inequality*

$$\int \rho \left|\nabla \frac{\delta F}{\delta \rho}\right|^2 \geq 2\lambda[F(\rho) - F(\rho_\infty)]. \tag{61}$$

As an example, let

$$F(\rho) = \frac{1}{m-1}\int \rho(x)^m \, dx + \int \rho(x)\frac{|x|^2}{2}\, dx$$

be defined on $L^1(\mathbb{R}^n)$. From a theorem of McCann [27], the first half of the right-hand side is displacement convex if $m \geq 1 - 1/n$. On the other hand, the second half is 1-uniformly convex. Moreover, if $m \geq 1 - 1/n$, then the functional F is bounded below on the set of probability densities, with the exception of the case $m = 1/2$, $n = 1/2$. In all the other cases it follows from the formal theorem above that

$$\int \rho \left|\frac{m}{m-1}[\nabla \rho^{m-1}](x) + x\right|^2 \geq 2[F(\rho) - F(\rho_\infty)]. \tag{62}$$

These inequalities are the key to the study of the asymptotic behavior to the solutions of the porous medium equations,

$$\frac{\partial \rho}{\partial t} = \Delta \rho^m,$$

and were found independently by several authors [13, 18, 32].

The following astonishing remark by Dolbeault and Del Pino establishes an unexpected connection between this field and the theory of optimal Sobolev-type inequalities. Expand the square in the left-hand side of (62) and notice that the terms in $\int \rho(x)|x|^2\, dx$ cancel out on both sides of the inequality. Then one is left with

$$\left(\frac{m}{m-1}\right)^2 \int \rho|\nabla \rho^{m-1}|^2 + \frac{2m}{m-1}\int \rho(x)\,(\nabla \rho^{m-1})(x)\cdot x \geq \frac{2}{m-1}\int \rho^m + C_\infty,$$

for some constant $C_\infty = -2F(\rho_\infty)$. It is not difficult to identify ρ_∞:

$$\rho_\infty(x) = \left(\sigma^2 + \frac{1-m}{2m}|x|^2\right)_+^{\frac{1}{m-1}}$$

(in physics, this profile is known as the Barenblatt profile), and therefore one can compute the constant C_∞ explicitly. Next, by chain-rule and integration by parts,

$$\frac{2m}{m-1}\int \rho(x)\,(\nabla \rho^{m-1})(x)\cdot x = -2\int \rho^m.$$

Moreover, by chain-rule,

$$\left(\frac{m}{m-1}\right)^2 \int \rho |\nabla \rho^{m-1}|^2 = \left(\frac{2m}{2m-1}\right)^2 \int |\nabla(\rho^{m-\frac{1}{2}})|^2.$$

On the whole, we see that

$$\left(\frac{2m}{2m-1}\right)^2 \int |\nabla(\rho^{m-\frac{1}{2}})|^2 \geq \frac{2m}{m-1}\int \rho^m + C_\infty,$$

for any probability density ρ. Or equivalently, for any nonnegative function ρ,

$$\left(\frac{2m}{2m-1}\right)^2 \left(\int \rho\right)^{1-m} \int |\nabla(\rho^{m-\frac{1}{2}})|^2 \geq \frac{2m}{m-1}\left(\int \rho^m\right) + C_\infty \left(\int \rho\right)^m .$$

If one now sets $u = \rho^{m-1/2}$, this transforms into

$$\left(\frac{2m}{2m-1}\right)^2 \|\nabla u\|_{L^2} \|u\|_{L^{2/(2m-1)}}^{1-m} \geq \frac{2m}{m-1}\|u\|^m_{L^{2m/(2m-1)}} + C_\infty \|u\|^m_{L^{2/(2m-1)}}.$$

By tedious but easy homogeneity argument, this transforms into a Gagliardo-Nirenberg inequality

$$\|u\|_{L^p} \leq C \|\nabla u\|^\theta_{L^2} \|u\|^{1-\theta}_{L^q}, \tag{63}$$

for some exponents p, q defined as follows: either $2 < p \leq 2n/(n-2)$ and $q = (p+2)/2$, or $0 < p < 2$ and $q = 2(p-1)$. Once p and q are given, the value of θ is completely determined by scaling homogeneity.

It turns out that the constants C obtained in this process are optimal ! and that the minimizers in these inequalities are obtained by a simple rescaling of the Barenblatt profile. Further note that the case $p = 2n/(n-2)$ corresponds to $\theta = 1$, and then (63) is nothing but the usual Sobolev embedding with optimal constants (for $n \geq 3$); it also corresponds to the critical exponent $m = 1 - 1/n$ found by McCann [27].

To conclude this lecture, let me mention an important application of Otto's interpretation: it suggests a natural approximation of gradient flow equations like (5) by a discrete time-step minimization procedure, see in particular [21]. This procedure can be extended to more general cost functions (in the definition of the mass transportation problem) than the quadratic one, see Otto [30], or Agueh's PhD thesis. For instance, the heat equation, but also the porous medium equations, the p-Laplace equations (or rather p-heat equations), or even a combination of both, can be obtained in this way. It can also be used for some Hamiltonian equations, see Cullen and Maroofi [17]. Otto has also worked on more degenerate examples [31] arising from fluid mechanics problems. Finally, Carlen and Gangbo have studied interesting variants of this time-step procedure in a kinetic context [10, 11].

References

[1] Ané, C., Blachère, S., Chafaï, D., Fougères, P., Gentil, I., Malrieu, F., Roberto, C., and Scheffer, G. *Sur les inégalités de Sobolev logarithmiques*, vol. 10 of *Panoramas et Synthèses*. Société Mathématique de France, 2000.

[2] Bakry, D. On Sobolev and logarithmic Sobolev inequalities for Markov semigroups. In *New trends in stochastic analysis (Charingworth, 1994), Taniguchi symposium* (1997), World Sci. Publishing.

[3] Bakry, D., and Emery, M. Diffusions hypercontractives. In *Sém. Proba. XIX*, no. 1123 in Lecture Notes in Math. Springer, 1985, pp. 177–206.

[4] Barthe, F. On a reverse form of the Brascamp-Lieb inequality. *Invent. Math. 134*, 2 (1998), 335–361.

[5] Benedetto, D., Caglioti, E., Carrillo, J. A., and Pulvirenti, M. A non-Maxwellian steady distribution for one-dimensional granular media. *J. Statist. Phys. 91*, 5-6 (1998), 979–990.

[6] Benedetto, D., Caglioti, E., and Pulvirenti, M. A kinetic equation for granular media. *RAIRO Modél. Math. Anal. Numér. 31*, 5 (1997), 615–641. Erratum in *M2AN Math. Model. Numer. Anal. 33*, 2 (1999), 439–441.

[7] Bobkov, S., Gentil, I., and Ledoux, M. Hypercontractivity of Hamilton-Jacobi equations. *J. Math. Pures Appl. 80*, 7 (2001), 669–696.

[8] Bobylev, A. V., Carrillo, J., and Gamba, I. On some properties of kinetic and hydrodynamic equations for inelastic interactions. Preprint.

[9] Brenier, Y. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math. 44*, 4 (1991), 375–417.

[10] Carlen, E., and Gangbo, W. Constrained steepest descent in the 2-Wasserstein metric. Preprint, 2001.

[11] Carlen, E., and Gangbo, W. On the solution of a model Boltzmann equation via steepest descent in the 2-Wasserstein metric. Preprint, 2001.

[12] Carrillo, J., McCann, R., and Villani, C. Kinetic equilibration rates for granular media. Preprint, 2001.

[13] Carrillo, J. A., and Toscani, G. Asymptotic L^1-decay of solutions of the porous medium equation to self-similarity. *Indiana Univ. Math. J. 49*, 1 (2000), 113–142.

[14] Cercignani, C. Recent developments in the mechanics of granular materials. In *Fisica matematica e ingeneria delle strutture : rapporti e compatibilità*, G. Ferrarese, Ed. Pitagora Ed., Bologna, 1995, pp. 119–132.

[15] Cordero-Erausquin, D. Some applications of mass transport to Gaussian type inequalities. To appear in *Arch. Rational Mech. Anal.*

[16] Cordero-Erausquin, D., Gangbo, W., and Houdré, C. Inequalities for generalized entropy and optimal transportation. Preprint, 2001.

[17] Cullen, M., and Maroofi, H. The fully compressible semi-geostrophic system from meteorology. Preprint, 2001.

[18] Del Pino, M., and Dolbeault, J. Generalized Sobolev inequalities and asymptotic behaviour in fast diffusion and porous medium problems. Preprint Univ. Paris IX-Dauphine, CEREMADE, n. 9905, 1999.

[19] Del Pino, M., and Dolbeault, J. Best constants for Gagliardo-Nirenberg inequalities and applications to nonlinear diffusions. To appear in *J. Funct. Anal.*, 2001.

[20] Gross, L. Logarithmic Sobolev inequalities and contractivity properties of semigroups. In *Dirichlet forms (Varenna, 1992)*. Springer, Berlin, 1993, pp. 54–88. Lecture Notes in Math., 1563.

[21] Jordan, R., Kinderlehrer, D., and Otto, F. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal. 29*, 1 (1998), 1–17.

[22] Ledoux, M. Inégalités isopérimétriques en analyse et probabilités. *Astérisque*, 216 (1993), Exp. No. 773, 5, 343–375. Séminaire Bourbaki, Vol. 1992/93.

[23] Ledoux, M. Concentration of measure and logarithmic Sobolev inequalities. In *Séminaire de Probabilités, XXXIII*. Springer, Berlin, 1999, pp. 120–216.
[24] Malrieu, F. Logarithmic Sobolev inequalities for some nonlinear PDE's. To appear in *Stochastic Process. Appl.*
[25] Malrieu, F. Convergence to equilibrium for granular media equations and their Euler schemes. Preprint, 2001.
[26] McCann, R. J. Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J. 80*, 2 (1995), 309–323.
[27] McCann, R. J. A convexity principle for interacting gases. *Adv. Math. 128*, 1 (1997), 153–179.
[28] McNamara, S., and Young, W. Kinetics of a one-dimensional granular medium. *Phys. Fluids A 5*, 1 (1993), 34–45.
[29] Murata, H., and Tanaka, H. An inequality for certain functional of multidimensional probability distributions. *Hiroshima Math. J. 4* (1974), 75–81.
[30] Otto, F. Double degenerate diffusion equations as steepest descent. Preprint Univ. Bonn, 1996.
[31] Otto, F. Lubrication approximation with prescribed nonzero contact angle. *Comm. Partial Differential Equations 23*, 11-12 (1998), 2077–2164.
[32] Otto, F. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations 26*, 1-2 (2001), 101–174.
[33] 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.
[34] Otto, F., and Villani, C. Comment on: "Hypercontractivity of Hamilton-Jacobi equations" [J. Math. Pures Appl. (9) **80** (2001), no. 7, 669–696; 1846020] by S. G. Bobkov, I. Gentil and M. Ledoux. *J. Math. Pures Appl. (9) 80*, 7 (2001), 697–700.
[35] Sznitman, A.-S. Topics in propagation of chaos. In *École d'Été de Probabilités de Saint-Flour XIX—1989*. Springer, Berlin, 1991, pp. 165–251.
[36] Talagrand, M. Transportation cost for Gaussian and other product measures. *Geom. Funct. Anal. 6*, 3 (1996), 587–600.
[37] 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* (1973), 47–52.
[38] Tanaka, H. Probabilistic treatment of the Boltzmann equation of Maxwellian molecules. *Z. Wahrsch. Verw. Gebiete 46*, 1 (1978/79), 67–105.
[39] Villani, C. A survey of mathematical topics in kinetic theory. To appear in *Handbook of fluid mechanics*, S. Friedlander and D. Serre, Eds.
[40] Villani, C. Mass transportation, transportation inequalities and dissipative equations. Notes for a summer school in Ponta Delgada, September 2000, 2001.
[41] Villani, C. Topics in optimal transportation. Forthcoming book by the AMS.

UMPA, Ecole Normale Supérieure de Lyon, F-69364 Lyon Cedex 07, FRANCE. e-mail: cvillani@umpa.ens-lyon.fr