Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates

(To appear in Revista Matematica Iberoamericana)

José A. Carrillo[*], Robert J. McCann[†], Cedric Villani[‡§]

March 2, 2002

Abstract

The long-time asymptotics of certain nonlinear, nonlocal, diffusive equations with a gradient flow structure are analyzed. In particular, a result of Benedetto, Caglioti, Carrillo and Pulvirenti [4] guaranteeing eventual relaxation to equilibrium velocities in a spatially homogeneous model of granular flow is extended and quantified by computing explicit relaxation rates. Our arguments rely on establishing generalizations of logarithmic Sobolev inequalities and mass transportation inequalities, via either the Bakry-Emery method or the abstract approach of Otto and Villani [28].

Mathematics Subject Classification: 35B40, 35K55, 35K65, 35Q72.
Keywords: rates of convergence, generalized log-Sobolev inequalities, Wasserstein distance, inelastic collision models.

Contents

[*]Departamento de Matemática Aplicada, Universidad de Granada, 18071 Granada, SPAIN.
[†]Department of Mathematics, University of Toronto, Toronto, Ontario, M5S 3G3, CANADA.
[‡]UMPA, Ecole Normale Supérieure de Lyon, F-69364 Lyon Cedex 07, FRANCE.
[§]e-mails: carrillo@ugr.es, mccann@math.toronto.edu, cvillani@umpa.ens-lyon.fr

1 Introduction

This paper is devoted to the asymptotic behavior of solutions of the equation

$$\frac{\partial \rho}{\partial t} = \nabla \cdot \left[\rho \nabla \left(U'(\rho) + V + W * \rho\right)\right], \tag{1.1}$$

where the unknown $\rho(t, \cdot)$ is a time-dependent probability measure on $\mathbb{R}^d$ ($d \geq 1$), $U : \mathbb{R}^+ \to \mathbb{R}$ is a density of internal energy, $V : \mathbb{R}^d \to \mathbb{R}$ is a confinement potential and $W : \mathbb{R}^d \to \mathbb{R}$ is an interaction potential. The symbol ∇ denotes the gradient operator and will always be applied to functions, while $\nabla \cdot$ stands for the divergence operator, and will always be applied to vector fields (or vector valued measures). In the sequel, we identify both the probability measure $\rho(t, \cdot) = \rho_t$ with its density $d\rho_t/dx$ with respect to Lebesgue, and thus, we use the notation $d\rho_t = d\rho(t, x) = \rho(t, x)\, dx$. We shall make precise convexity assumptions about U, V, W later on; for the moment we just mention that it is convenient to require W to be symmetric ($\forall z \in \mathbb{R}^d$, $W(-z) = W(z)$), and U to satisfy the following dilation condition, introduced in McCann [25]:

$$\lambda \longmapsto \lambda^d U(\lambda^{-d}) \qquad \text{is convex nonincreasing on } \mathbb{R}^+. \tag{1.2}$$

The most important case of application is $U(s) = s \log s$, which identifies the internal energy with Boltzmann's entropy, and yields a linear diffusion term, $\Delta\rho$, in the right-hand side of (1.1).

Equations such as (1.1) appear in various contexts; our interest for them arose from their use in the modelling of granular flows: see the works of Benedetto, Caglioti, Carrillo, Pulvirenti, Toscani [3, 4, 31] and the references there for physical background and mathematical study (a short mathematical review is provided in [34, chapter 5]). Let us just recall the most important facts.

To equation (1.1) is associated an entropy, or *free energy*:

$$F(\rho) = \int_{\mathbb{R}^d} U(\rho)\, dx + \int_{\mathbb{R}^d} V(x)\, d\rho(x) + \frac{1}{2} \int_{\mathbb{R}^d \times \mathbb{R}^d} W(x - y)\, d\rho(x)\, d\rho(y). \tag{1.3}$$

This functional can be split into the sum of an internal energy $\mathcal{U}$, a potential energy $\mathcal{V}$ and an interaction energy $\mathcal{W}$, corresponding respectively to the three terms on the right-hand side of (1.3). A simple computation shows that, at least for classical solutions, the time-derivative of $F(\rho)$ along solutions of (1.1) is the negative of

$$D(\rho) \equiv \int_{\mathbb{R}^d} |\xi|^2 \, d\rho, \tag{1.4}$$

where

$$\xi \equiv \nabla \left[U'(\rho) + V + W * \rho \right]. \tag{1.5}$$

The functional D will henceforth be referred to as the *entropy dissipation functional*. Since D is obviously nonnegative, the free energy F acts as a *Lyapunov functional* for equation (1.1).

In many cases of interest, the competition between $\mathcal{U}$, $\mathcal{V}$ and $\mathcal{W}$ determines a unique minimizer ρ_∞ for F, as shown in [25]. In this paper, our conditions on U, V, W will ensure that this is indeed true — except in certain situations where the minimizer will only be unique up to translation. A natural question is of course to determine whether solutions to (1.1) do converge to this minimizer as $t \to +\infty$, and how fast. To formulate this problem more precisely, one first needs to decide how to measure the distance between ρ and ρ_∞. In this paper this will usually be achieved by the relative free energy, improperly called *relative entropy*:

$$F(\rho|\rho_\infty) = F(\rho) - F(\rho_\infty). \tag{1.6}$$

Thus we intend to prove that $F(\rho(t, \cdot)|\rho_\infty)$ converges to 0 as $t \to +\infty$, and estimate the speed of convergence.

Let us mention here one of our main results, and relate it to previous work. A few years ago, Benedetto, Caglioti, Carrillo and Pulvirenti [4] studied equation (1.1) in the case (arising in the modelling of granular material) when $U(s) = \sigma s \log s$, $V(x) = \lambda |x|^2/2$, $W(z) = |z|^3$, $d = 1$ $(\lambda, \sigma > 0)$. Via the study of the free energy F, they proved convergence to equilibrium in large time, without obtaining any rate; here we shall prove exponential convergence at an explicit rate. Moreover, our result holds for any dimension of space, for interaction potentials which are perturbations of $|z|^3$ (while the method in [4] needs $W(z) = |z|^3$ heavily and $d = 1$ in order that F be a convex functional), and we shall also prove exponential convergence when $\lambda = 0$.

Our proofs are based on the so-called *entropy dissipation method*, which consists in *bounding below the entropy dissipation functional* (1.4) *in terms of the relative entropy* (1.6). At first sight this is a quite technical task in view of the complexity of functionals (1.4) and (1.6). Furthermore, the value of $F(\rho_\infty)$ is not explicitly known, since the Euler-Lagrange equation for the minimizer of (1.3) seems to be unsolvable — thus our strategy may seem doomed from the very beginning. But the particular structure of equation (1.1) will allow the use of powerful methods taking their roots in the theory of logarithmic Sobolev inequalities.

At this point we may recall one of the most fundamental results in this theory, due to Bakry and Emery [2]. Consider the case when $U(s) = s \log s$, $W = 0$, and assume that V is uniformly convex, in the sense that there exists $\lambda > 0$ such that

$$D^2 V \geq \lambda I, \tag{1.7}$$

where I is the identity matrix on $\mathbb{R}^d$, and the inequality holds in the sense of symmetric matrices. Then

$$D(\rho) \geq 2\lambda\, F(\rho|\rho_\infty) \tag{1.8}$$

Assuming without loss of generality that $\int e^{-V} = 1$, this can be rewritten as

$$\int_{\mathbb{R}^d} |\nabla (\log \rho + V)|^2 \, d\rho \geq 2\lambda \int_{\mathbb{R}^d} (\log \rho + V)\, d\rho. \tag{1.9}$$

This is one of the many forms taken by *logarithmic Sobolev inequalities*. The task we undertake here is to generalize the functional inequality (1.8) to handle the nonlocal nonlinearity introduced by an interaction potential W. Up to now, the most noticeable generalization of (1.8) had been the replacement of the Boltzmann entropy $U(s) = s \log s$ by other functionals associated with nonlinear but still local diffusion, as in works of Carrillo, Jüngel, Markowich, Toscani, Unterreiter, Dolbeault, del Pino and Otto [12, 9, 18, 27].

There are at least two general methods to prove inequalities such as (1.9) (and only two, so far as we know, which are robust enough to be used in our context); we shall work out both of them. The first one, inherited from the seminal work of Bakry and Emery, goes via the study of *the second derivative* of the relative entropy functional. Indeed, in many cases of interest, a direct comparison of the entropy with its dissipation is a very difficult task, but *a comparison of the entropy dissipation with the dissipation of entropy dissipation is much easier*. The surprising success of this method [1, 11, 12, 9] (see [23] for a tentative user-friendly review on these techniques), which may seem hard to believe at first, can be explained at a heuristical level by the conceptual work of Otto [27], who showed that the relation of the equation (1.1) to the free energy (1.3) has the structure of a gradient flow. Since it is well-known that the asymptotic behavior of the trajectories of a gradient flow are closely linked to the convexity properties of the corresponding functional, this suggests that differentiating twice is a natural thing to do. In fact, the relevant notion of convexity, in this context, is the *displacement convexity* introduced by McCann [25], whose definition will be recalled below.

In Section 3 of Otto and Villani [28] it is proven in an abstract framework that, at least from the formal point of view, *uniform* displacement convexity implies an inequality of the same type as (1.8). The illumination provided by this point of view is suggested briefly by the following example. Assume a convex function $f : \mathbb{R} \longrightarrow [0, \infty)$ attains its minimum value $f(w) = 0$ at $w = 0$. From uniform convexity, $f''(w) \geq \lambda$, it is easy to deduce the inequalities

$$f(w) - \frac{1}{2\lambda} (\operatorname{grad} f)^2(w) \leq 0, \tag{1.10}$$

$$f(w) - |w| |\operatorname{grad} f(w)| + \frac{\lambda}{2}|w|^2 \leq 0, \tag{1.11}$$

since both expressions are maximized at $w = 0$, where they vanish, and

$$f(w) \geq \frac{\lambda}{2}|w|^2. \tag{1.12}$$

Even in higher dimensions, convexity of a function f is an assumption about its behaviour along line segments. Thus it is not difficult to understand that inequalities (1.10–1.11) extend to any function $f : M \longrightarrow [0, \infty]$ uniformly convex with respect to arclength along the minimizing geodesics of a Riemannian manifold M; only $|w| = \operatorname{dist}(w, w_\infty)$ must be replaced by the geodesic distance from the point w to the point $w_\infty \in M$ minimizing $f(w_\infty) = 0$. Following Otto in identifying $w \leftrightarrow \rho$, $f(\,\cdot\,) \leftrightarrow F(\,\cdot\,|\rho_\infty)$, and $(\operatorname{grad} f)^2(w) \leftrightarrow D(\rho)$ we recognize (1.10) as the log Sobolev inequality (1.8) in disguise!

To discuss the analog for (1.11) to which we now turn, identify $|w| \leftrightarrow W_2(\rho, \rho_\infty)$ with the Wasserstein distance between two probability measures μ and ν on $\mathbb{R}^d$:

$$W_2(\mu, \nu) = \inf\Big\{ \sqrt{E|X - Y|^2};\ \operatorname{law}(X) = \mu,\ \operatorname{law}(Y) = \nu \Big\}, \tag{1.13}$$

the infimum being taken with respect to all couples of random variables X and Y with respective law μ and ν. (We hope there will be no confusion between the potential W and the distance W_2). This identification is natural both because $\sqrt{D(\rho)}$ measures the rate of change of $F(\rho|\rho_\infty)$ with respect to Wasserstein distance along the geodesic joining ρ to ρ_∞ [27], and also because $F(\rho|\rho_\infty)$ is a convex function of arclength along such geodesics according to McCann's displacement convexity inequalities [25][26, Remark 3.3] — the prerequisites for (1.10–1.11).

The second known approach towards inequalities such as (1.9) is what we shall refer to as the "HWI method", because it was first worked out in the context of inequality (1.9) via the HWI inequalities introduced by Otto and Villani [27, 28]. These are interpolation inequalities involving the entropy functional H, the Wasserstein distance W_2 and the relative Fisher information I. In particular, under assumption (1.7),

$$\int_{\mathbb{R}^d} (\log \rho + V)\ d\rho \leq W_2(\rho, e^{-V}) \sqrt{\int_{\mathbb{R}^d} |\nabla (\log \rho + V)|^2\ d\rho} - \frac{\lambda}{2} W_2(\rho, e^{-V})^2. \tag{1.14}$$

This is but a particular case of the more general inequality

$$F(\rho|\rho_\infty) \leq W_2(\rho, \rho_\infty)\sqrt{D(\rho)} - \frac{\lambda}{2} W_2(\rho, \rho_\infty)^2, \tag{1.15}$$

which we recognize as (1.11) in disguise.

Of course, inequality (1.8) follows immediately by relaxing the bound (1.15) via maximization with respect to $W_2(\rho, \rho_\infty) \geq 0$. Also we may identify inequality (1.12) as

$$W_2(\rho, \rho_\infty) \leq \sqrt{\frac{2}{\lambda} F(\rho|\rho_\infty)} \tag{1.16}$$

which corresponds to a generalized Talagrand inequality [28]. This point of view introduces the Wasserstein distance in a problem in which it is not explicitly present; however, this is not so surprising since the Wasserstein distance is intimately linked with the notion of displacement convexity, which we like to think about as the key concept underlying the Bakry-Emery theorem.

As another connection, our companion paper [10] uses the Wasserstein distance to measure the distance between ρ and ρ_∞ in the problem of asymptotic behavior for equation (1.1). It gives a different approach to this problem, by exploiting directly quantitative versions of displacement convexity developed in a length space setting. However, here we shall show that in many cases, trend to equilibrium in relative entropy implies trend to equilibrium in Wasserstein distance, by displaying some *transportation inequalities*, or functional inequalities comparing the Wasserstein distance with the relative entropy. The proof of these inequalities is patterned on Otto and Villani [28].

When applicable, the HWI proof is certainly preferable to the Bakry-Emery argument, for it is more direct and does not presuppose a priori knowledge that $\rho_t \longrightarrow \rho_\infty$. However, the Bakry-Emery argument has the advantage not to explicitly use mass transportation, with the resulting difficulties about smoothness issues sometimes associated with it. Even if formally complicated, the manipulations leading to the computation of the dissipation of entropy dissipation are very easy to justify if solutions to (1.1) are smooth. And in any case, the Bakry-Emery method is presently familiar to a much wider mathematical audience than the HWI method, which justifies our implementing both schemes here.

As a further remark on the subject, we mention the recent simplified proof of (1.14) in [13, 14]. It will be interesting to see if this argument can be extended to provide a simplified proof for the inequalities of the type (1.15) which will show up in the present study.

Finally, let us explain a little bit about the plan of this paper. Since the main object of this paper is by no means the Cauchy problem associated with (1.1), we shall not search for optimal conditions or refined existence/uniqueness theorems, but simply state at the beginning of Section 2 some conditions which ensure the existence of weak solutions; the discussion of these topics is postponed to the appendix. The rest of Section 2 is devoted to the presentation of our main results about trend to equilibrium and related functional inequalities. In Section 3, we shall present some crucial preliminary computations which are at the basis of our proofs. Most of the material there is not new, except for the few computations dealing with the interaction energy $\mathcal{W}$. Then, in Section 4, we present the proofs of our main results. Finally, in Section 5, we show how to apply these techniques in order to prove exponential convergence in a more traditional sense, namely L^1 norm. Readers may wish to consult Theorem 5.1 to see a concrete example of a purely PDE problem which can be solved quite satisfactorily by the use of our mass transportation techniques.

2 Main results

In the sequel we shall make the following **technical assumptions**, which ensure the existence of "well-behaved" solutions to (1.1):

- $U(s) = 0$ (no diffusion), or $U(s) = \sigma s \log s$ for some $\sigma > 0$ (linear diffusion), or U is a strictly convex function on $\mathbb{R}^+$, such that $U(0) = 0$, of class at least C^4 on $(0, +\infty)$, with a right-derivative at 0, superlinear at infinity in the sense that

$$\frac{U(s)}{s} \xrightarrow[s \to +\infty]{} +\infty. \tag{2.1}$$

  Moreover, $s \mapsto sU''(s)$ should be nondecreasing for $s > 0$ small enough.

  In particular, $U(s) = s^m$ for $m > 1$ is convenient. The case $U(s) = s \log s$ could be seen as a limit case of this family as $m \to 1$. It should also be possible to include singular cases such as $U(s) = -s^m$ for $m \in (1 - d^{-1}, 1)$, but we shall not discuss this here.

- The non-negative potentials V and W lie in $W^{2,\infty}_{\mathrm{loc}}(\mathbb{R}^d)$, grow at most polynomially at infinity, and satisfy

$$\forall x \in \mathbb{R}^d \qquad x \cdot \nabla V(x) \geq -C, \qquad |\nabla V(x)| \leq C(1 + |x|)^\alpha,$$

  and either

$$\forall z \in \mathbb{R}^d \qquad z \cdot \nabla W(z) \geq K(1 + |z|)^{1+\beta} - C, \qquad |\nabla W(z)| \leq C(1 + |z|)^\beta$$

  or

$$\forall z \in \mathbb{R}^d \qquad z \cdot \nabla W(z) \geq -C, \qquad |\nabla W(z)| \leq C(1 + |z|)$$

  for some constants α, β, C, K.

- We also impose that the interaction potential W is symmetric: $W(-z) = W(z)$. In the case of the nonlinear diffusion, we impose for simplicity that $W(z)$ be a function of $|z|$ and that V be strictly convex.

Before we state a theorem, let us make the meaning of solutions more precise. In the case of linear diffusion, where equation (1.1) will take the form

$$\partial_t \rho_t = \sigma \Delta \rho_t + \nabla \cdot (\rho_t \nabla V) + \nabla \cdot (\rho_t \nabla (W * \rho_t)),$$

with $\sigma \geq 0$, we define a solution as a mapping $t \mapsto \rho_t \in C(\mathbb{R}^+; \mathcal{D}'(\mathbb{R}^d))$, with values in the set of probability measures, such that $\nabla W * \rho_t \in L^\infty_{\mathrm{loc}}(\mathbb{R}^+ \times \mathbb{R}^d)$, and such that for all $T > 0$ and smooth, compactly supported test-functions $\varphi \in \mathcal{D}(\mathbb{R}^d)$,

$$\int \varphi \, d\rho_T - \int \varphi \, d\rho_0 = \sigma \int_0^T dt \int \Delta\varphi \, d\rho_t - \int_0^T dt \int \nabla\varphi \cdot \nabla(V + W * \rho_t) \, d\rho_t. \tag{2.2}$$

On the other hand, in the case of nonlinear diffusion, when the equation can be rewritten as

$$\partial_t \rho_t = \Delta P(\rho_t) + \nabla \cdot (\rho_t \nabla V) + \nabla \cdot (\rho_t \nabla (W * \rho_t)),$$

then we require in addition ρ_t to be absolutely continuous with respect to Lebesgue measure for a.a. $t \geq 0$, $P(\rho_t) \in L^1_{\mathrm{loc}}(\mathbb{R}^+ \times \mathbb{R}^d)$, and replace formula (2.2) by

$$\int \varphi \, d\rho_T - \int \varphi \, d\rho_0 = \int_0^T dt \int \Delta\varphi \, P(\rho_t) \, dx - \int_0^T dt \int \nabla\varphi \cdot \nabla(V + W * \rho_t) \, d\rho_t, \quad (2.3)$$

where the pressure function P is given by (3.3).

Proposition 2.0. *Let the technical assumptions above be satisfied. Then, there exists $s_0 > 0$, depending on V, W, such that whenever ρ_0 is an initial probability measure such that $\int |x|^{s_0} \, d\rho_0(x) < +\infty$, $F(\rho_0) < +\infty$, and $d\rho_0/dx \in L^\infty(\mathbb{R}^d)$ in the case of nonlinear diffusion, then there exists a solution $(\rho_t)_{t \geq 0}$ of* (1.1)*, such that $t \mapsto \rho_t$ is continuous in distribution sense, and $t \mapsto F(\rho_t)$ is nonincreasing on $\mathbb{R}^+$ hence a.e. differentiable, and*

$$\frac{d}{dt} F(\rho_t) \leq -D(\rho_t) \qquad \text{for a.a. } t > 0, \quad (2.4)$$

where the entropy dissipation functional $D(\rho)$ is defined by (1.4)*.*

We admit here this proposition, and sketch its proof in the appendix. We insist that this result is *not* optimal, and that, depending on the cases, more general situations are allowed, and many of the above requirements can be relaxed (for instance, the finiteness of all moments....) Moreover, under very mild additional assumptions these solutions are smooth, and one can prove *equality* in (2.4). We do not treat all these extensions in order to limit the size of the present paper, and because this is not at all our main subject of interest here. We also point out that, in most of the cases to be examined, by the method of Otto and Villani [29] one can prove that $F(\rho_t)$ immediately becomes finite even if it is infinite at initial time.

In the sequel, we shall implicitly assume that the assumptions of Proposition 2.0 are satisfied, and we now concentrate on the problem of trend to equilibrium for these solutions. We insist that all our results below do not use the above technical assumptions, but are valid as soon as strong solutions to (1.1) can be constructed.

Our main results can be summarized by some heuristic rules, whose precise meaning is contained in the theorems below. We denote by I the $d \times d$ identity matrix.

Rule # 1: *A uniformly convex confinement potential implies an exponential decay to equilibrium. Moreover, if the convexity of the confinement potential is strong enough, it can overcome a lack of convexity of the interaction potential.*

This rule is illustrated by:

Theorem 2.1. *Assume that U satisfies the dilation condition* (1.2)*, and V, W are convex. Assume moreover that $V \geq 0$ is uniformly convex, in the sense that*

$$D^2 V \geq \lambda I$$

for some $\lambda > 0$. Then,

(i) There exists a unique minimizer ρ_∞ of F, which also turns out to be the unique stationary state for equation (1.1)*;*

(ii) Whenever ρ is a probability density satisfying $F(\rho) < +\infty$, then

$$F(\rho|\rho_\infty) \leq W_2(\rho, \rho_\infty)\sqrt{D(\rho)} - \frac{\lambda}{2} W_2(\rho, \rho_\infty)^2, \tag{2.5}$$

$$D(\rho) \geq 2\lambda F(\rho|\rho_\infty), \tag{2.6}$$

$$W_2(\rho, \rho_\infty) \leq \sqrt{\frac{2F(\rho|\rho_\infty)}{\lambda}}; \tag{2.7}$$

(iii) Solutions to (1.1) *which are provided by Proposition* 2.0 *satisfy the decay estimate*

$$F(\rho_t|\rho_\infty) \leq e^{-2\lambda t} F(\rho_0|\rho_\infty). \tag{2.8}$$

(iv) Moreover, if W is not convex but the negative part $(D^2W)^-$ of its Hessian satisfies

$$\|(D^2W)^-\|_{L^\infty} < \lambda/2$$

then points (ii)-(iii) above still hold true, with λ replaced by $\lambda - 2\|(D^2W)^-\|_{L^\infty}$.

Remark: Of course, the standard logarithmic Sobolev inequality (1.9) is just a particular case of this theorem (choose $U(s) = s \log s$, $W = 0$). Also, generalized Log-Sobolev inequalities [9, 18] are obtained as particular cases (choose $U(s) = s^m$, $W = 0$).

Rule # 2: *When the center of mass is fixed, then a uniformly convex interaction potential implies an exponential decay to equilibrium. In this situation, a strongly convex interaction can also overcome a lack of convexity of the confinement.*

Thus, in terms of trend to equilibrium, the uniform convexity of the interaction potential is just as good as the uniform convexity of the confinement potential, except that the interaction potential needs a fixed center of mass. This is due to the fact that the interaction energy,

$$\mathcal{W}(\rho) = \frac{1}{2} \int_{\mathbb{R}^{2d}} W(x-y)\, d\rho(x)\, d\rho(y),$$

is invariant by translation, and therefore fails to be strictly displacement convex under shift of probability measures. Rule # 2 is the content of the following

Theorem 2.2. *Assume that U satisfies the dilation condition* (1.2)*, and that V and W are convex. Assume moreover that the center of mass of ρ_t,*

$$\theta_t = \int_{\mathbb{R}^d} x \, d\rho_t(x) \tag{2.9}$$

is invariant by the evolution along the equation (1.1)*, and that W is uniformly convex, in the sense that*

$$D^2 W \geq \lambda I$$

for some $\lambda > 0$. Then,

(i) For any $\theta \in \mathbb{R}^d$ there exists a unique minimizer ρ_∞ of F among probability measures ρ such that $\int x \, d\rho_\infty(x) = \theta$. In the same class of probability measures, ρ_∞ also turns out to be the unique stationary state for equation (1.1)*;*

(ii) Whenever ρ is a probability density satisfying $F(\rho) < +\infty$, $\int x \, d\rho(x) = \theta$, then the HWI inequalities (2.5–2.7) *are satisfied;*

(iii) Solutions to (1.1) *provided by Proposition* 2.0 *satisfy the decay estimate* (2.8)*.*

(iv) Moreover, if V is not convex but $\|(D^2V)^-\|_{L^\infty} < \lambda$, then (2.5–2.8) *above still hold true, with λ replaced by $\lambda - \|(D^2V)^-\|_{L^\infty}$.*

Remarks:

1. Simple sufficient conditions for the center of mass to be fixed are either that $V = 0$, or that V and W are radially symmetric and that we restrict equation (1.1) to radially symmetric initial data (in which case (2.6) holds for radially symmetric probability densities).

2. We shall see later that in presence of diffusion, similar results hold true without the restriction on the center of mass.

Rule # 3: *When the interaction potential is only degenerately convex and when there is no diffusion, then the decay to equilibrium is in general only algebraic.*

This situation is exemplified in the situation when the interaction potential is of the form $W(z) = |z|^{\gamma+2}$ for some exponent $\gamma > 0$, either in the whole space or locally, close to 0. If there is no diffusion and V is strictly convex, then the minimizer is just a Dirac mass located at the minimum of V (if there is no confinement, then the position of the Dirac mass is determined by the center of mass of the initial probability distribution). Then it is possible to construct explicit solutions for which the trend to equilibrium is only algebraic with rate $t^{-1/\gamma}$ for convergence in the Wasserstein sense and with rate $t^{-(\gamma+2)/\gamma}$ for convergence in the relative entropy sense.

Theorem 2.3. *Assume that $U = 0$, and that V, W are convex. Assume moreover that the center of mass θ_t of ρ_t is fixed by the evolution along the equation* (1.1)*, and that W is degenerately (strictly) convex at the origin, in the sense that*

$$D^2W \geq A \min(|z|^\alpha, 1), \qquad A > 0, \quad 0 < \alpha < 2.$$

Further assume that

$$|D^2W(z)| \leq B(1 + |z|^\beta), \qquad \beta > 0.$$

Then,

(i) There exists a unique minimizer ρ_∞ for F, in the class of probability measures ρ with center of mass θ, and this minimizer turns out to be the unique steady state for (1.1) *in the same class of probability measures;*

(ii) For all probability density ρ with finite moments of sufficiently high order, there exists a constant $K > 0$ and exponent $\kappa > (1 - \alpha/2)^{-1} > 1$, depending only on A, B, α, β, the dimension d, and moments of ρ of order large enough, such that

$$D(\rho) \geq K F(\rho|\rho_\infty)^\kappa; \tag{2.10}$$

If $\kappa < 2$, then also

$$W_2(\rho, \rho_\infty) \leq \frac{1}{(1 - \kappa/2)K^{1/2}} F(\rho|\rho_\infty)^{1-\kappa/2}; \tag{2.11}$$

(iii) Solutions to (1.1) *which are provided by Proposition* 2.0 *satisfy the decay estimate*

$$\begin{aligned} F(\rho_t|\rho_\infty) &\leq \frac{F(\rho_0|\rho_\infty)}{[1 + (\kappa - 1)F(\rho_0|\rho_\infty)^{\kappa-1}Kt]^{1/(\kappa-1)}} && (2.12) \\ &\leq \frac{1}{((\kappa - 1)Kt)^{1/(\kappa-1)}}. && (2.13) \end{aligned}$$

Remarks:

1. This theorem is the only one in the series in which we shall only apply the Bakry-Emery strategy, and not the "HWI strategy".

2. Since the constant K can be chosen uniformly for probability measures ρ whose moments of high enough order are uniformly bounded, and since moments of ρ_t are nonincreasing with t under our assumptions, (2.12) follows from (2.4) and (2.10) via Gronwall's inequality.

3. This result is also true in the diffusive case, for smooth solutions of (1.1), if we allow the constant K to depend on $\sup_{t\geq 0} \|D^2U'(\rho_t)\|_{L^\infty}$ and on $\sup_{t\geq 0} \int |x|^s d\rho_t(x)$ for s large enough, $(\rho_t)_{t\geq 0}$ standing for the solution of (1.1) with initial datum $\rho_0 = \rho$.

4. The exponent κ provided by our proof is in general not optimal. If ρ decays fast enough, then κ can be chosen arbitrarily close to $(1-\alpha/2)^{-1}$. A complicated variant of the proof of Theorem 2.4 below enables to treat any exponent α not necessarily smaller than 2; we do not reproduce it here.

5. We do not consider here the case when the convexity is degenerate at infinity. A reason for this is that in cases of applications known to us, a typical behavior for W is $|z|^3$, hence degenerately convex at the origin but not at infinity. Let us however mention that the influence of a degeneracy at infinity of the convexity of the confinement potential V was discussed in Section 2 of Toscani and Villani [32], in the case when $W = 0$ and $U(s) = s \log s$. The results there are in the same spirit as here: an algebraic rate of decay is derived via an inequality like (2.10). However, the main difference is that when there is some degeneracy at infinity, then it is not necessarily true that the solution has uniformly (in time) bounded moments. In other words, the confinement may be too weak to ensure a good localization of solutions; as a consequence, a careful study of the behavior of moments has to be performed. It is easy to check that such a study is also possible in our situation, under some precise assumptions on W.

Rule # 4: *In presence of (linear or superlinear) diffusion, a degenerately convex interaction potential induces an exponential trend to equilibrium.*

This rule is much more surprising than rule # 3, because the decay is exponential even though one would like to consider the degenerate interaction as the driving mechanism for equilibration. The idea is that if the interaction potential is degenerately convex, then the associated energy $\mathcal{W}$ fails to be uniformly displacement convex, but this failure only matters for probability densities which are very concentrated. The presence of diffusion associated with a *superlinear* internal energy density compels the probability density to be *spread* enough that the interaction energy behave just as if it were uniformly displacement convex. This is the content of the following

Theorem 2.4. *Assume that U satisfies the dilation condition* (1.2)*, that $U(s)/s \to +\infty$ as $s \to +\infty$, and that $V \geq 0$ and $W \geq 0$ are convex. Further assume that the center of mass of ρ_t is fixed by the evolution along the equation* (1.1)*, and that W is degenerately (strictly) convex, in the sense that*

$$D^2 W(z) \geq \psi_W(z) I, \qquad \psi_W \in C(\mathbb{R}^d), \tag{2.14}$$

for some modulus of convexity ψ_W which is allowed to degenerate near $z = 0$ only:

$$\psi_W(z) > 0 \textit{ if } z \neq 0, \qquad \psi_W(z) \textit{ is uniformly bounded below as } |z| \to \infty. \tag{2.15}$$

Then,

(i) There exists a unique minimizer ρ_∞ for F, in the class of probability measures ρ with center of mass θ, and this minimizer turns out to be the unique steady state for (1.1) *in the same class of probability measures;*

(ii) For all probability densities ρ such that $F(\rho) < +\infty$, there exists a constant $\lambda = \lambda(\rho)$, depending only on $F(\rho)$, U, V, ψ_W, d, such that the HWI inequalities (2.5–2.7) *are satisfied;*

(iii) If λ_0 is the constant associated to ρ_0 by (ii), then solutions to (1.1) *provided by Proposition* 2.0 *satisfy the decay estimate*

$$F(\rho_t|\rho_\infty) \le e^{-2\lambda_0 t} F(\rho_0|\rho_\infty). \tag{2.16}$$

Remarks:

1. Unlike in preceding theorems, we have explicitly restated the superlinear growth condition $U(s)/s \to +\infty$ with s, because here it is not merely a technical assumption; instead, it seems to play a crucial role in our proof. As mentioned before, typical examples of U satisfying the assumptions of Theorem 2.4 are $U(s) = s\log s$, $U(s) = s^m$ for $m > 1$.

2. For any value $F_0 < \infty$, the constant $\lambda(\rho)$ can be chosen uniform in the class of probability measures ρ such that $F(\rho) \le F_0$. Since $F(\rho(t,\cdot))$ is nonincreasing with t, point (ii) implies point (iii) at once, yielding a rate of convergence which only depends on $F(\rho)$ at the initial time.

3. If the reader is interested in practical computations, then we must warn him that the constants provided by the proof may be rather poor. The reason is that the free energy is usually very bad at controlling concentration, especially in the case $U(s) = s\log s$. Depending on the situation, it may be much better to use L^∞ bounds on ρ for instance, and perform the proof again. As will be clear from our argument, all that one needs to control is the concentration of the solution (and of the decay at infinity). We also mention that in [16], a spectacular improvement of scales of time for decay were obtained by using (in a rather unexpected way) the *entropy dissipation* itself for the control of concentration.

4. In the case of logarithmic Sobolev inequalities: $U(s) = s\log s$, $W = 0$, then it is well-known by the Holley-Stroock perturbation lemmas [20] that an inequality such as (2.6) holds true when V is degenerately (strictly) convex, with a universal constant (depending only on V). However, the only known proof of this result is a perturbation argument at the level of (1.9), and it is an open problem to recover it directly by a Bakry-Emery type argument. On the contrary, here we are able to prove Theorem 2.4 by a variant of the Bakry-Emery strategy. Even if in the end the constant is not universal, since it depends on $F(\rho)$ (and on the dimension d) this shows an unexpectedly nice behavior of the interaction energy.

Rule #5: *In presence of diffusion, the interaction potential is able to drive the system to equilibrium even if the center of mass is moving.*

This rule is exemplified in the situation when the system is confined by a degenerately convex potential, so that the center of mass may be moving with time (if we are not in a radially symmetric situation), but the confinement potential lacks the uniform convexity required to drive the system to equilibrium:

Theorem 2.5. *Assume that U satisfies the dilation condition* (1.2)*, that $U(s)/s \to +\infty$ as $s \to +\infty$, and that V and W are convex. Further assume that W is degenerately (strictly) convex, in the sense of Theorem 2.4, and that either $V = 0$ or*

$$D^2 V \geq \psi_V(x), \qquad \psi_V \in C(\mathbb{R}^d), \quad \psi_V(x) > 0 \quad \text{if } x \neq 0.$$

Then,

(i) There exists a unique minimizer ρ_∞ for F, which also turns out to be the unique stationary state for (1.1)*;*

(ii) For all probability densities ρ, there exists a constant $\lambda = \lambda(\rho)$, depending only on $F(\rho)$, U, ψ_W, ψ_V, d, such that the HWI inequalities (2.5–2.7) *are satisfied;*

(iii) If λ_0 is the constant which is associated to ρ_0 by (ii), then solutions of (1.1) *provided by Proposition* 2.0 *satisfy the decay estimate* (2.16)*.*

Remarks:

1. Of course, Theorem 2.5 is more general than Theorem 2.4, but we have presented it separately for the sake of clarity.

2. Again, the main idea is that the bound on the superlinear internal energy prevents concentration of ρ. And again, better constants may be obtained in practice by using L^∞ bounds for instance.

The remainder of this paper is devoted to the proofs of Theorems 2.1 to 2.5.

Important remark: By combining the estimates of the next sections with the method of Otto and Villani [29], one can prove the a priori estimate, in *all* the cases above,

$$F(\rho_t|\rho_\infty) \leq \frac{W_2(\rho_0, \rho_\infty)^2}{4t} \leq \frac{C(\rho_0)}{t}, \tag{2.17}$$

where the constant $C(\rho_0)$ depends on ρ_0 only via $\int |x|^2 \, d\rho_0(x)$. This allows one to **extend all the present results to the case where the initial datum has infinite free energy**. We shall skip this extension, only because we do not wish to discuss the Cauchy problem in this case.

3 Preliminary computations

In this section we prepare for the proofs of Theorems 2.1 to 2.5 by presenting two crucial computations. The first one is the formula for the second variation of the functional F under "displacement", which is, crudely speaking, geodesic variation in the sense of optimal transportation with respect to Wasserstein L^2 distance. The second one is the formula for the dissipation of the entropy dissipation, which is at the basis of the Bakry-Emery argument.

The computations here are formal and we shall not endeavor to justify them; only in the next section will we be concerned with rigorous justification. Nearly all of the forthcoming discussion follows Otto and Villani [28].

3.1 Second variation of entropy $F(\rho)$ under displacement

Let ρ_0, ρ_1 be two L^1 probability measures. As we know from [6] and [24], there exists a ($d\rho_0$-a.e.) unique gradient of convex function $\nabla\varphi$, such that

$$\nabla\varphi\#\rho_0 = \rho_1.$$

Here # denotes the push-forward operation, defined by the formula

$$\int f\, d(T\#\mu) = \int (f \circ T)\, d\mu.$$

Moreover, $\nabla\varphi$ is an optimal transportation for the Monge-Kantorovich problem with quadratic cost, in the sense that

$$W_2(\rho_0, \rho_1)^2 = \int_{\mathbb{R}^d} |x - \nabla\varphi(x)|^2\, d\rho_0(x).$$

The family of probability measures

$$\rho_s = [(1-s)\mathrm{Id} + s\nabla\varphi]\#\rho_0$$

plays the role of a geodesic path interpolating between ρ_0 and ρ_1. Whenever G is a functional such that, for all ρ_0, ρ_1, the map $s \mapsto G(\rho_s)$ is convex, one says that G is *displacement convex*. It is known since the work of McCann [25] that $\mathcal{U}$ is displacement convex whenever U satisfies the dilation condition (1.2), that $\mathcal{V}$ is displacement convex whenever V is convex, and that $\mathcal{W}$ is displacement convex whenever W is convex.

An easy computation shows that, if smoothness issues are disregarded, then ρ_s satisfies the following differential system representing conservation of mass and momentum:

$$\begin{cases} \dfrac{\partial\rho_s}{\partial s} + \nabla\cdot(\rho_s v_s) = 0, \\[2ex] \dfrac{\partial(\rho_s v_s)}{\partial s} + \nabla\cdot(\rho_s v_s \otimes v_s) = 0, \end{cases} \tag{3.1}$$

with $v_0(x) = \nabla\varphi(x) - x$.

From this one deduces after a long calculation that

$$\frac{d^2}{ds^2}\mathcal{U}(\rho_s) = \int_{\mathbb{R}^d} \big[P'(\rho_s)\rho_s - P(\rho_s)\big](\nabla \cdot v_s)^2 + \int_{\mathbb{R}^d} P(\rho_s)\,\mathrm{tr}(Dv_s)^2, \tag{3.2}$$

where

$$P(\rho) = \int_0^\rho \sigma U''(\sigma)\,d\sigma = \rho U'(\rho) - U(\rho) \tag{3.3}$$

is the (nonnegative) "pressure" associated to the equation (1.1) (in the case of linear diffusion, just replace $P(\rho)$ by ρ). Here Dv stands for the matrix $(\partial_i v_j)_{\leq i,j \leq d}$. Note that v_s remains a gradient vector field for all $s \in [0,1]$, so that $\mathrm{tr}(Dv_s)^2$ coincides with $\mathrm{tr}(Dv_s)^T(Dv_s)$, which is the Hilbert-Schmidt square-norm of the matrix Dv_s. The pressure function P is rather naturally associated to the evolution, since the first (i.e. nonlinear diffusion) term on the right-hand side of (1.1) can be rewritten as $\Delta[P(\rho)]$; further note the dilation condition (1.2) is equivalent to: $P(\rho)/\rho^{1-1/d}$ nondecreasing, or

$$\rho P'(\rho) \geq (1 - 1/d)P(\rho). \tag{3.4}$$

A by now standard argument [27, 9] uses the Cauchy-Schwarz inequality for the Hilbert-Schmidt norm to show that, under assumption (3.4), the right-hand side of (3.2) is nonnegative.

The second variation of $\mathcal{V}$ is easier to compute:

$$\frac{d^2}{ds^2}\mathcal{V}(\rho_s) = \int_{\mathbb{R}^d} \langle D^2 V \cdot v_s, v_s\rangle\,d\rho_s. \tag{3.5}$$

As for the interaction energy $\mathcal{W}$, it requires a little bit more work,

$$\frac{d}{ds}\mathcal{W}(\rho_s) = -\frac{1}{2}\int \nabla W(x-y)\cdot[v_s(x) - v_s(y)]\,d\rho_s(x)\,d\rho_s(y),$$

$$\frac{d^2}{ds^2}\mathcal{W}(\rho_s) = \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} \Big\langle D^2W(x-y)\cdot[v_s(x)-v_s(y)], [v_s(x)-v_s(y)]\Big\rangle\,d\rho_s(x)\,d\rho_s(y). \tag{3.6}$$

From these equations we deduce that whenever U satisfies the condition (1.2) and when V, W are both *convex*, then

$$\frac{d^2}{ds^2}F(\rho_s) \geq 0,$$

which indicates that F is displacement convex in the sense of McCann [25]. We think of all our results below as consequences of some form of *strict displacement convexity*. One can quantify displacement convexity as follows: as explained at length in [27, 28], the structure of the Monge-Kantorovich problem allows to use these second derivatives along optimal transportation in order to define a formal Hessian operator.

$$\left\langle \mathrm{Hess}_{W_2} F(\rho) \cdot \frac{d\rho}{ds}, \frac{d\rho}{ds} \right\rangle_{W_2} = \left. \frac{d^2}{ds^2} \right|_{s=0} F(\rho_s), \tag{3.7}$$

where ρ_s is the solution of (3.1), $v = v_0$ being a gradient vector field satisfying $\partial\rho/\partial s = -\nabla \cdot (\rho v)$. What we shall use in the sequel is that when U satisfies (1.2) and when everything is smooth enough, then

$$\begin{aligned} &\left\langle \mathrm{Hess}_{W_2} F(\rho) \cdot \frac{d\rho}{ds}, \frac{d\rho}{ds} \right\rangle_{W_2} \geq \\ &\int_{\mathbb{R}^d} \langle D^2 V \cdot v, v\rangle \, d\rho + \frac{1}{2} \int_{\mathbb{R}^d \times \mathbb{R}^d} \Big\langle D^2 W(x-y) \cdot [v(x) - v(y)], [v(x) - v(y)] \Big\rangle \, d\rho(x) \, d\rho(y). \end{aligned} \tag{3.8}$$

3.2 Dissipation of entropy dissipation

Let $(\rho_t)_{t\geq 0}$ be a smooth solution to (1.1). Then, as we mentioned earlier, $dF(\rho_t)/dt = -D(\rho_t)$, where D is the entropy dissipation functional,

$$D(\rho) = -\int_{\mathbb{R}^d} |\xi(x)|^2 \, d\rho(x),$$

with ξ given by (1.5). Similarly, one can define the *dissipation of entropy dissipation*, DD, by the formula

$$DD(\rho_0) = - \left. \frac{d}{dt} \right|_{t=0} D(\rho_t) = \left. \frac{d^2}{dt^2} \right|_{t=0} F(\rho_t). \tag{3.9}$$

The explicit form of this functional must be computed for later use.

Proposition 3.0. *Let ρ be a smooth probability measure, and assume that equation* (1.1) *has smooth solutions with initial datum ρ. Then, with notations* (1.5), (3.3) *still in use,*

$$\begin{aligned} DD(\rho) = \;& 2\int_{\mathbb{R}^d} [\rho P'(\rho) - P(\rho)] \, (\nabla \cdot \xi)^2 \, dx + \int_{\mathbb{R}^d} P(\rho) \, \mathrm{tr}(D\xi)^T (D\xi) \, dx \\ & +2 \int_{\mathbb{R}^d} \langle D^2 V \cdot \xi, \xi \rangle \, d\rho \\ & + \int_{\mathbb{R}^{2d}} \Big\langle D^2 W(x-y) \cdot [\xi(x) - \xi(y)], [\xi(x) - \xi(y)] \Big\rangle \, d\rho(x) \, d\rho(y). \end{aligned} \tag{3.10}$$

In all the sequel, we shall only use the following corollary.

Corollary 3.1. *Under the previous assumptions, if the dilation condition* (1.2)*, or equivalently* (3.4)*, is fulfilled, then*

$$\begin{aligned} DD(\rho) \geq \;& 2\int_{\mathbb{R}^d} \langle D^2V\cdot\xi, \xi\rangle \, d\rho \\ & + \int_{\mathbb{R}^{2d}} \Big\langle D^2W(x-y)\cdot[\xi(x)-\xi(y)], [\xi(x)-\xi(y)]\Big\rangle \, d\rho(x)\, d\rho(y). \end{aligned} \tag{3.11}$$

Remark: Of course, the internal energy U has not disappeared: it is hidden in the definition of ξ (see formula (1.5)).

The proof of proposition 3.0 is by direct calculation in the same spirit used by Arnold, Carrillo, Jüngel, Markowich, Toscani and Unterreiter in [1, 12, 9]; it only relies on differential calculus and integration by parts. Let us just explain a way to arrive at the result by the formal considerations developed in [28] and sketched in the paragraphs above. Consider an abstract gradient flow

$$\frac{d\rho}{dt} = -\mathrm{grad}\, F(\rho),$$

so that

$$-\frac{d}{dt}F(\rho) = \|\mathrm{grad}\, F(\rho)\|^2.$$

This is formula (1.4), if we take into account

$$\frac{\partial\rho}{\partial t} = \nabla\cdot(\rho\xi) \tag{3.12}$$

and define, whenever (3.12) is satisfied with ξ a gradient vector field,

$$\left\|\frac{d\rho}{dt}\right\|^2 = \int |\xi|^2\, d\rho. \tag{3.13}$$

Then, the second derivative of the free energy F is

$$\frac{d^2}{dt^2}F(\rho) = -\frac{d}{dt}\big\|\mathrm{grad}\, F(\rho)\big\|^2 = 2\Big\langle \mathrm{Hess}\, F(\rho)\cdot \mathrm{grad}\, F(\rho), \mathrm{grad}\, F(\rho)\Big\rangle.$$

This turns out to be formula (3.10): the three terms appearing in the right-hand side of (3.10) correspond respectively to $\mathrm{Hess}\,\mathcal{U}$, $\mathrm{Hess}\,\mathcal{V}$, $\mathrm{Hess}\,\mathcal{W}$.

4 The proofs

We now explain the proofs of our main results. Since the proofs of Theorems 2.1 to 2.5 bear a lot of common points, and since much of the work has in fact already been done in previous work, we shall only give the first proof in some detail, and be content with providing the new estimates for the other ones.

4.1 Uniformly convex confinement

Proof of Theorem 2.1. (i) The first step of the proof is to study the behavior of F under displacement. Let us consider the case where there is diffusion, so that minimizers, and solutions to (1.1), have to be L^1 functions. The existence of minimizers can be shown by standard techniques of weak convergence, or by the arguments of McCann [25]. From this reference we also know that $\mathcal{U}$ is displacement convex. Let ρ_0, ρ_1 be two given L^1 probability measures, and let $(\rho_s)_{0\le s\le 1}$ be the interpolation described in Section 3. From the procedure there, one finds

$$\mathcal{V}(\rho_s) = \int V((1-s)x + s\nabla\varphi(x))\, d\rho(x). \tag{4.1}$$

Our hypothesis $D^2V \ge \lambda I$ now implies that

$$\frac{d^2}{ds^2}\mathcal{V}(\rho_s) \ge \lambda \int |\nabla\varphi(x) - x|^2\, d\rho(x) = \lambda W_2(\rho_0,\rho_1)^2.$$

Similarly,

$$\mathcal{W}(\rho_s) = \frac{1}{2}\int W\big((1-s)(x-y) + s(\nabla\varphi(x) - \nabla\varphi(y))\big)\, d\rho(x)\, d\rho(y), \tag{4.2}$$

so the displacement convexity of $\mathcal{W}$ is easily checked. On the whole, F is therefore *uniformly* displacement convex, in the sense that $d^2F(\rho_s)/ds^2 \ge \lambda W_2(\rho_0,\rho_1)^2$. This immediately implies uniqueness in the minimizer (as in McCann [25]), for if ρ_0, ρ_1 were two distinct minimizers, then $\rho_{\frac{1}{2}}$ would satisfy $F(\rho_{\frac{1}{2}}) < [F(\rho_0) + F(\rho_1)]/2$.

In the case where there is no diffusion, the only difference is that the minimizers will not in general have a L^1 density. However, it is easy to modify the argument in the following way: let (X,Y) be a couple of random variables achieving the infimum in (1.13), then define ρ_s to be the law of $sX + (1-s)Y$.

(ii) The next step consists in writing an adequate Taylor-like formula for $F(\rho_s)$; here the proof follows Otto and Villani [28, section 5]. We would like ρ_1 and ρ_0 to be an arbitrary L^1 probability measures with finite entropy $F(\rho_0), F(\rho_1) < +\infty$. Instead, to simplify we shall take ρ_0 and ρ_1 to be smooth and compactly supported. This implies that the optimal map $\nabla\varphi$ transporting ρ_0 onto ρ_1 is L^∞ on the support of ρ_0. By approximation, in the end our results will hold true for arbitrary ρ_0 and ρ_1 with finite entropy.

Thanks to (4.1), one sees that

$$\left.\frac{d}{ds}\right|_{s=0} \mathcal{V}(\rho_s) = \int \langle \nabla V(x), \nabla\varphi(x) - x\rangle\, d\rho_0(x).$$

Similarly,

$$\left.\frac{d}{ds}\right|_{s=0} \mathcal{W}(\rho_s) \;=\; \frac{1}{2}\int \langle \nabla W(x-y), \nabla\varphi(x) - x - \nabla\varphi(y) + y\rangle\, d\rho_0(x) d\rho_0(y)$$

$$= \int \langle \nabla\varphi(x) - x, \int \nabla W(x-y)\, d\rho_0(y)\rangle\, d\rho_0(x).$$

The only delicate part is the $\mathcal{U}$ functional. As shown in [25], one can write

$$\mathcal{U}(\rho_s) = \int_{\mathbb{R}^d} U\left(\frac{\rho_0(x)}{\det((1-s)I + sD^2\varphi(x))}\right) \det((1-s)I + sD^2\varphi(x))\, dx,$$

with $D^2\varphi$ being understood in the sense of Aleksandrov. Thanks to our assumptions on U, one can check that the integrand of previous identity is a convex function of s, so that, combining Lebesgue's theorem with Aleksandrov's theorem (as in [28, Section 5]), one easily proves that

$$\begin{aligned}
\lim_{s\to 0} \frac{\mathcal{U}(\rho_s) - \mathcal{U}(\rho_0)}{s} &= -\int_{\mathbb{R}^d} P(\rho_0)\, \mathrm{tr}[D^2\varphi(x) - I]\, dx && (4.3)\\
&\geq \int \langle \nabla P(\rho_0), \nabla\varphi(x) - x\rangle\, dx && (4.4)\\
&= \int \langle \nabla U'(\rho_0), \nabla\varphi(x) - x\rangle\, d\rho_0(x). && (4.5)
\end{aligned}$$

Here inequality (4.4) follows from the fact that P is nonnegative, and the Aleksandrov (i.e. pointwise a.e.) Laplacian $\mathrm{tr}D^2\varphi$ of a convex function is always less than the distributional Laplacian $\Delta\varphi$. Summarizing, and using the notation $\xi_0 = \xi$ from (1.5) with ρ replaced by ρ_0,

$$\left.\frac{d}{ds}\right|_{s=0} F(\rho_s) \geq \int \langle \xi_0, \nabla\varphi(x) - x\rangle\, d\rho_0(x).$$

By Cauchy-Schwarz inequality, this expression is bounded below by

$$-\sqrt{\int |\xi_0|^2\, d\rho_0}\sqrt{\int |\nabla\varphi(x) - x|^2\, d\rho_0(x)} = -D(\rho_0)^{1/2} W_2(\rho_0, \rho_1).$$

Combining this with the bound on the second derivative, Taylor's formula yields

$$F(\rho_0) - F(\rho_1) \leq W_2(\rho_0, \rho_1)\sqrt{D(\rho_0)} - \frac{\lambda}{2} W_2(\rho_0, \rho_1)^2. \tag{4.6}$$

These formulas, proven for smooth densities, can be extended by density arguments as soon as $D(\rho_0) < +\infty$. One can now prove the uniqueness of the stationary state in the class of probability densities such that $D(\rho) < +\infty$. Assume that $D(\rho) = 0$ (or $\xi = 0$, which is the same), and set $\rho = \rho_0$ in (4.6). Since ρ_1 is arbitrary, it can be seen that ρ_0 is then a minimizer for F, concluding the proof of (i).

The assertions of (ii) also follow immediately: by setting ρ_1 to be the minimizer ρ_∞, one gets formula (2.5). Choosing W_2 to maximize the bound (4.6) for $\rho_0 = \rho_\infty$ implies (2.6). By setting ρ_0 to be the minimizer so $D(\rho_0) = 0$, one obtains formula (2.7).

(iii) The decay rate 2λ for the relative entropy follows immediately from (2.4) and (2.6) by Gronwall's inequality.

Before turning to (iv), let us now sketch the alternative proof for (2.6), which is the *Bakry-Emery strategy.* We first assume that everything is very smooth, in the sense which is explained in Appendix A (smooth enough solutions, etc.) It is clear from (3.11) and the hypothesis $D^2 V \geq \lambda I$ that

$$DD(\rho) \geq 2\lambda D(\rho). \tag{4.7}$$

Let us fix a smooth density ρ_0, and consider the solution $(\rho_t)_{t\geq 0}$ of equation (1.1) starting from ρ_0. Since $DD(\rho_t)$ gives the time derivative (3.9) of $-D(\rho_t)$, Gronwall's inequality applied to (4.7) yields exponential decay of $D(\rho_t) \to 0$ as $t \to \infty$. Integrating in time from 0 to $+\infty$ we get

$$D(\rho_0) = \int_0^{+\infty} DD(\rho_t)\, dt \geq 2\lambda \int_0^{+\infty} D(\rho_t)\, dt. \tag{4.8}$$

Using a priori bounds on ρ and the same method as [9, Theorem 11] (for smooth problems), one can also prove that

$$F(\rho_t) \xrightarrow[t\to+\infty]{} F(\rho_\infty)$$

(without any a priori knowledge about the rate, of course). Since $D(\rho_t)$ in turn is the time derivative of $F(\rho_t)$, we can now evaluate (4.8) explicitly: it reduces to

$$D(\rho_0) \geq 2\lambda[F(\rho_0) - F(\rho_\infty)],$$

which is formula (2.6).

Once Theorem 2.1 is proven in the smooth case, it can be extended to the general case by a density argument (described in Appendix A).

An alternative proof of (2.7) follows the lines of [28, Theorem 1, Proposition 1]. By using (2.6), one proves that

$$\eta(t) \ \equiv\ W_2(\rho_0, \rho_t) + \sqrt{\frac{2F(\rho_t|\rho_\infty)}{\lambda}}.$$

is nonincreasing. Then the desired inequality is just $\eta(\infty) \leq \eta(0)$.

Remark: In the case when the functional $\mathcal{V} + \mathcal{W}$ is ϕ-uniformly convex (see [10]) one can prove similarly that

$$F(\rho|\rho_\infty) \ \leq\ \sqrt{D(\rho)}\, W_2(\rho, \rho_\infty) - \phi(W_2(\rho, \rho_\infty)).$$

In the sequel, we shall systematically use either the HWI procedure or the Bakry-Emery method without mentioning it explicitly. It should be clear from the considerations above that both rely on the same key estimate, which is a lower bound on $\mathrm{Hess}_{\mathrm{W}_2}\mathrm{F}$.

(iv) Let us now turn to the proof of the last statement in Theorem 2.1; for instance, in the formalism of the dissipation of entropy dissipation. We want to estimate the error coming from the contribution of the negative part of the Hessian of W to $DD(\rho)$ in (3.10). By expanding the scalar product and applying Cauchy-Schwarz inequality on cross-terms, we find

$$\begin{aligned}
& \int \Big\langle D^2W(x-y)\cdot[\xi(x)-\xi(y)],[\xi(x)-\xi(y)]\Big\rangle\, d\rho(x)\, d\rho(y) \\
\geq\ & -\|(D^2W)^-\|_{L^\infty} \int \Big\langle \xi(x)-\xi(y),\xi(x)-\xi(y)\Big\rangle\, d\rho(x)\, d\rho(y) \\
\geq\ & -4\|(D^2W)^-\|_{L^\infty} \int_{\mathbb{R}^d} |\xi|^2\, d\rho, \\
=\ & -4\|(D^2W)^-\|_{L^\infty} D(\rho).
\end{aligned}$$

From 3.10 we see that (4.7) will still hold true, except that λ must be replaced by $\lambda - 2\|(D^2W)^-\|_{L^\infty}$. This implies (iv), thereby concluding the proof of Theorem 2.1. $\square$

4.2 Uniformly convex interaction, fixed center of mass

Next we turn to the proof of Theorem 2.2. Its proof is exactly the same as before, except that now we want to take advantage of the uniform convexity coming from the interaction potential in (3.10). Assuming that $D^2W \geq \lambda I$, we find (for any measurable vector field v)

$$\begin{aligned}
\int \Big\langle D^2W(x-y)^-\cdot[v(x)-v(y)] \quad , \quad & [v(x)-v(y)]\Big\rangle\, d\rho(x) d\rho(y) \\
\geq\ & \lambda \int |v(x)-v(y)|^2\, d\rho(x)\, d\rho(y) \\
=\ & 2\lambda \int |v(x)|^2\ d\rho(x) - 2\left(\int v\, d\rho\right)^2.
\end{aligned}$$

where the last equality is obtained by expanding the square and using symmetry.

But the fact that we restrict ourselves to probability measures whose center of mass is given entails that whenever we consider a transportation vector field, say $T(x)$, push-forwarding ρ_0 onto ρ_1, it satisfies the identity

$$\int_{\mathbb{R}^d} [T(x)-x]\, d\rho_0 = 0. \tag{4.9}$$

Indeed, this follows easily from $\int x\, d\rho_1(x) = \int T(x) d\rho_0(x)$ and $\int x\, d\rho_0(x) = \int x\, d\rho_1(x)$. Similarly, if ξ is the vector field appearing in (1.5), and since the evolution equation (1.1) can be rewritten

$$\frac{\partial \rho}{\partial t} = \nabla\cdot\big(\rho\xi\big),$$

the invariance of the center of mass implies

$$0 = \frac{d}{dt} \int_{\mathbb{R}^d} x \, d\rho(x) = d \int_{\mathbb{R}^d} \xi \, d\rho.$$

Therefore, the term $\int v \, d\rho$ will always vanish, be it in the HWI argument or in the Bakry-Emery method. Now that

$$DD(\rho) \geq 2\lambda D(\rho), \tag{4.10}$$

is established the rest of the proof of (i)-(iii) of Theorem 2.2 follows exactly as in the previous section. As for point (iv) of Theorem 2.2, it is a consequence of the estimate

$$\int \big\langle D^2V(x) \cdot v(x), v(x) \big\rangle \, d\rho(x) \geq -\|(D^2V)^-\|_{L^\infty} \int |v|^2 \, d\rho$$

which yields $DD(\rho) \geq 2(\lambda - \|(D^2V)^-\|_{L^\infty})D(\rho)$ instead of (4.10).

4.3 Degenerately convex interaction, perturbative argument

In this paragraph, we prove Theorem 2.3 by a perturbative argument. Here we only use the Bakry-Emery strategy. As already mentioned, the results below also apply in presence of some diffusion, but under the restriction that ρ, and its time-evolution along (1.1), satisfy sufficient smoothness and localization bounds. This is not a restrictive assumption in the linear diffusion case, or in the sublinear diffusion case if the initial datum is smooth enough; on the contrary, in the superlinear degenerate diffusion case it is not, in general, satisfied, since solutions may fail to be smooth at the boundary of their support.

The idea behind the proof, which is quite natural, is to isolate the degeneracy at the origin by cutting out small values of $|x-y|$ in the integral

$$\int_{\mathbb{R}^{2d}} \big\langle D^2W(x-y) \cdot [\xi(x) - \xi(y)], [\xi(x) - \xi(y)] \big\rangle \, d\rho(x) \, d\rho(y). \tag{4.11}$$

On the set $|x-y| \leq \varepsilon$, uniform convexity of W cannot be used; but on this set, $\xi(x)-\xi(y)$ will be very small, so that this part should contribute very little in (4.11).

Similar ideas have already been used by Carlen and Carvalho [7], Desvillettes and Villani [16] for the Boltzmann and the Landau equation respectively. The main difference between these works and ours is the following: for Boltzmann-type equations, the entropy functional is just the same ($H(f) = \int f \log f$) for all kind of interactions (hard, soft....), and it is only at the level of the entropy dissipation that the choice of interaction matters. And at this level, the property of *monotonicity* with respect to the cross-section can be exploited for error estimates (the importance of this fact was first pointed out by Carlen and Carvalho).

On the other hand, here we do not have any monotonicity properties at the level of the entropy dissipation (1.4)! Consider two different interaction potentials W_1 and

W_2, satisfying $W_1 \geq W_2$, or $D^2W_1 \geq D^2W_2$, either condition *does not* imply that the entropy dissipation functionals respectively associated to W_1 and W_2 can be compared. This forces our perturbation argument to be performed *at the level of the dissipation of entropy dissipation.*

In the sequel we assume that there is no confinement ($V = 0$). Although the theorem was stated under the condition $U = 0$, it was also remarked that the proof requires merely

$$D^2W(z) \geq A\min(|z|^\alpha, 1), \qquad A > 0,\ 0 < \alpha < 2; \tag{4.12}$$

$$|D^2W(z)| \leq B(1 + |z|^\beta), \qquad \beta > 0, \tag{4.13}$$

$$\sup_{t\geq 0} \int_{\mathbb{R}^d} |x|^s \, d\rho_t(x) \leq B \tag{4.14}$$

$$\sup_{t\geq 0} \left\| D^2U'(\rho_t) \right\| \leq B \tag{4.15}$$

for some $s > 0$ sufficiently large. In the last two formulas, $(\rho_t)_{t\geq 0}$ stands for the solution of (1.1) with initial datum $\rho_0 = \rho$. The last assumption is of course void if there is no diffusion ($U = 0$). As for assumption (4.14), it is also automatically satisfied if there is no diffusion and if ρ has a finite moment of order s.

Proof of Theorem 2.3. Under assumptions (4.12–4.15) we shall prove that

$$D(\rho) \geq KF(\rho|\rho_\infty)^\kappa, \tag{4.16}$$

where K, κ depend only on the constants A, B, α, β, s and d. If s is very large, then κ can be chosen arbitrarily close to $(1 - \alpha/2)^{-1}$.

The way towards (4.16) goes as follows. Let $\varepsilon < 1$ to be chosen later on, we write

$$D^2W(z) \geq A(\varepsilon^\alpha - \varepsilon^\alpha 1_{|z|\leq\varepsilon})I,$$

and accordingly,

$$\begin{aligned} DD(\rho) &\geq \int_{\mathbb{R}^{2d}} \Big\langle D^2W(x-y)\cdot[\xi(x)-\xi(y)], [\xi(x)-\xi(y)]\Big\rangle \, d\rho(x)\, d\rho(y) \\ &\geq A\varepsilon^\alpha \left[\int_{\mathbb{R}^{2d}} |\xi(x)-\xi(y)|^2 \, d\rho(x)\, d\rho(y) - \int_{|x-y|\leq\varepsilon} |\xi(x)-\xi(y)|^2 \, d\rho(x)\, d\rho(y)\right]. \end{aligned}$$

The first term can be bounded below as in the previous section, so that

$$DD(\rho) \geq A\varepsilon^\alpha \left[D(\rho) - \int_{|x-y|\leq\varepsilon} |\xi(x)-\xi(y)|^2 \, d\rho(x)\, d\rho(y)\right]. \tag{4.17}$$

We now proceed to bound the second term in the right-hand side of (4.17). Using assumptions (4.12–4.15), and the convolution structure, one easily shows that

$$\begin{cases} |\xi(z)| = |\nabla W * \rho + \nabla U'(\rho)|\,(z) \leq C(1 + |z|^{\beta+1}), \\ \\ |\xi(x) - \xi(y)| \leq C(1 + |x| + |y|)^\beta |x-y|, \end{cases} \tag{4.18}$$

where C depends on B, β and moments of ρ of order $\beta+1$. As a consequence, Chebyshev's inequality yields

$$\int_{|x-y|\leq\varepsilon} |\xi(x)-\xi(y)|^2\,d\rho(x)\,d\rho(y)$$
$$\leq \int_{|x|,|y|\leq R,\ |x-y|\leq\varepsilon} |\xi(x)-\xi(y)|^2\,d\rho(x)\,d\rho(y) + 4\int_{|x|\geq R-1} |\xi(x)|^2\,d\rho(x)$$
$$\leq C\left[\iint_{|x|,|y|\leq R,|x-y|\leq\varepsilon} (1+|x|+|y|)^{2\beta}|x-y|^2 d\rho(x)d\rho(y) + \int_{|x|\geq R-1} (1+|x|)^{2(\beta+1)}d\rho(x)\right]$$
$$\leq C\left[(1+R)^{2\beta}\int_{|x-y|\leq\varepsilon} |x-y|^2\,d\rho(x)\,d\rho(y) + \frac{4}{(1+R)^{s-2(\beta+1)}}\int_{\mathbb{R}^d}(1+|x|^s)\,d\rho(x)\right]$$
$$\leq C\left[(1+R)^{2\beta}\varepsilon^2 + \frac{1}{(1+R)^{s-2(\beta+1)}}\right],$$

where C stands for various constants depending only on B, β, s, and moments of ρ of order $s > 2(\beta+1)$. Optimizing in R, we find

$$\int_{|x-y|\leq\varepsilon} |\xi(x)-\xi(y)|^2\,d\rho(x)\,d\rho(y) \leq C\varepsilon^{2\delta}, \tag{4.19}$$

with $\delta = (s-2\beta-2)/(s-2)$, which is very close to 1 if s is large. Inserting (4.19) in (4.17), we get

$$DD(\rho) \geq A\varepsilon^\alpha[D(\rho) - C\varepsilon^{2\delta}],$$

and a convenient choice of ε yields

$$DD(\rho) \geq K D(\rho)^{1+\frac{\alpha}{2\delta}}, \tag{4.20}$$

for some constant K depending on α, s, β, A, B, d. Similarly, for all $t\geq 0$,

$$DD(\rho_t) \geq K D(\rho_t)^{1+\frac{\alpha}{2\delta}}. \tag{4.21}$$

Once (4.21) is obtained, it is very easy to build a variant of the Bakry-Emery trick: one just rewrites (4.20) as

$$DD(\rho)D(\rho)^{-\frac{\alpha}{2\delta}} \geq K D(\rho),$$

and notes that the left-hand side is the negative of the time-derivative of $D(\rho)^{1-\alpha/(2\delta)}$ along the evolution equation (1.1). If $\alpha < 2\delta$ (which is always the case if s is large enough), then one can integrate as in the Bakry-Emery procedure, and recover

$$D(\rho)^{1-\frac{\alpha}{2\delta}} \geq K F(\rho|\rho_\infty),$$

which is precisely (4.16) — the entropy-information (HI) part of the HWI inequality.

To motivate the form of the transportation inequality (2.11) — the entropy-Wasserstein (HW) part of the HWI inequality, we sketch a formal argument relying on Otto's

Riemannian calculus on the space of probability measures metrized by W_2; it can be made rigorous by mimicking the proof of [28, Theorem 1 and Proposition 1]. Fix a smooth density ρ_0 and let $(\rho_t)_{t\geq 0}$ be a solution of (1.1). Assume $\kappa \in (1,2)$, and set

$$\eta(t) = W_2(\rho_0, \rho_t) + \frac{1}{K^{1/2}(1-k/2)} F(\rho_t, \rho_\infty)^{1-k/2}.$$

Differentiating formally using $\dot{\rho}_t = -\mathrm{grad}\, F$ yields

$$\begin{aligned}
\eta'(t) &= \left\langle \mathrm{grad}\, W_{2,\rho_0} + \frac{F(\rho_t,\rho_\infty)^{-k/2}}{K^{1/2}} \mathrm{grad}\, F, \frac{d\rho_t}{dt} \right\rangle_{W_2} \\
&\leq \|\mathrm{grad}\, W_{2,\rho_0}\|_{\rho_t} \left\| \frac{d\rho_t}{dt} \right\| - \frac{F(\rho_t,\rho_\infty)^{-k/2}}{K^{1/2}} \left\| \frac{d\rho_t}{dt} \right\|^2 \\
&\leq \sqrt{D(\rho_t)} - \frac{F(\rho_t,\rho_\infty)^{-k/2}}{K^{1/2}} D(\rho_t),
\end{aligned}$$

where the last inequality follows from the identity (3.13) and the fact that the gradient of the metric distance W_{2,ρ_0} to ρ_0 cannot exceed unit magnitude. Now $\eta'(t) \leq 0$ in view of (2.10), whence $\eta(0) \geq \eta(\infty)$. But this translates into (2.11).

(iii) Finally, to deduce (2.12) from (2.4) and (2.10), integrate the inequality

$$-\frac{dF(\rho_t|\rho_\infty)}{dt} \geq KF(\rho_t|\rho_\infty)^\kappa.$$

□

As we already mentioned, assumption (4.15) might be too restrictive in certain cases of application for the diffusive case. We shall circumvent this problem in the next section by a different argument which exploits the presence of diffusion.

4.4 Degenerately convex interaction with diffusion

In this paragraph we prove Theorem 2.4. Again we use the formalism of dissipation of entropy dissipation, though the HWI method works perfectly well with the very same estimates. We assume that V is convex and that

$$D^2W(z) \geq \psi_W(z) I, \qquad \psi_W \in C^0(\mathbb{R}^d), \tag{4.22}$$

where the modulus of convexity ψ_W is allowed to degenerate at the origin:

$$\psi_W(z) > 0 \text{ if } z \neq 0, \qquad \psi_W(z) \text{ is uniformly bounded below as } |z| \to \infty. \tag{4.23}$$

Without loss of generality we assume ψ_W to be bounded.

As a consequence of Corollary 3.1,

$$DD(\rho) \geq \int_{\mathbb{R}^{2d}} \psi_W(x-y) |\xi(x) - \xi(y)|^2 \, d\rho(x)\, d\rho(y). \tag{4.24}$$

Most of the troubles caused by small values of $|x-y|$ will be eliminated by the following simple

Lemma 4.1. *Let ρ be a probability measure, and ψ_W a nonnegative continuous function. Then*

$$\int_{\mathbb{R}^{2d}} \psi_W(x-y)|\xi(x)-\xi(y)|^2\,d\rho(x)\,d\rho(y) \geq L\int |\xi(x)-\xi(y)|^2\,d\rho(x)\,d\rho(y), \tag{4.25}$$

where

$$L = \frac{1}{4}\inf_{x,z\in\mathbb{R}^d}\int_{\mathbb{R}^d}\min\bigl[\psi_W(x-y),\psi_W(z-y)\bigr]\,d\rho(y). \tag{4.26}$$

Proof. The proof relies on a trick which was used in Villani [35] with limited success. Let X be the left-hand side of (4.25). We introduce an artificial parameter $z\in\mathbb{R}^d$ and integrate the constant function X against $d\rho(z)$, to obtain

$$X = \int_{\mathbb{R}^d} X\,d\rho(z) = \int_{\mathbb{R}^{3d}} \psi_W(x-y)|\xi(x)-\xi(y)|^2\,d\rho(x)\,d\rho(y)\,d\rho(z) \tag{4.27}$$

$$\geq \int_{\mathbb{R}^{3d}} \min\bigl[\psi_W(x-y),\psi_W(z-y)\bigr]|\xi(x)-\xi(y)|^2\,d\rho(x)\,d\rho(y)\,d\rho(z). \tag{4.28}$$

Just by exchanging x and z,

$$X \geq \int_{\mathbb{R}^{3d}} \min\bigl[\psi_W(x-y),\psi_W(z-y)\bigr]|\xi(y)-\xi(z)|^2\,d\rho(x)\,d\rho(y)\,d\rho(z). \tag{4.29}$$

Since $|\xi(x)-\xi(z)|^2 \leq 2|\xi(x)-\xi(y)|^2 + 2|\xi(y)-\xi(z)|^2$, by adding (4.28) and (4.29) we deduce

$$\begin{aligned}
2X &\geq \frac{1}{2}\int_{\mathbb{R}^{3d}} \min\bigl[\psi_W(x-y),\psi_W(z-y)\bigr]|\xi(x)-\xi(z)|^2\,d\rho(x)\,d\rho(y)\,d\rho(z)\\
&= \frac{1}{2}\int_{\mathbb{R}^{2d}} |\xi(x)-\xi(z)|^2\left(\int_{\mathbb{R}^d}\min\bigl[\psi_W(x-y),\psi_W(z-y)\bigr]\,d\rho(y)\right)d\rho(x)\,d\rho(z)\\
&\geq 2L\int_{\mathbb{R}^{2d}} |\xi(x)-\xi(z)|^2\,d\rho(x)\,d\rho(z).
\end{aligned}$$

□

It is now very easy to prove Theorem 2.4. Since V and W are non-negative, the assumptions of Theorem 2.4 imply that

$$\mathcal{U}(\rho) = \int_{\mathbb{R}^d} U(\rho)\,dx < +\infty, \tag{4.30}$$

$$\frac{U(s)}{s} \xrightarrow[s\to\infty]{} +\infty. \tag{4.31}$$

The following lemma of equi-integrability is well-known:

Lemma 4.2. *Let ρ be a probability density on $\mathbb{R}^d$ and U satisfy* (4.31)*. Then, there is $\varepsilon > 0$, depending only on U, $\mathcal{U}(\rho)$, and d, such that for all balls B of radius ε,*

$$\int_B d\rho \le \frac{1}{4}. \tag{4.32}$$

Corollary 4.3. *The constant L in formula* (4.26) *admits a lower bound depending only on U, $\mathcal{U}(\rho)$, ψ_W and d.*

Proof of the corollary. Let ε be given by Lemma 4.2, and $S = \{y \in \mathbb{R}^d;\ |x - y| \le \varepsilon \text{ or } |z - y| \le \varepsilon\}$. Then $\int_{S^c} d\rho \ge 1/2$, hence

$$L \ge \frac{1}{8} \min \{\psi_W(z);\ |z| \ge \varepsilon\}.$$

□

Putting together formula (4.24), Lemma 4.1 and Corollary 4.3, we can bound $DD(\rho)$ by

$$DD(\rho) \ge 2\lambda \int_{\mathbb{R}^{2d}} |\xi(x) - \xi(y)|^2 \, d\rho(x)\, d\rho(y)$$

for some $\lambda = \lambda(\rho)$ which depends only on $F(\rho)$, *and which is uniform in the class of probability densities such that $F(\rho) \le F_0$.* The remainder of the proof of Theorem 2.4 is now as in Section 4.2.

Note that when transposing these estimates to the HWI method, one needs to know that $F(\rho_s) \le F(\rho_0)$ when ρ_s is the displacement interpolation between ρ_0 and $\rho_1 = \rho_\infty$. This is a consequence of the following elementary observation: a convex function of $s \in [0, 1]$, which achieves its minimum value at $s = 1$, has to be maximum at $s = 0$.

4.5 Treatment of moving center of mass

In this paragraph, we relax the assumptions that the center of mass be fixed by the evolution equation (1.1), and we prove Theorem 2.5. Once again, we only consider the formalism of dissipation of the entropy dissipation, but mention that the very same estimates work out for the HWI method. Using the same arguments as in the previous section we know that

$$DD(\rho) \ge 2\lambda \int_{\mathbb{R}^{2d}} |\xi(x) - \xi(y)|^2 \, d\rho(x)\, d\rho(y) + 2 \int_{\mathbb{R}^d} \langle D^2 V \cdot \xi, \xi \rangle \, d\rho \tag{4.33}$$

for some $\lambda = \lambda(\rho)$, depending on $\mathcal{U}(\rho)$. Moreover, we assume that

$$D^2 V(x) \ge \psi_V(x),$$

where ψ_V is continuous, $\psi_V(x) > 0$ if $x \ne 0$ (no assumption on the behavior of ψ_V at infinity).

Then,

$$
\begin{aligned}
DD(\rho) &\geq 2\lambda \int_{\mathbb{R}^{2d}} |\xi(x) - \xi(y)|^2 \, d\rho(x) \, d\rho(y) + 2 \int_{\mathbb{R}^d} \psi_V |\xi|^2 \, d\rho \\
&\geq 2\lambda \int_{\mathbb{R}^d} \left| \xi - \left(\int_{\mathbb{R}^d} \xi \, d\rho \right) \right|^2 \, d\rho + 2 \int_{\mathbb{R}^d} \psi_V |\xi|^2 \, d\rho \\
&= 2\lambda \left[\int |\xi|^2 \, d\rho - \left(\int \xi \, d\rho \right)^2 \right] + 2 \int_{\mathbb{R}^d} \psi_V |\xi|^2 \, d\rho \\
&= 2\lambda \left[\int_{\mathbb{R}^d} |\xi|^2 \left(1 + \frac{\psi_V}{\lambda} \right) \, d\rho - \left(\int_{\mathbb{R}^d} \xi \, d\rho \right)^2 \right].
\end{aligned}
$$

By the Cauchy-Schwarz inequality,

$$
\left(\int \xi \, d\rho \right)^2 \leq \left\{ \int |\xi|^2 \left(1 + \frac{\psi_V}{\lambda} \right) \, d\rho \right\} \left\{ \int \frac{d\rho}{1 + \frac{\psi_V}{\lambda}} \right\},
$$

so

$$
\begin{aligned}
DD(\rho) &\geq 2\lambda \left(\int |\xi|^2 \left(1 + \frac{\psi_V}{\lambda} \right) \, d\rho \right) \left(1 - \int \frac{d\rho}{1 + \frac{\psi_V}{\lambda}} \right) \\
&\geq 2\lambda \left(\int |\xi|^2 \, d\rho \right) \left(1 - \int \frac{d\rho}{1 + \frac{\psi_V}{\lambda}} \right) \\
&= 2\lambda \left(1 - \int \frac{d\rho}{1 + \frac{\psi_V}{\lambda}} \right) D(\rho) \\
&\geq 2\lambda\delta D(\rho)
\end{aligned}
$$

where $\delta > 0$ depending only on ψ_V, λ, U and $\mathcal{U}(\rho_0)$, is provided by the following lemma:

Lemma 4.4. *Let χ be a continuous function on $\mathbb{R}^d$, $\chi(z) < 1$ if $z \neq 0$. Let*

$$
\mathcal{U}(\rho) = \int_{\mathbb{R}^d} U(\rho) \, dx, \qquad \lim_{s \to +\infty} \frac{U(s)}{s} = +\infty.
$$

Then for all probability measure ρ on $\mathbb{R}^d$ there exists $\delta > 0$, depending only on χ, U, d and an upper bound for $\mathcal{U}(\rho)$, such that

$$
\int \chi(x) \, d\rho(x) \leq 1 - \delta. \tag{4.34}
$$

5 Rates of convergence in L^1

In this section, we indicate how to recover rates of convergence in a more traditional sense (total variation) as an application of the machinery developed in the present work and in Otto and Villani [28, 29]. To illustrate the method, we shall sketch the proof of the following model result, which is stated in a self-contained form:

Theorem 5.1. *Let $\rho = (\rho_t)_{t\geq 0}$ be a solution of*

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

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

$$F(\rho) = \int \rho \log \rho \, dx + \frac{1}{2}\int W(x-y)\, d\rho(x)\, d\rho(y),$$

and let ρ_∞ be the unique minimizer of F with the same center of mass as ρ. Then, for any $t_0 > 0$ there exist constants $C_0, \lambda_0 > 0$, explicitly computable and depending on ρ only via an upper bound for $\int |x|^2\, d\rho_0(x)$, such that

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

Remarks:

1. The analysis would be the same in presence of a convex external potential V.

2. This theorem improves in several respects on [4] (in which L^1 convergence was proven for the particular case of the cubic potential in dimension 1, without any rate) and on Malrieu [22] who was the first to prove exponential trend to equilibrium in L^1 sense for equations of the same type, although only in presence of a uniformly convex confining potential, and under stronger assumptions on the initial datum.

Sketch of proof. In order to limit the size of the present paper, we do not give a fully detailed proof, but show precisely how to perform all the steps and explain all the new ingredients.

From the displacement convexity of F follows, just as in Otto and Villani [29],

$$F(\rho_t) \leq \frac{W_2(\rho_0, \rho_\infty)^2}{4t} \qquad \forall t > 0.$$

Since $W_2(\rho_0,\rho_\infty)^2 \leq 2(\int |x|^2\, d\rho_0(x) + \int |x|^2\, d\rho_\infty(x))$, this gives an a priori bound on $F(\rho_t)$ for $t \geq t_0 > 0$, which only depends on $\int |x|^2\, d\rho_0(x)$. Combining this with Theorem 2.4, we recover exponential convergence of $F(\rho_t)$ towards $F(\rho_\infty)$, and also

$$W_2(\rho_t, \rho_\infty) \leq C_1 e^{-\lambda_1 t}, \qquad t \geq t_0. \tag{5.35}$$

The rest of the argument consists in transforming this information of weak convergence into strong convergence by means of an appropriate interpolation.

Now introduce the entropy dissipation functional

$$D(\rho) = \int |\nabla(\rho * W + \log \rho)|^2 \, d\rho(x).$$

Copying again the proof in [29], one also has the estimate

$$D(\rho_t) \leq \frac{W_2(\rho_0, \rho_\infty)^2}{t^2} \tag{5.36}$$

which gives a uniform bound on $D(\rho_t)$ for $t \geq t_0$.

On the other hand, an analysis of the Euler-Lagrange equation associated with the minimization of F shows that ρ_∞ has a strictly positive density satisfying

$$\log \rho_\infty + \rho_\infty * W = \mu \in \mathbb{R}. \tag{5.37}$$

In particular,

$$\nabla \log \rho_\infty = -\nabla(\rho_\infty * W),$$

and

$$\begin{aligned} \int \left| \nabla \log \left(\frac{\rho_t}{\rho_\infty} \right) \right|^2 d\rho_t(x) &= \int \left| \nabla \log \rho_t + \nabla(\rho_\infty * W) \right|^2 \, d\rho_t(x) \\ &\leq 2 \int \left[\left| \nabla \left(\log \rho_t + \rho_t * W \right) \right|^2 + \left| (\rho_t - \rho_\infty) * \nabla W \right|^2 \right] \, d\rho_t(x) \\ &\leq 2D(\rho_t) + C \left(\int (1 + |x|^{2\beta}) \, d\rho_t(x) \right)^2 . \end{aligned} \tag{5.38}$$

An easy moment estimate, in the same manner as in Appendix A.1 or in [15], leads to the differential inequality

$$\begin{aligned} \frac{d}{dt} \int (1 + |x|^2)^\beta \, d\rho_t(x) &\leq C - K \int |x|^{2\beta+\gamma} \, d\rho_t(x) \\ &\leq C - K \left[\int (1 + |x|^2)^\beta \, d\rho_t(x) \right]^{1+\frac{\gamma}{2\beta}}, \end{aligned}$$

with C and K denoting various constants depending on ρ only via an upper bound on $\int |x|^2 \, d\rho_0(x)$. This implies a uniform bound on $\int |x|^{2\beta} \, d\rho_t(x)$ as $t \to \infty$. Combining this with (5.38) and (5.36), we obtain the estimate

$$\int \left| \nabla \log \frac{\rho_t}{\rho_\infty} \right|^2 \, d\rho_t(x) \leq C, \qquad t \geq t_0 > 0. \tag{5.39}$$

Now, from (5.37) we also deduce that $\log(d\rho_\infty/dx)$ is concave (semi-concave would be sufficient for the argument). By the HWI inequality of [28],

$$\int \log \frac{\rho_t}{\rho_\infty}\, d\rho_t \leq W_2(\rho_t, \rho_\infty) \sqrt{\int \left|\nabla \log \frac{\rho_t}{\rho_\infty}\right|^2 d\rho_t},$$

and in view of (5.35) and (5.39), this shows that the left-hand side converges to 0 exponentially fast as $t \to \infty$. We conclude by the well-known inequality

$$\|\rho_t - \rho_\infty\|_{L^1}^2 \leq 2 \int \log \frac{\rho_t}{\rho_\infty}\, d\rho_t. \tag{5.40}$$

□

Remark: Due to the nonlocal nature of F, it is not obvious whether a functional inequality such as $F(\rho) - F(\rho_\infty) \geq \text{const.} \|\rho - \rho_\infty\|_{L^1}^2$ holds true; the preceding argument sidesteps this difficulty, at the expense of the loss by a factor $1/2$ in the rate of convergence.

A The Cauchy problem and smooth approximations

The goal of this appendix is to discuss the Cauchy problem for equation (1.1). This serves two purposes: first, to ensure that our theorems about trend to equilibrium are relevant and non-vacuous, secondly to complete those arguments which rely on the Bakry-Emery method and *demand sufficient regularity for the Cauchy problem, that one is allowed to twice differentiate the entropy functional with respect to time*, and in particular establish formulae like (2.4) and (3.10). In fact what is needed for the proofs to be complete, is not that equation (1.1) possess smooth solutions, but rather that it can be approximated, in a sense to be made precise, by an equation with smooth solutions. Many of the technical details involved in this study are in fact a reworking of previous ideas, which is why we shall not display them in full detail, but only explain how to deal with them and give precise references.

We shall distinguish three situations, according to the form of the diffusion. The first one, and most natural for many applications, is the linear diffusion case, when $U(s) = s \log s$. With the help of some regularity and convexity assumptions on V and W it is possible to adapt the proofs of Desvillettes and Villani [15] to prove C^∞ smoothness and rapid decay at infinity of the solutions, assuming if necessary that the initial datum be replaced by a rapidly decaying, C^∞ approximation. In reference [15], the proofs are carried out for the so-called spatially homogeneous Landau equation of plasma physics, which bears a lot of similarity with (1.1) for $U(s) = s \log s$, except that the diffusion is nonlocal; the arguments are a bit tedious but rely on a few simple principles which are readily adapted to our situation.

A second important case is the case without diffusion. This one is immediately included as a limit case of the previous one. A direct argument is also possible, with the tools used by Benedetto, Caglioti and Pulvirenti [3]. Note that in this case, no smoothness at all is required in the computations (everything makes sense with combinations of Dirac masses...).

Finally, we also consider the case with nonlinear diffusion. Here we want to include in our analysis degenerate cases such as $U(s) = s^m$ with $m > 1$; we do not treat the (more difficult) singular case $m < 1$. A detailed (and delicate !) study for problems of this type when $W = 0$ was performed in [9]. In the degenerate case, it is well-known that compactly supported initial data may lead to compactly supported solutions, for which the regularity of the boundary of the support is not sufficient to prove (3.10) directly. Thus, in this case it is natural to approximate the degenerate diffusion by a non-degenerate diffusion. This approach has been carried out recently by Otto [27] and Carrillo and Toscani [12] for equation (1.1) in the pure power-law case (when $U(s) = s^m$) without interaction potential and it was later generalized in [9] for general nonlinearities satisfying (1.2) without interaction potential.

We insist on one of the advantages of our "functional" approach: once inequalities such as (2.6) are established for smooth initial data and regularized equations (smooth V, smooth W...) then they can be extended by density to cover most of the relevant situations. See the appendix of Otto and Villani [28] for a typical such density argument.

A.1 Linear diffusion

Let us first focus on the Cauchy problem for equation (1.1) with $U(s) = \sigma s \log s$, that is,

$$\frac{\partial \rho}{\partial t} = \nabla \cdot [\rho \nabla (V + W * \rho)] + \sigma \Delta \rho. \tag{A.1}$$

We shall consider potentials V, W which are bounded below and lie in $W^{2,\infty}_{\rm loc}(\mathbb{R}^d)$. Moreover,

$$V(x) \le C(1+|x|)^{\alpha_0}, \qquad W(z) \le C(1+|z|)^{\beta_0}.$$

We shall assume that there exist constants $\alpha, \beta, C, K > 0$ such that V satisfies

$$\forall x \in \mathbb{R}^d \qquad x \cdot \nabla V(x) \ge -C, \qquad |\nabla V(x)| \le C(1+|x|)^{\alpha}, \tag{A.2}$$

and W satisfies either

$$\forall z \in \mathbb{R}^d \qquad z \cdot \nabla W(z) \ge K(1+|z|)^{1+\beta} - C, \qquad |\nabla W(z)| \le C(1+|z|)^{\beta}, \tag{A.3}$$

or

$$\forall z \in \mathbb{R}^d \qquad z \cdot \nabla W(z) \ge -C, \qquad |\nabla W(z)| \le C(1+|z|). \tag{A.4}$$

In the last case, we set $\beta = 1$. As for the initial datum ρ_0, we shall assume that it satisfies

$$\int |x|^{s_0} \, d\rho_0(x) < +\infty, \qquad F(\rho_0) < +\infty \tag{A.5}$$

for some s_0 such that

$$s_0 \geq 2\max(\alpha, \beta, 1), \qquad s_0 > \max(\alpha_0, \beta_0).$$

Under these assumptions, let us show how to perform the basic a priori estimates on the Cauchy problem for (1.1). Combining these estimates with standard results of existence and regularity for linear parabolic equations (e.g. Friedman [19]), Schauder's theorem, and an approximation procedure (as in Desvillettes and Villani [15] or Carrillo [8]; see also paragraph A.2 where a similar construction is performed in some detail), it is easy to construct smooth, rapidly decaying solutions if V, W and ρ_0 are smooth and well-behaved (say, if V is growing faster than W at infinity, and uniformly convex...).

In the sequel we just indicate how to construct weak solutions when V, W, ρ_0 are not so smooth, by passing to the limit in a sequence of regularized problems. Even in this case, those weak solutions are actually strong under some mild additional assumptions, but we do not care about this here. We shall only sketch the main steps and perform a priori estimates as if we were dealing with smooth solutions. The notations C, K will stand for various positive constants throughout which never depend on ρ_t; they may depend on t however, but are always uniformly controlled on any compact interval $t \in [0, T]$.

Mass conservation:

$$\int d\rho_t(x) = 1.$$

Kinetic energy estimate: from (1.1) and $U(\rho) = \sigma\rho\log\rho$

$$\frac{d}{dt}\int |x|^2\,d\rho_t(x) = -2\int \nabla V(x)\cdot x\,d\rho_t(x) - 2\int (\nabla W * \rho_t)\cdot x\,d\rho(x) + 2d\sigma.$$

Since $z\cdot\nabla V(z) \geq -C$ by (A.2) and mass is conserved, the first term on the right-hand side is bounded above. As for the the second one, since $\nabla W(-z) = -\nabla W(z)$, it can be rewritten as

$$-\int \langle\nabla W(x-y), x-y\rangle\,d\rho_t(x)\,d\rho_t(y).$$

Now $z\cdot\nabla W(z) \geq -C$ either by (A.3) or (A.4), so mass conservation implies this term is also bounded above. On the whole, we have proven the a priori bound

$$\int |x|^2\,d\rho_t(x) \leq \int |x|^2\,d\rho_0(x) + Ct.$$

Remark: Substituting hypotheses (A.2) and (A.4) by

$$\forall x \in \mathbb{R}^d \qquad x\cdot\nabla V(x) \geq -C|x|^2, \qquad |\nabla V(x)| \leq C(1+|x|)^\alpha,$$

and

$$\forall z \in \mathbb{R}^d \qquad z\cdot\nabla W(z) \geq -C|z|^2, \qquad |\nabla W(z)| \leq C(1+|z|).$$

respectively, we can also prove apriori bounds of the energy which grows exponentially in time. This will include concave cases as $V(x) = -\lambda|x|^2$.

Free energy estimate: From the decrease of $F(\rho_t)$ with t and from the fact that V and W are non-negative, one finds

$$\int \rho_t \log \rho_t \, dx \leq \frac{F(\rho_0)}{\sigma} < +\infty.$$

This, together with the kinetic energy estimate, ensures **weak compactness in** L^1 for sequences of approximate solutions to (1.1) (via the well-known Dunford-Pettis criterion).

Moment estimate: Let $s \geq 2\max(\alpha, \beta, 1)$, then

$$\begin{aligned} \frac{d}{dt} \int (1+|x|^2)^{s/2} \, d\rho_t(x) &= -s \int \nabla V(x) \cdot x (1+|x|^2)^{\frac{s-2}{2}} \, d\rho_t(x) \\ &\quad - s \int \nabla(W * \rho_t)(x) \cdot x(1+|x|^2)^{\frac{s-2}{2}} \, d\rho_t(x) + \sigma \int \Delta(1+|x|^2)^{\frac{s}{2}} \, d\rho_t(x). \end{aligned} \tag{A.6}$$

The first and last terms on the right-hand side are immediately bounded from above by $C \int (1+|x|^2)^{\frac{s-2}{2}} \, d\rho_t(x)$, so let us focus on the middle one.

Using assumption (A.3) and our bound on the growth of kinetic energy, one finds

$$\begin{aligned} \nabla(W * \rho_t)(x) \cdot x &= \int \nabla W(x-y) \, d\rho_t(y) \cdot x \\ &= \int \nabla W(x-y) \, d\rho_t(y) \cdot (x-y) + \int \nabla W(x-y) \, d\rho_t(y) \cdot y \\ &\geq K \int (1+|x-y|)^{1+\beta} d\rho_t(y) - C(1+|x|)^{\beta} - C \int (1+|y|)^{1+\beta} d\rho_t(y) \\ &\geq K(1+|x|)^{1+\beta} - C \int (1+|y|)^{1+\beta} \, d\rho_t(y) \end{aligned}$$

where as noted above, C and K may depend on time, but remain uniformly bounded on $t \in [0, T]$. The last inequality follows from the elementary inequality

$$-|x-y|^{1+\beta} \leq -K_\beta(1+|x|)^{1+\beta} + C_\beta(1+|y|)^{1+\beta} \qquad (x, y \in \mathbb{R}^d).$$

After some easy calculations, it follows that the second term in (A.6) can be bounded from above by

$$-K \int (1+|x|)^{s+\beta-1} \, d\rho(x) + C \left[\int (1+|y|)^{\beta+1} \, d\rho_t(y) \right] \left[\int (1+|x|)^{s-2} \, d\rho_t(x) \right].$$

From elementary Lebesgue interpolation and use of the kinetic energy bound, the last term can be bounded by

$$\left(\int (1+|x|)^{s+\beta-1} \, d\rho(x) \right)^{\theta} \left(\int (1+|x|^2) \, d\rho(x) \right)^{1-\theta}$$

for some exponent $\theta \in (0,1)$; hence it can be bounded by $\varepsilon \int (1+|x|)^{s+\beta-1}\, d\rho_t(x) + C_\varepsilon$, where ε is arbitrarily small and C_ε depends on ε and on $\int (1+|x|^2)\, d\rho_t(x)$. Putting together all these estimates, we conclude that, under assumption (A.3),

$$\frac{d}{dt}\int (1+|x|^2)^{s/2}\, d\rho_t(x) \leq -K \int (1+|x|)^{s+\beta-1}\, d\rho_t(x) + C.$$

On the other hand, under assumption (A.4), one finds

$$-\nabla(W*\rho_t)(x)\cdot x \leq C + C\int (1+|x|+|y|)|y|\, d\rho_t(y),$$

which is bounded by $C(1+|x|)$ in view of the kinetic energy estimate; again, the constant C may grow with time but remains bounded on any compact interval $[0,T]$. By (A.6) this easily leads to

$$\frac{d}{dt}\int (1+|x|^2)^{s/2}\, d\rho_t(x) \leq C + C\int (1+|x|)^{s-1}\, d\rho_t(x).$$

In both cases, we obtain an a priori bound of $\int |x|^s\, d\rho_t(x)$, which only depends on t and $\int |x|^s\, d\rho_0(x)$.

In particular, if $\int |x|^\beta\, d\rho_0(x) < +\infty$, then one has a $L^\infty_{\rm loc}$ bound on $\nabla W * \rho_t$. Thus the above estimates are already sufficient to ensure the stability of solutions to (1.1) under weak convergence. Now we have to work just a little bit more to ensure the convergence of the free energy in this process.

Remark: Moment estimates are also available in concave cases as for the kinetic energy estimate under assumptions written in the remark above.

Entropy dissipation estimate: Any smooth solution $(\rho_t)_t \geq 0$ of (1.1) which decays rapidly in space satisfies $dF(\rho_t)/dt = -D(\rho_t) \leq 0$, whence

$$\int_0^T \int \rho_t \left| \sigma\frac{\nabla\rho_t}{\rho_t} + \nabla V + (\nabla W)*\rho_t \right|^2 dt \quad = \quad F(\rho_0) - F(\rho_t). \tag{A.7}$$

Such solutions therefore enjoy the a priori bounds

$$(1-\epsilon)\sigma^2 \int_0^T \int \frac{|\nabla\rho_t|^2}{\rho_t}\, dt \leq F(\rho_0|\rho_\infty) + (\epsilon^{-1}-1)\int_0^T \int \left(|\nabla V|^2 + |(\nabla W)*\rho_t|^2\right) d\rho_t$$

for all $\epsilon > 0$. If we assume that $\int |x|^{s_0}\, d\rho_0(x) < +\infty$ for $s_0 = \max(2\alpha, 2\beta)$, then any convenient choice of ϵ couples with previous steps to yield

$$\int_0^T \int \frac{|\nabla\rho_t|^2}{\rho_t}\, dt \leq C(T,\rho_0).$$

An easy regularization argument shows that this is the same as

$$\int_0^T \|\nabla\sqrt{\rho_t}\|_{L^2}\,dt \le C(T,\rho_0).$$

Now assume a sequence of solutions $(\rho_t^n)_{t\ge 0}$ of (1.1) converges weakly to $(\rho_t)_{t\ge0}$ and satisfyies the bounds proved above for ρ_t, namely

$$\int_0^T \|\nabla\sqrt{\rho_t^n}\|_{L^2}\,dt \le C$$

together with entropy and moment estimates independent of n. We claim that $\int \rho_t^n \log \rho_t^n$ converges to $\int \rho_t \log\rho_t$ for almost all t. This could be achieved by standard but intricate PDE estimates; here, we shall show it by an elementary method, as an application of HWI inequalities. Let $I(\rho) = 4\int|\nabla\sqrt{\rho}|^2$ denote the Fisher information functional, and $H(\rho) = \int \rho\log\rho$ the Boltzmann H functional. Since I is convex and lower semi-continuous on L^1, it is weakly lower semi-continuous, hence

$$\int_0^T \|\nabla\sqrt{\rho_t}\|_{L^2}\,dt \le C.$$

As a consequence of the results in Otto and Villani [28] (c.f. (4.6)), one has, for all $\delta>0$,

$$|H(\rho_t^n)-H(\rho_t)| \le W_2(\rho_t,\rho_t^n)(\sqrt{I(\rho_t^n)}+\sqrt{I(\rho_t)}) \le \delta[I(\rho_t^n)+I(\rho_t)] + \frac{1}{2\delta}W_2^2(\rho_t,\rho_t^n).$$

Now among measures with uniformly bounded moments, the Wasserstein metric is bounded and topologizes weak convergence (see for instance [30]). Thus we deduce that $W_2(\rho_t,\rho_t^n)$ converges to 0 as $n\to\infty$, and

$$\limsup_{n\to\infty}\int_0^T |H(\rho_t^n)-H(\rho_t)|\,dt \le 2\delta \limsup_{n\to\infty}\int_0^T I(\rho_t^n)\,dt,$$

for all $\delta>0$, so that actually

$$\int_0^T |H(\rho_t^n)-H(\rho_t)|\,dt \longrightarrow 0.$$

This implies that $H(\rho_t^n)$ converges towards $H(\rho_t)$ for almost all $t\in[0,T]$. As a consequence of moment bounds, also $F(\rho_t^n)$ converges towards $F(\rho_t)$ for almost all $t\in[0,T]$.

On the other hand, by expanding the square one easily shows that in the same situation,

$$D(\rho_t) \le \liminf_n D(\rho_t^n).$$

Consequently, these a priori estimates, Fatou's lemma and (A.7) combine to allow one to construct weak solutions of (1.1) such that for almost all $s<t$,

$$0 \le \int_s^t D(\rho_\tau)\,d\tau \le F(\rho_s)-F(\rho_t). \tag{A.8}$$

This shows $F(\rho_t)$ to be nonincreasing, and proves (2.4) wherever $F(\rho_t)$ is differentiable.

Remark: In all the cases which are considered in the theorems of Section 2, one has the additional a priori estimates

$$D(\rho_t) \leq \frac{C}{t^2}, \qquad F(\rho_t) \leq \frac{C}{t},$$

which are sufficient to prove equality in (A.8) and hence in (2.4).

Finally, the argument for the construction of very smooth solutions follows by similar a priori estimates, in the manner of Desvillettes and Villani [15]. For instance, when V is smooth and convex, and $W(z) = |z|^{2+\gamma}$ for some $\gamma > 0$, then one can prove C^∞ smoothness and rapid decay of the solutions (uniformly in time). With this result at hand, one can easily work out, for instance, the Bakry-Emery strategy for generalized logarithmic Sobolev inequalities in the same way as in Carrillo, Jüngel, Markowich, Toscani and Unterreiter [9] (in this reference, the authors always perform the Bakry-Emery strategy on very smooth solutions of approximate problems).

The case when there is no diffusion is even easier to treat: the moment bounds are even stronger than the diffusive case, they yield convergence of $F(\rho_t^n) \to F(\rho_t)$ directly, without a discussion of the Fisher information or entropy $H(\rho)$, but the weak compactness takes place in the space of Borel probability measures rather than in L^1. Alternately, let $\sigma \to 0$ in the previous estimates, and dispense with those that become irrelevant. (In fact, one could prove propagation of smoothness in the no-diffusion case, at least if V and W are regular enough: note that in this case, ρ_t is just the push-forward of ρ_0 by the characteristic field which is associated to the smooth velocity field $-\nabla(V + W * \rho_t)$.

A.2 Nonlinear diffusion

We finally give some details about the treatment of the case with nonlinear diffusion, which is more delicate, in particular because of the lack of smoothness which is associated with degenerate nonlinearities — we shall only focus on this case, even though it is certainly possible to also consider the case of nondegenerate singular diffusions. We recommend Vazquez [33] as a good source for the theory of porous-medium type equations.

We now assume that the non-negative confinement potential V and interaction potential W still satisfy the same assumptions as before, but in addition we impose that

$$V \text{ is strictly convex,} \tag{A.9}$$

$$W(z) \text{ is a function of } |z|. \tag{A.10}$$

Let $R > 0$ be a large positive number, we denote by B_R the set $\{x \mid V(x) \leq R\}$. This is a strictly convex domain of $\mathbb{R}^d$. For $x \in \partial B_R$ we shall denote by $n(x)$ the unit outwards normal vector to ∂B_R at x. Note in particular that

$$n(x) \cdot \nabla V(x) \geq 0 \tag{A.11}$$

on ∂B_R, since $n = \nabla V / |\nabla V|$.

In addition, we assume that U is a strictly convex function on $\mathbb{R}^+$, $U(0) = 0$, $U \in C^4(0, +\infty)$, with a finite right-derivative at 0,

$$\frac{U(s)}{s} \xrightarrow[s \to +\infty]{} +\infty. \tag{A.12}$$

Moreover, $s \mapsto sU''(s)$ should be nondecreasing for $s > 0$ small enough. We define

$$P(\rho) = \int_0^\rho sU''(s)\,ds, \tag{A.13}$$

$$h(\rho) \equiv \int_1^\rho \frac{P'(s)}{s}\,ds; \tag{A.14}$$

from our assumptions it follows that $h \in L^1_{loc}([0, +\infty))$ and $h(\rho) = U'(\rho) - U'(1)$ for all $\rho \geq 0$. In particular,

$$-\infty < h(0+), \qquad h(+\infty) = +\infty, \qquad \exists s_0 > 0;\ \forall s \in (0, s_0),\ P''(s) \geq 0. \tag{A.15}$$

Let ρ_0 be a nonnegative initial datum satisfying

$$\rho_0 \in L^\infty(B_R). \tag{A.16}$$

Here as in the previous section, we identify the measure ρ_0 with its density $d\rho_0/dx$. We consider the initial boundary value problem for equation (1.1) in B_R

$$\frac{\partial \rho}{\partial t} = \nabla \cdot \left[\rho \nabla \left(U'(\rho) + V + W * \bar\rho\right)\right], \tag{A.17}$$

with no-flux boundary conditions

$$n(x) \cdot \Big(\rho \nabla (V + W * \bar\rho) + \nabla P(\rho)\Big)(t, x) = 0, \qquad (x \in \partial B_R,\ t > 0), \tag{A.18}$$

where $\bar\rho$ indicates the extension to $\mathbb{R}^d$ by zero of the function ρ defined in B_R. Hence

$$\forall x \in B_R \quad W * \bar\rho(x) = \int_{B_R} W(x - y)\rho(y)\,dy.$$

Problem (A.17)–(A.18), supplemented with the initial condition ρ_0, will be denoted by $(IBVP_R)$. Without additional assumptions on the potentials we are able to prove existence and uniqueness of a weak solution to $(IBVP_R)$. However, we also need to show that smooth solutions can be constructed for regularized problems.

Regularize V, W, ρ_0 into V^ϵ, W^ϵ, ρ_0^ϵ by convolving them with a radially symmetric mollifier. Also, let P^ϵ be a sequence of nondegenerate diffusion functions approximating

P; here "nondegenerate" means that $h^\epsilon(0+) = -\infty$. The construction of such an approximation is performed in Carrillo, Jüngel, Markowich, Toscani and Unterreiter [9] or Otto [27]. Now, we consider the regularized problem $(IBVP_R^\epsilon)$ consisting of

$$\frac{\partial \rho^\epsilon}{\partial t} \;=\; \nabla \cdot \left[\rho^\epsilon \nabla \left(U^{\epsilon\prime}(\rho^\epsilon) + V^\epsilon + W^\epsilon * \bar{\rho}^\epsilon\right)\right], \tag{A.19}$$

$$n(x) \cdot (\rho^\epsilon \nabla (V^\epsilon + W^\epsilon * \bar{\rho}^\epsilon) + \nabla P^\epsilon(\rho^\epsilon))(x,t) = 0, \qquad (x \in \partial B_R, t > 0)$$

with initial datum ρ_0^ϵ.

Let us prove first the existence of a solution to problem $(IBVP_R^\epsilon)$ by standard quasilinear parabolic theory. Let us consider the set $\mathcal{S}_T$ of positive continuous functions $\varphi \in C([0,T] \times \bar{B_R})$ such that

$$\|\varphi(t)\|_{L^1(B_R)} \;=\; \|\rho_0^\epsilon\|_{L^1(B_R)}$$

and

$$\|\varphi\|_{L^\infty((0,T)\times B_R)} \;\leq\; M(T, \rho_0)$$

for any $t \geq 0$ and any $T > 0$, where M only depends on ρ_0 and T and is to be specified later. Given $\varphi \in \mathcal{S}_T$ we define $Z(\varphi)$ to be the unique classical solution (see Ladyzenskaja, Solonnikov and Ural'ceva [21, Chapter V, Section 7]) of problem

$$\frac{\partial \rho^\epsilon}{\partial t} = \nabla \cdot \left[\rho^\epsilon \nabla \left((U^\epsilon)'(\rho^\epsilon) + V^\epsilon + W^\epsilon * \bar{\varphi}\right)\right],$$

$$n(x) \cdot \Big(\rho^\epsilon \nabla (V^\epsilon + W^\epsilon * \bar{\varphi}) + \nabla P^\epsilon(\rho^\epsilon)\Big)(t,x) = 0, \qquad (x \in \partial B_R, t > 0)$$

with initial condition ρ_0^ϵ. Note that this is now a quasilinear problem, since we have frozen the self-interaction term.

From (A.10–A.11), (A.15–A.16) and the convexity of B_R, one can prove that φ is positive and

$$g^\epsilon(t) = C \exp\left(\|\Delta V^\epsilon\|_{L^\infty(B_R)} t + \|\Delta W^\epsilon\|_{L^\infty(B_R)} \|\rho_0^\epsilon\|_{L^1(B_R)} t\right)$$

is a supersolution of this linear problem for C large enough. Therefore, we can choose $M(T, \rho_0) = \sup\{g^\epsilon(T)\,,\, \epsilon > 0\} < +\infty$. Now, since the mass of the solution is preserved and 0 is obviously a subsolution, then we have proved that Z maps $\mathcal{S}_T$ into $\mathcal{S}_T$. Furthermore, by standard quasilinear parabolic theory (see [21, Theorem 7.2]), the map Z is continuous and (due to the uniform L^∞ estimate for any $\varphi \in \mathcal{S}_T$) its range is a family of uniformly Hölder continuous (in particular equicontinuous) functions. Therefore, Schauder's fixed point theorem implies the existence of a classical solution for problem $(IBVP_R^\epsilon)$. The uniqueness for problem $(IBVP_R^\epsilon)$ follows from an L^1 contraction estimate contained in Bertsch and Hilhorst [5, Theorem 4.1].

Now, let us take the limit $\epsilon \to 0$. First, we show that $\nabla P^\epsilon(\rho^\epsilon)$ is uniformly (with respect to ϵ) bounded in L^2. Multiplying equation (A.19) by $P^\epsilon(\rho^\epsilon)$ and using the divergence theorem, one finds

$$\frac{\partial}{\partial t}\int_{B_R} \Pi^\epsilon(\rho^\epsilon)\,dx \;=\; -\int_{B_R} |\nabla P^\epsilon(\rho^\epsilon)|^2\,dx + \int_{B_R} \Phi^\epsilon(\rho^\epsilon)(\Delta V^\epsilon + \Delta W^\epsilon * \bar{\rho^\epsilon})\,dx \tag{A.20}$$

$$-\int_{\partial B_R} \Phi^\epsilon(\rho^\epsilon)(\nabla(V^\epsilon + W^\epsilon * \bar{\rho^\epsilon})\cdot n(x))\,dS(x) \tag{A.21}$$

where $\Pi^\epsilon(\rho) = \int_0^\rho P^\epsilon(s)\,ds$ and $\Phi^\epsilon(\rho) = \int_0^\rho sP'^\epsilon(s)$. By using (A.10–A.11), the convexity of B_R, and the positivity of φ, we deduce

$$\frac{\partial}{\partial t}\int_{B_R} \Pi^\epsilon(\rho^\epsilon)\,dx \le \int_{B_R} \Phi^\epsilon(\rho^\epsilon)(\Delta V + \Delta W * \bar{\rho^\epsilon})\,dx - \int_{B_R} |\nabla P^\epsilon(\rho^\epsilon)|^2\,dx. \tag{A.22}$$

By the L^∞ uniform bound we deduce that the first term in the right-hand side of (A.22) is bounded uniformly in ϵ for any $0 < t < T$. Therefore, we have a uniform bound

$$\int_{B_R} \Pi^\epsilon(\rho^\epsilon)\,dx \le C(t, \rho_0)$$

and by integrating over $[\tau, T]$ we obtain a uniform bound on

$$\int_\tau^T \int_{B_R} |\nabla P^\epsilon(\rho^\epsilon)|^2\,dx\,dt.$$

Again, by the same L^1-contraction estimate as above, one can prove the uniqueness of the solution to problem $(IBVP_R)$.

Now, we can use an equicontinuity and regularity result of E. DiBenedetto [17] for general quasilinear parabolic equations with degenerate diffusion. Since the sequence ρ^ε of approximate solutions is uniformly bounded in $L^\infty(Q_T)$ with $\nabla P^\epsilon(\rho^\epsilon)$ uniformly bounded in $L^2(Q_T)$, the result of DiBenedetto implies that for any τ, T $(0 < \tau < T)$, ρ^ϵ is an equicontinuous family of functions in $C([\tau, T] \times \bar{B_R})$. With this auxiliary result at hand, it is easy to prove that the limit of a subsequence is a weak solution (in distributional sense) of problem $(IBVP_R)$. For more details, we refer to [5, 17, 9].

As we mentioned earlier, from this point, the proof of a generalized Log-Sobolev inequality by the Bakry-Emery method follows exactly the lines laid out by Carillo, Jüngel, Markowich, Toscani and Unterreiter in [9]. Formula (3.10) is proven for $(IBVP_R^\epsilon)$ directly since the solutions are smooth enough. The existence of equilibrium solutions for the problem in the bounded convex domain B_R is ensured by Theorem 3.1 in [25]. Generalized Logarithmic Sobolev inequalities and entropy decay estimates are shown for $(IBVP_R^\epsilon)$. Taking the limit $\epsilon \to 0$, one proves generalized Log-Sobolev inequalities; the entropy decay estimates for $(IBVP_R)$ follow by the compactness result discussed above. In the end, one takes the limit $R \to \infty$. Generalized Sobolev inequalities are finally shown as in [9, Theorem 17]. For the existence of solution and entropy decay estimates one either proves L^∞ estimates (for instance, if the initial data is compactly supported and bounded by a stationary state, see [9, Theorem 19]) or moment bounds. Also, once generalized Sobolev inequalities are proven, asymptotic stability theorems stating minimal requirements over solutions of problem (1.1) can be obtained (see [9, Theorems 20–21]).

Remarks:

1. Truncating the problem to a sublevel set B_R of V amounts to replace V by $+\infty$ outside of B_R (infinitely strong confinement). The argument of McCann [25] which we used to prove uniqueness of the minimizer of F still holds in this case.

2. In the nondegenerate singular diffusion case, in which $-\infty = h(0+)$ and $h(\infty) = \infty$ (fast-diffusion type equations), the same procedure sketched above can be performed, but it is not so clear that there exists a unique equilibrium ρ_∞ with finite free energy (see assumptions **(HV4)**-**(HV6)** in [9]).

Acknowledgements: These results were first presented at the 2000 TMR Summer School "Mass Transportation Methods in Kinetic Theory and Hydrodynamics" organized by Maria Carvalho in Azores. We are grateful to Maria and the TMR for having the vision and energy to arrange this beautiful event. JAC and CV acknowledge support from the TMR-project "Asymptotic Methods in Kinetic Theory", No. ERBFRMXCT 970157, funded by the European Community. JAC also acknowledges support from DGES (Spain) PB98-1281 and PB98-1294, while RJM acknowledges the support of Natural Sciences and Engineering Research Council of Canada grant 217006-99 RGPIN and United States National Science Foundation grant DMS 0074037. RJM and CV extend their thanks to Wilfrid Gangbo and members of the Georgia Institute of Technology School of Mathematics for their kind hospitality and the stimulating environment they provided during various stages of this work. Both profited from a semester spent at Georgia Tech: in particular, the first ideas for this work arose during the fall of 1999 when CV was teaching a graduate course on mass transportation techniques there. Finally, we thank Dario Cordero-Erausquin for a simple but important remark which enabled us to improve the rate constant in Theorem 2.2 by a factor of 2.

References

[1] Arnold, A., Markowich, P., Toscani, G., and Unterreiter, A. On convex Sobolev inequalities and the rate of convergence to equilibrium for Fokker-Planck type equations. *Comm. Partial Differential Equations 26*, 1–2 (2001), 43–100.

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

[3] Benedetto, D., Caglioti, E., and Pulvirenti, M. A kinetic equation for granular media. *RAIRO Modél. Math. Anal. Numér. 31*, 5 (1997), 615–641.

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

[5] Bertsch, M., and Hilhorst, D. A density dependent diffusion equation in population dynamics: stabilization to equilibrium. *SIAM J. Math. Anal. 17* (1986), 863-883.

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

[7] Carlen, E., and Carvalho, M. Entropy production estimates for Boltzmann equations with physically realistic collision kernels. *J. Statist. Phys. 74*, 3-4 (1994), 743–782.

[8] Carrillo, J. A. On a 1-D granular media immersed in a fluid. *Fields Inst. Comm. 27* (2000), 43-56.

[9] Carrillo, J. A., Jüngel, A., Markowich, P. A., Toscani, G., and Unterreiter, A. Entropy dissipation methods for degenerate parabolic systems and generalized Sobolev inequalities. *Monatsh. Math.* 133 (2001) 1–82.

[10] Carrillo, J. A., McCann, R., and Villani, C. Kinetic equilibration rates for granular media: geometric approach. Work in progress.

[11] Carrillo, J. A., and Toscani, G. Exponential convergence toward equilibrium for homogeneous Fokker-Planck-type equations. *Math. Meth. Appl. Sci. 21*, (1998), 1269-1286.

[12] 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*, (2000), 113-141.

[13] Cordero-Erausquin, D. Some applications of mass transport. To appear in *Arch. Rational Mech. Anal.*

[14] Cordero-Erausquin, D., Gangbo, W., Houdre, C. Inequalities for generalized entropy and optimal transportation. To appear in Proceedings of the Workshop: Mass Transportation Methods in Kinetic Theory and Hydrodynamics.

[15] Desvillettes, L., and Villani, C. On the spatially homogeneous Landau equation for hard potentials. Part I : Existence, uniqueness and smoothness. *Comm. Partial Differential Equations 25*, 1–2 (2000), 179–259.

[16] DESVILLETTES, L., AND VILLANI, C. On the spatially homogeneous Landau equation for hard potentials. Part II : H-theorem and applications *Comm. Partial Differential Equations 25*, 1–2 (2000), 261–298.

[17] DIBENEDETTO, E. Continuity of weak solutions to a general porous medium equation. *Indiana Math. Univ. J.* **32** (1983), 83-118.

[18] DOLBEAULT, J., AND DEL PINO, M. Generalized Sobolev inequalities and asymptotic behaviour in fast diffusion and porous medium problems. Preprint, 1999.

[19] FRIEDMAN, A. Partial differential equations of parabolic type. Prentice-Hall, Englewood Cliffs, N.J. 1964.

[20] HOLLEY, R., AND STROOCK, D. Logarithmic Sobolev inequalities and stochastic Ising models. *J. Statist. Phys. 46*, 5–6 (1987), 1159–1194.

[21] LADYZENSKAJA, O. A., SOLONNIKOV, V. A., AND URAL'CEVA, N. N. Linear and Quasilinear Equations of Parabolic Type. AMS, Providence, Rhode Island. 1968.

[22] MALRIEU, F. Logarithmic Sobolev inequalities for some nonlinear PDE's. Preprint, 2001.

[23] MARKOWICH, P., AND VILLANI, C. On the trend to equilibrium for the Fokker-Planck equation: an interplay between physics and functional analysis. Proceedings, VI Workshop on Partial Differential Equations, Part II (Rio de Janeiro, 1999). *Mat. Contemp. 19* (2000), 1–29.

[24] MCCANN, R. J. Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J. 80* (1995), 309–323.

[25] MCCANN, R. J. A convexity principle for interacting gases. *Adv. Math. 128*, 1 (1997), 153–179.

[26] MCCANN, R. J. Equilibrium shapes for planar crystals in an external field. *Comm. Math. Phys. 195*, (1998) 699–723.

[27] OTTO, F. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations 26* (2001) 101–174.

[28] 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.

[29] OTTO, F., AND VILLANI, C. Comment on "Hypercontractivity of Hamilton-Jacobi equations", by S. Bobkov, I. Gentil and M. Ledoux. To appear in *J. Maths Pures Appl.*

[30] Rachev, S. and Rüschendorf, L. *Mass Transportation Problems*, Springer-Verlag, Coll. Probability and its applications (1998).

[31] Toscani, G. One-dimensional kinetic models of granular flows. *RAIRO Modél. Math. Anal. Numér. 34*, 6 (2000), 1277–1291.

[32] Toscani, G., and Villani, C. On the trend to equilibrium for some dissipative systems with slowly increasing a priori bounds. *J. Statist. Phys. 98*, 5–6 (2000), 1279–1309.

[33] Vázquez, J. L. An introduction to the mathematical theory of the porous medium equation. Shape optimization and free boundaries (Montreal, PQ, 1990). *NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci. 380*, 347–389, Kluwer Acad. Publ., 1992.

[34] Villani, C. A survey of mathematical topics in the collisional kinetic theory of gases. To appear in *Handbook of Fluid Mechanics*, S. Friedlander and D. Serre, Eds.

[35] Villani, C. Regularity estimates via the entropy dissipation for the spatially homogeneous Boltzmann equation. *Rev. Mat. Iberoam. 15*, 2 (1999), 335–352.