WORKING PAPERS

N° 17-787

March 2017

“A family of functional inequalities: lojasiewicz inequalities and displacement convex functions"

Adrien Blanchet and Jérôme Bolte

Toulouse School of Economics

A FAMILY OF FUNCTIONAL INEQUALITIES: ŁOJASIEWICZ INEQUALITIES AND DISPLACEMENT CONVEX FUNCTIONS

ADRIEN BLANCHET & JÉRÔME BOLTE

Abstract. For displacement convex functionals in the probability space equipped with the Monge-Kantorovich metric we prove the equivalence between the gradient and functional type Łojasiewicz inequalities. In a second part, we specialise these inequalities to some classical geodesically convex functionals. For the Boltzmann entropy, we obtain the equivalence between logarithmic Sobolev and Talagrand's inequalities. On the other hand, the non-linear entropy and the Gagliardo-Nirenberg inequality provide a Talagrand inequality which seems to be a new equivalence. Our method allows also to recover some results on the asymptotic behaviour of the associated gradient flows.

1. Introduction

Łojasiewicz inequalities have long been known to be extremely powerful tools for studying the long-time behaviour of a dynamical system in an Euclidean or Hilbert space, see e.g., [29, 19, 14, 7] and references therein. On the other hand, in the optimal transport community a lot of excitation followed the discovery of the "Riemannian structure" behind the set of probability measures endowed with the Monge-Kantorovich distance, see [26, 1, 18, 32, 21]. In this article we aim at making some connections between these two research fields through flows and functional inequalities. We indeed formulate Łojasiewicz inequalities in the probability space equipped with the Monge-Kantorovich distance and we prove their equivalence in the case of displacement convex functionals. Both types of Łojasiewicz inequalities can be viewed as families of abstract functional inequalities. Hence for specific choices of functionals their equivalence translates into the equivalence between functionals inequalities as famous as the logarithmic Sobolev, Talagrand type or the Gagliardo-Nirenberg inequalities. Some of these connections were known, see e.g., [27], we recover them here in a transparent and unified way. To the best of our knowledge the equivalence between Gagliardo-Sobolev inequality and the Talagrand type inequality (23) is new.

On the other hand, it is well known that functional inequalities are crucial tools to study the large time behaviour of numerous partial differential equations or of their corresponding stochastic processes, see e.g., [3, 12, 11, 24, 23, 2, 15]. In optimal transport, the interplay between functional inequalities and dynamical systems has also been the subject of many studies see e.g., [5, 10, 15, 17]. Our approach brings new insights into these stabilization phenomena and provides general convergence rate results in the presence of a continuum of equilibria (see Theorem 2).

In Section 2, we state two fundamental types of Łojasiewicz inequalities and prove that they are equivalent for displacement convex functions. General convergence rate results for subgradient systems are also provided. Our results do not subsume the uniqueness of a minimiser. In Section 3 we translate these inequalities into functional inequalities in the case of the Boltzmann and non-linear entropies.

Notations, definitions and classical results on Monge-Kantorovich metric and optimal transportation are provided in Section 4. For a thorough exposition see [1].

Date: February 24, 2017.

Key words and phrases. Łojasiewicz inequality, Functional inequalities, Gradient flows.

2. Łojasiewicz inequalities for displacement convex functions

2.1. **Main results.** Let $\mathcal{P}_2(\mathbb{R}^d)$ be the set of probability measures in $\mathbb{R}^d$ with bounded second moments. In the sequel $\mathcal{J} : \mathcal{P}_2(\mathbb{R}^d) \to (-\infty, +\infty]$ is a lower semi-continuous displacement convex functional. This implies in particular that the domain of $\mathcal{J}$, denoted by $\operatorname{dom} \mathcal{J}$, is convex in the sense of optimal transport:

$$\text{For any } \mu, \nu \text{ in } \operatorname{dom} \mathcal{J}, \text{ the McCann interpolant between } \mu \text{ and } \nu \text{ exists and lies in } \operatorname{dom} \mathcal{J}. \quad (1)$$

The functional $\mathcal{J}$ is called *proper* if $\operatorname{dom} \mathcal{J} \neq \emptyset$.

Two inequalities of Łojasiewicz type. Assume that $\mathcal{J}$ has at least a minimiser, set $\hat{\mathcal{J}} = \min\{\mathcal{J}[\rho] : \rho \in \mathcal{P}_2(\mathbb{R}^d)\}$ and fix $r_0 \in (\hat{\mathcal{J}}, +\infty]$, $\theta \in (0,1]$. We consider the two following properties:

Property 1 (Łojasiewicz gradient property). There exists $c_g > 0$ such that for all $\rho \in \mathcal{P}_2(\mathbb{R}^d)$,

$$\mathcal{J}[\rho] - \hat{\mathcal{J}} < r_0 \quad \Rightarrow \quad \forall \nu \in \partial \mathcal{J}[\rho]\,,\ c_g \left(\mathcal{J}[\rho] - \hat{\mathcal{J}}\right)^{1-\theta} \leq \|\nu\|. \tag{2}$$

Property 2 (Functional Łojasiewicz inequality). There exists $c_f > 0$ such that for all $\rho \in \mathcal{P}_2(\mathbb{R}^d)$,

$$\mathcal{J}[\rho] - \hat{\mathcal{J}} < r_0 \quad \Rightarrow \quad c_f\, \mathcal{W}_2(\rho, \operatorname{Argmin} \mathcal{J})^{1/\theta} \leq \mathcal{J}[\rho] - \hat{\mathcal{J}}. \tag{3}$$

Remark 1. (a) Original inequalities for analytic and subanalytic functions can be found in the IHES lectures [30] by Łojasiewicz. Several generalisations of gradient inequalities followed, see in particular [29, 22, 6].
(b) In this infinite dimensional setting, the above Łojasiewicz inequalities should be thought as families of functional inequalities. Formal connections with the Talagrand, logarithmic Sobolev and Gagliardo-Nirenberg inequalities are provided in Section 3.

The displacement convexity of $\mathcal{J}$ induces existence and uniqueness of the associated gradient flow but also convergence of the minimising scheme, stability, contraction of the semi-group, regularising effects (see [1, Chapter 11] and references therein). These properties are also true without assuming (1) but for a more restrictive notion of convexity associated to generalised geodesics, this is the subject of a forthcoming work.

We now state our two main theorems.

Theorem 1 (Equivalence between Łojasiewicz inequalities in $\mathcal{P}_2(\mathbb{R}^d)$)**.** *Let $\mathcal{J}$ be a proper lower semi-continuous displacement convex functional which has at least a minimiser. Then the Łojasiewicz gradient property* (2) *and the Łojasiewicz functional property* (3) *are equivalent.*

Remark 2. (a) Theorem 1 does not provide a simple relationship between c_f and c_g and the optimality of the constant might be lost. Indeed our proof only gives $c_g = c_f^\theta$ when one establishes (3) $\Rightarrow$ (2), while $c_f = (\theta c_g)^{1/\theta}$ when (2) $\Rightarrow$ (3) is proved.
Yet, when $\theta = 1$ the equivalence between (2) and (3) holds with $c_f = c_g$.
(b) Assume for simplicity that $\hat{\mathcal{J}} = 0$. The extension of the gradient Łojasiewicz inequality by Kurdyka [22] in our setting writes

$$\mathcal{J}[\rho] \in (0, r_0) \quad \Rightarrow \quad \forall \nu \in \partial \mathcal{J}[\rho]\,,\ 1 \leq \varphi'(\mathcal{J}[\rho]) \|\nu\|, \tag{4}$$

for some $\varphi \in C^0[0, r_0) \cap C^1(0, r_0)$ is such that $\varphi(0) = 0$ and $\varphi' > 0$ on $(0, r_0)$. Note that (4) is nothing but inequality (2) with $\varphi(s) = s^\theta / (\theta c_g)$. As in Theorem 1 one can prove that this implies:

$$\mathcal{J}[\rho] < r_0 \quad \Rightarrow \quad \varphi^{-1}(\mathcal{W}_2(\rho, \operatorname{Argmin} \mathcal{J})) \leq \mathcal{J}[\rho]. \tag{5}$$

Yet (4) and (5) are not in general equivalent. One can build a C^2 convex coercive function in $\mathbb{R}^2$ which satisfies (5) but not (4) (see [8]). The construction is fairly complex, involves highly oscillatory behaviour of level sets and originates in [7]. This limitation helps to understand the discrepancy between constants mentioned in item (a) of this remark.

Our second result concerns the flow of $-\partial\mathcal{J}$ and is simply a Monge-Kantorovich version of the classical Hilbertian results. For a given proper lower semi-continuous functional $\mathcal{J} : \mathcal{P}_2(\mathbb{R}^d) \to (-\infty, +\infty]$ which is displacement convex, and $\rho_0 \in \operatorname{dom}\mathcal{J}$, we consider absolutely continuous solutions $\rho : [0, +\infty) \to \mathcal{P}_2(\mathbb{R}^d)$ to the subgradient dynamics:

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho(t) + \partial\mathcal{J}[\rho(t)] \ni 0, \text{ for almost all } t \text{ in } (0, +\infty), \tag{6}$$

with initial condition $\rho(0) = \rho_0$. Such a solution exists and is unique see [1, Theorem 11.3.2, p.305] and [1, Theorem 11.1.4, p.285]. Moreover by the energy identity [1, Theorem 11.3.2, (11.3.7), p.305], one has:

$$\mathcal{J}[\rho(\cdot)] \text{ is non-increasing, absolutely continuous,} \tag{7}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{J}[\rho(t)] = -\left\|\frac{\mathrm{d}}{\mathrm{d}t}\rho(t)\right\|^2 \text{ a.e. on } (0, +\infty). \tag{8}$$

As recalled in the introduction it is well known that functional inequalities are key tools for the asymptotic study of dissipative systems of gradient type. If Łojasiewicz inequalities are seen as families of functional inequalities, the theorem below could be considered as an abstract principle to deduce the convergence of a gradient flow from a given functional inequality[1].

Theorem 2 (Global convergence rates and functional inequalities). *Assume that $\mathcal{J}$ is a proper lower semi-continuous displacement convex function, has at least a minimiser and satisfies* (3). *Consider a trajectory of* (6) *starting at a measure ρ_0 such that $\mathcal{J}(\rho_0) < r_0$.*

Then this trajectory has a finite length and converges to a minimiser ρ_∞ of $\mathcal{J}$ for the Monge-Kantorovich metric. Moreover the following estimations hold:

(i) If $\theta \in (0, c1/2)$,

$$\mathcal{J}[\rho(t)] - \hat{\mathcal{J}} \leq \left\{(\mathcal{J}[\rho_0] - \hat{\mathcal{J}})^{2\theta-1} + c_g^2(1-2\theta)t\right\}^{-\frac{1}{1-2\theta}},$$

$$\mathcal{W}_2(\rho(t), \rho_\infty) \leq \frac{1}{c_g\theta}\left\{(\mathcal{J}[\rho_0] - \hat{\mathcal{J}})^{2\theta-1} + c_g^2(1-2\theta)t\right\}^{-\frac{\theta}{1-2\theta}}.$$

(ii) If $\theta = 1/2$,

$$\mathcal{J}[\rho(t)] - \hat{\mathcal{J}} \leq \mathcal{J}[\rho_0]\exp[-c_g^2 t],$$

$$\mathcal{W}_2(\rho(t), \rho_\infty) \leq \frac{2}{c_g}\sqrt{\mathcal{J}[\rho_0]}\exp[-c_g^2 t/2].$$

(iii) If $\theta \in (1/2, 1]$ we observe a finite time stabilisation: The final time is smaller than

$$T = \frac{(\mathcal{J}[\rho_0] - \hat{\mathcal{J}})^{2\theta-1}}{c_g^2(2\theta-1)}.$$

When t is in $[0, T]$

$$\mathcal{J}[\rho(t)] - \hat{\mathcal{J}} \leq \left((\mathcal{J}[\rho_0] - \min\mathcal{J})^{2\theta-1} - c_g^2(2\theta-1)t\right)^{\frac{1}{2\theta-1}},$$

$$\mathcal{W}_2(\rho(t), \rho_\infty) \leq \frac{1}{c_g\theta}\left((\mathcal{J}[\rho_0] - \min\mathcal{J})^{2\theta-1} - c_g^2(2\theta-1)t\right)^{\frac{\theta}{2\theta-1}}.$$

Remark 3. (a) The above theorem relies on a "zero-order" functional inequality in the sense that inequality (3) only involves the value of the function $\mathcal{J}$ and no higher order information.
(b) A second fundamental feature of this convergence result is that *it does not subsume the uniqueness of a minimiser* which is uncommon in the domain.
(c) It can be proved that the generalised Łojasiewicz gradient inequality of Remark 2 (b), often called Kurdyka-Łojasiewicz property, allows as well to study the convergence of gradient system (6).

[1]See Section 3 in which some classical functional inequalities are interpreted as Łojasiewicz inequalities

2.2. **Proofs of the main results.**

Lemma 1 (Slope and convexity)**.** *Let $\mathcal{J}$ be a proper lower semi-continuous displacement convex functional and $\mu \in \operatorname{dom} \partial J$, $\nu \in \operatorname{dom} \mathcal{J}$ two distinct probabilities in $\mathbb{R}^d$. Then*

$$\frac{\mathcal{J}[\mu]-\mathcal{J}[\nu]}{\mathcal{W}_2(\mu,\nu)} \le \|\partial^0 \mathcal{J}[\mu]\|. \tag{9}$$

Proof. Set $\alpha = \mathcal{W}_2(\mu,\nu)$. Let $[0,\alpha] \ni t \to \rho_t := \mu_{t/\alpha}$ where μ_t is McCann's interpolant between μ and ν (recall assumption (1)). Since $t \to \mathcal{J}[\rho_t]$ is convex, we have

$$\frac{\mathcal{J}[\rho_\alpha]-\mathcal{J}[\rho_0]}{\alpha} \le \limsup_{\tau\to\alpha} \frac{\mathcal{J}[\rho_\alpha]-\mathcal{J}[\rho_\tau]}{\alpha-\tau}.$$

Recalling that $\mathcal{W}_2(\rho_t,\rho_\tau) = |t/\alpha - \tau/\alpha|\mathcal{W}_2(\mu,\nu) = |t-\tau|$ for all time $(t,\tau) \in [0,\alpha]^2$, the above can be rewritten as

$$\frac{\mathcal{J}[\nu]-\mathcal{J}[\mu]}{\mathcal{W}_2(\mu,\nu)} \le \limsup_{\tau\to\alpha} \frac{\mathcal{J}[\nu]-\mathcal{J}[\rho_\tau]}{\mathcal{W}_2(\nu,\rho_\tau)} \le |\nabla|\mathcal{J}[\nu] = \|\partial^0\mathcal{J}[\nu]\|,$$

where the last equality follows from (29).

□

Proof of Theorem 1 With no loss of generality, we assume that $\min \mathcal{J} = 0$.

• (3) ⇒ (2). If $\rho \in \operatorname{Argmin} \mathcal{J}$ there is nothing to prove. Assume that $\mathcal{J}[\rho] \in (0, r_0)$, take ν in $\partial\mathcal{J}[\rho]$. By Lemma 1, for any $\bar\rho$ in Argmin $\mathcal{J}$, we have

$$\mathcal{J}[\rho] \le \mathcal{W}_2(\rho,\bar\rho)\,\|\nu\|\,,$$

and thus

$$\mathcal{J}[\rho] \le \mathcal{W}_2(\rho, \operatorname{Argmin} \mathcal{J})\,\|\nu\|\,.$$

By Łojasiewicz functional property (3), we obtain

$$\mathcal{J}[\rho] \le \left[\frac{1}{c_f}\mathcal{J}[\rho]\right]^\theta \|\nu\|\,,$$

or equivalently $c_f^\theta \mathcal{J}[\rho]^{1-\theta} \le \|\nu\|$. This is the claimed result with $c_g = c_f^\theta$.

• (2) ⇒ (3). Take ρ_0 with $\mathcal{J}(\rho_0) \in (0, r_0)$ and consider the dynamics

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho(t) + \partial\mathcal{J}[\rho(t)] \ni 0 \;\text{ with } \rho(0) = \rho_0. \tag{10}$$

Set $\bar t = \sup\{t : \mathcal{J}[\rho(\tau)] > 0, \forall\tau \in [0,t)\}$. If $\bar t < +\infty$, the continuity of $\mathcal{J}[\rho(\cdot)]$ ensures that $\mathcal{J}[\rho(\bar t)] = 0$. By (7), one has $\mathcal{J}[\rho(t)] = 0$ for all $t \ge \bar t$. Thus by integrating (8) over $(\bar t, \tau)$ (with $\tau \ge \bar t$)

$$\mathcal{J}[\rho(\bar t)] - \mathcal{J}[\rho(\tau)] = \int_{\bar t}^{\tau} \left\|\frac{\mathrm{d}}{\mathrm{d}t}\rho(s)\right\|^2 \,\mathrm{d}s = 0 \tag{11}$$

hence $\rho(t) = \rho(\bar t)$ for all $t \ge \bar t$.

We now consider the case when $t < \bar t$ so that $\mathcal{J}[\rho(t)] > 0$. By the chain rule, we have

$$-\frac{\mathrm{d}}{\mathrm{d}t}\left[\mathcal{J}[\rho(t)]^\theta\right] = -\theta\mathcal{J}[\rho(t)]^{\theta-1}\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{J}[\rho(t)] \text{ a.e. on } (0,\bar t).$$

As ρ satisfies the dynamics (10), we have by (8)

$$-\frac{\mathrm{d}}{\mathrm{d}t}\left[\mathcal{J}[\rho(t)]^\theta\right] = \theta\mathcal{J}[\rho(t)]^{\theta-1}\left\|\frac{\mathrm{d}}{\mathrm{d}t}\rho(t)\right\|^2 \text{ a.e. on } (0,\bar t).$$

Using Łojasiewicz gradient property (2), we obtain

$$-\frac{\mathrm{d}}{\mathrm{d}t}\left[\mathcal{J}[\rho(t)]^\theta\right] \ge c_g\theta\left\|\frac{\mathrm{d}}{\mathrm{d}t}\rho(t)\right\|^2 \|\nu(t)\|^{-1}\,,$$

for any $\nu(t) \in \partial\mathcal{J}[\rho(t)]$. Using the gradient dynamics one may take $\nu(t) = \mathrm{d}\rho/\mathrm{d}t$ for almost all t, to obtain

$$-\frac{\mathrm{d}}{\mathrm{d}t}\left[\mathcal{J}[\rho(t)]^{\theta}\right] \geq c_g\theta\left\|\frac{\mathrm{d}}{\mathrm{d}t}\rho(t)\right\| . \tag{12}$$

Integrating between 0 and $t \in [0, \bar{t})$ and using the absolute continuity of $\mathcal{J}[\rho(\cdot)]$, we obtain

$$0 \leq \int_0^t \left\|\frac{\mathrm{d}}{\mathrm{d}t}\rho(s)\right\| \,\mathrm{d}s \leq \frac{1}{c_g\theta}\left[\mathcal{J}^{\theta}[\rho_0] - \mathcal{J}^{\theta}[\rho(t)]\right] .$$

As a consequence of (11) this yields

$$\int_0^{\infty} \left\|\frac{\mathrm{d}}{\mathrm{d}t}\rho(t)\right\| \,\mathrm{d}s \leq \frac{1}{c_g\theta}\mathcal{J}^{\theta}[\rho_0], \tag{13}$$

which implies in particular that $\|\,\mathrm{d}\rho(t)/\,\mathrm{d}t\|$ is bounded in $L^1(0,\infty)$.

Claim 1. *If an absolutely continuous curve $\mathbb{R}_+ \ni t \to \mu(t) \in \mathcal{P}_2(\mathbb{R}^d)$ satisfies*

$$\int_0^{\infty} \left\|\frac{\mathrm{d}}{\mathrm{d}t}\mu(t)\right\| \,\mathrm{d}t < \infty$$

then μ converges to some $\bar{\mu}$ as $t \to \infty$ in the sense of the Monge-Kantorovich metric.

Proof of Claim. Simply observe that the absolute continuity implies that

$$\mathcal{W}_2(\mu_t, \mu_s) \quad \leq \quad \int_t^s \left\|\frac{\mathrm{d}}{\mathrm{d}t}\mu(\tau)\right\| \,\mathrm{d}\tau \tag{14}$$

and thus μ is a Cauchy curve for the Monge-Kantorovich distance. Since $\mathbb{R}^d$ is complete so is $\mathcal{P}_2(\mathbb{R}^d)$ (see [1, Proposition 7.1.5, p.154]) and $t \mapsto \mu(t)$ converges to some $\bar{\mu}$ as t goes to infinity.□

If we did not have $\lim_{t\to+\infty} \mathcal{J}[\rho(t)] = 0$, property (2) would imply that the subgradients along $\rho(\cdot)$ are bounded away from zero and thus there would exist a positive constant $c > 0$ such that $\|\,\mathrm{d}\rho/\,\mathrm{d}t\| > c$. This would contradict the integrability property (13).

Thus $\lim \mathcal{J}[\rho(t)] = \min \mathcal{J} = 0$. Since $\mathcal{J}$ is lower semi-continuous the limit $\bar{\rho}$ of ρ satisfies $\bar{\rho} \in \operatorname{Argmin} \mathcal{J}$.

Combining (13) and (14), we obtain

$$\mathcal{W}_2(\rho_0, \operatorname{Argmin} \mathcal{J}) \leq \mathcal{W}_2(\rho_0, \bar{\rho}) \leq \int_0^{\infty} \left\|\frac{\mathrm{d}}{\mathrm{d}t}\rho_t\right\| \,\mathrm{d}t \leq \frac{1}{c_g\theta}\mathcal{J}^{\theta}[\rho_0] \tag{15}$$

which was the stated result with $c_f = (\theta c_g)^{\frac{1}{\theta}}$. This concludes the proof of Theorem 1. □

Let us now proceed with the study of subgradient curves.

Proof of Theorem 2 In view of the estimation of convergence rates, observe that inequality (15) implies

$$\mathcal{W}_2(\rho(t), \bar{\rho}) \leq \int_t^{\infty} \left\|\frac{\mathrm{d}}{\mathrm{d}t}\rho_t\right\| \,\mathrm{d}t \leq \frac{1}{c_g\theta}\mathcal{J}^{\theta}[\rho(t)] \tag{16}$$

and $\bar{\rho} = \rho_\infty$. From the above results $\lim_{t\to\infty} \mathcal{J}[\rho(t)] = \hat{\mathcal{J}} = 0$ and ρ converges to a minimiser $\rho_\infty := \bar{\rho}$ of $\mathcal{J}$. Using (12):

$$-\frac{\mathrm{d}}{\mathrm{d}t}\left[\mathcal{J}[\rho(t)]^{\theta}\right] \geq c_g\theta\left\|\frac{\mathrm{d}}{\mathrm{d}t}\rho(t)\right\|,$$

and applying once more (2) we obtain

$$-\frac{\mathrm{d}}{\mathrm{d}t}\left[\mathcal{J}[\rho(t)]^{\theta}\right] \geq c_g^2\theta\,\mathcal{J}[\rho(t)]^{1-\theta}. \tag{17}$$

Setting $z(t) = \mathcal{J}[\rho(t)]^{\theta}$ for $t \in (0, \bar{t})$, this gives the following differential inequality

$$-\dot{z}(t) \geq c_g^2\theta z(t)^{1/\theta-1} \text{ on } [0, \bar{t}).$$

Integrating the above inequality we obtain the desired estimates for $\mathcal{J}[\rho(t)]$. Those for the Monge-Kantorovich distance follow from (16).

3. Applications to functional inequalities

Let $V : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ be a lower semi-continuous convex potential such that

$$\mathcal{U} = \operatorname{int} \operatorname{dom} V \neq \emptyset,$$
$$\int_{\mathbb{R}^d} \exp(-V) = 1.$$

This defines a log-concave probability measure $\rho^* = \exp(-V)\ dx$. Consider a lower semi-continuous convex function $f : [0, +\infty) \to [0, +\infty)$ such that the map $s \in (0, +\infty) \mapsto f(s^{-d})s^d$ is convex and non-increasing. Standard examples are $f(s) = s \log s$ or $f(s) = s^m/(m-1)$ with $m \geq 1 - 1/d$. The relative internal energy is defined as

$$\mathcal{J}[\rho] = \begin{cases} \displaystyle\int_{\mathbb{R}^d} f(\rho/\rho^*) d\rho^* \text{ if } \rho \text{ is absolutely continuous w.r.t } d\rho^* \\ +\infty \text{ otherwise.} \end{cases}$$

It is known that $\mathcal{J}$ is lower semi-continuous and displacement convex in $\mathcal{P}_2(\mathbb{R}^d)$ [1, Theorem 9.4.12, p.224]. Let us introduce $P_f(s) = sf'(s) - f(s)$ for $s \geq 0$. Denote by $L^2_\rho(\mathbb{R}^d)$ the space of square ρ-integrable functions in $\mathbb{R}^d$. By [1, Theorem 10.4.9, p.265]

$$\operatorname{dom} \partial \mathcal{J} = \left\{ \rho \in \mathcal{P}_2(\mathbb{R}^d) : P_f[\rho] \in W^{1,1}_{\text{loc}}(\mathcal{U}), \frac{\rho^*}{\rho} \nabla \left(P_f(\rho/\rho^*) \right) \in L^2_\rho(\mathbb{R}^d) \right\}$$

and

$$\|\partial^0 \mathcal{J}[\rho]\|^2 = \int_{\mathbb{R}^d} \left| \frac{\rho^*}{\rho} \nabla \left(P_f(\rho/\rho^*) \right) \right|^2 d\rho. \tag{18}$$

In this context Łojasiewicz gradient inequality (2) would write

$$\sqrt{\int_{\mathbb{R}^d} \left| \frac{\rho^*}{\rho} \nabla \left(P_f(\rho/\rho^*) \right) \right|^2 d\rho} \geq c_g \left(\int_{\mathbb{R}^d} f(\rho/\rho^*) d\rho^* \right)^{1-\theta}, \forall \rho \in \mathcal{P}_2(\mathbb{R}^d), \tag{19}$$

for some $c_g > 0$ and $\theta \in (0, 1]$. Theorem 1 asserts that this inequality is equivalent to the functional Łojasiewicz inequality (3):

$$\mathcal{W}_2(\rho, \operatorname{Argmin} \mathcal{J})^{1/(1-\theta)} \leq c_f \int_{\mathbb{R}^d} f(\rho/\rho^*) d\rho^*, \forall \rho \in \mathcal{P}_2(\mathbb{R}^d), \tag{20}$$

for some $c_f > 0$.

Whether such inequalities are satisfied for a general f is not clear. However in the case of Boltzmann and non-linear entropies *i.e.*, $f(s) = s \log s$ or $f(s) = s^m/(m-1)$ with $m \geq 1 - 1/d$, much more can be said as shown in the next sections.

3.1. **The logarithmic Sobolev inequality is equivalent to Talagrand type's inequality.**

Consider the case when $f(s) = s \log s$, so that for all ρ in the domain of $\mathcal{J}$,

$$\mathcal{J}[\rho] = \int_{\mathbb{R}^d} \rho \log \rho + \int_{\mathbb{R}^d} \rho V .$$

We have

$$\operatorname{dom} \partial \mathcal{J} = \left\{ \rho \in W^{1,1}_{\text{loc}}(\mathbb{R}^d),\ \nabla \log \left(\frac{\rho}{\rho^*} \right) \in L^2_\rho(\mathbb{R}^d) \right\}.$$

In this case Łojasiewicz gradient inequality is actually the following logarithmic Sobolev inequality, see e.g., [32, Formula (9.27), p.279]:

$$\int_{\mathbb{R}^d} \left| \nabla \left[\log \left(\frac{\rho}{\rho^*} \right) \right] \right|^2 \mathrm{d}\rho \geq c_g \left| \int_{\mathbb{R}^d} \log \left(\frac{\rho}{\rho^*} \right) \mathrm{d}\rho \right|, \forall \rho \in \operatorname{dom} \partial \mathcal{J},$$

with $c_g > 0$ (and $\theta = 1/2$, $r_0 = +\infty$).

On the other hand the functional Łojasiewicz inequality (3) is exactly a Talagrand type inequality:

$$c_f \mathcal{W}_2(\rho,\rho^*)^2 \le \int_{\mathbb{R}^d} \log\left(\frac{\rho}{\rho^*}\right)\, \mathrm{d}\rho^*, \forall \rho \in L^1(\mathbb{R}^d) \text{ with } c_f > 0.$$

Therefore Theorem 1 ensures that, up to a multiplicative constant, Talagrand and logarithmic Sobolev inequalities are equivalent. This result was already known, see [27]. The corresponding gradient system is the classical linear Fokker-Planck equation describing the evolution of the density within Ornstein–Uhlenbeck process. Exponential stabilization rates are of course recovered through Theorem 2.

3.2. **The Gagliardo-Nirenberg inequality is equivalent to a non-linear Talagrand type inequality.** Let $d \ge 1$ and consider now that $f(s) = s^m/(m-1)$, with $m \ge 1 - 1/d$. The unique minimiser of $\mathcal{J}$ is a Barenblatt profile, see [31]

$$\rho^*(x) = \left(\sigma - \frac{m-1}{2m}|x|^2\right)_+^{\frac{1}{m-1}}, \forall x \in \mathbb{R}^d, \tag{21}$$

where $\sigma > 0$ is such that $\int_{\mathbb{R}^d} \rho^* = 1$. Besides, we have

$$\operatorname{dom} \partial\mathcal{J} = \left\{ \rho \in W^{1,1}_{\text{loc}}(\mathbb{R}^d) : \nabla\left(\frac{\rho}{\rho^*}\right)^{m-1} \in L^2_\rho(\mathbb{R}^d) \right\}.$$

and

$$\|\partial\mathcal{J}^0[\rho]\| = \frac{m^2}{(m-1)^2}\int_{\mathbb{R}^d} \left|\nabla\left(\frac{\rho}{\rho^*}\right)^{m-1}\right|^2 d\rho.$$

Inequality (2) thus writes

$$\frac{m}{(m-1)}\sqrt{\int_{\mathbb{R}^d} \left|\nabla\left(\frac{\rho}{\rho^*}\right)^{m-1}\right|^2 d\rho} \ge c_g \left(\int_{\mathbb{R}^d} \frac{1}{m-1}\left[\left(\frac{\rho}{\rho^*}\right)^m - 1\right] d\rho^*\right)^{1-\theta}. \tag{22}$$

while (3) is

$$\int_{\mathbb{R}^d} \frac{1}{m-1}\left[\left(\frac{\rho}{\rho^*}\right)^m - 1\right] d\rho^* \ge c_f \mathcal{W}_2(\rho,\rho^*)^{1/\theta}.$$

By [16] inequality (22) is an instance of Gagliardo-Nirenberg inequality and holds true for $c_g = 2$ and $\theta = 1/2$. Therefore we have the following seemingly Talagrand type inequality:

For all $\rho \in W^{1,1}_{\text{loc}}(\mathbb{R}^d)$ such that

$$\int_{\mathbb{R}^d} \left|\nabla\left(\frac{\rho}{\rho^*}\right)^{m-1}\right|^2 \mathrm{d}\rho < \infty$$

we have

$$\frac{1}{m-1}\int_{\mathbb{R}^d} \left[\left(\frac{\rho}{\rho^*}\right)^m - 1\right] d\rho^* \ge 2\mathcal{W}_2(\rho,\rho^*)^2\,, \tag{23}$$

where ρ^* is the Barenblatt profile of (21).

Remark 4. Similar results were obtained by very different means in [20] and also in [4] for the $\mathcal{W}_1$ distance.

Observe finally that the application of Theorem 2 in this framework allows to recover convergence rate results of nonlinear Fokker-Planck dynamics

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho = \Delta\rho^m + \operatorname{div}\rho \mathrm{V},\ \rho(0) \in \operatorname{dom}\mathcal{J},$$

see e.g., [16, 13, 11].

4. Appendix: Notations and fundamental results

Let us remind here a few elements of formal geometry of the probability measures with the Monge-Kantorovich distance. General monographs on the subject are [1, 32, 28].

4.1. **Monge and Kantorovich's problems.** Let X and Y be two metric spaces equipped respectively with the Borel probability measures $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. For $\mu \in \mathcal{P}(X)$ and a Borel map $T : X \to Y$, $T_\# \mu$ denotes the *push forward* of μ on ν through T which is defined by $T_\#\mu(B) = \mu(T^{-1}(B))$ for every Borel subset B of Y or equivalently by the change of variables formula

$$\int_Y \varphi \ \mathrm{d}T_\#\mu = \int_X \varphi(T(x))\,\mathrm{d}\mu(x),\ \forall \varphi \in \mathcal{C}_b(X). \tag{24}$$

A transport map between μ and ν is a Borel map such that $T_\#\mu = \nu$. For $(\mu,\nu) \in \mathcal{P}(X)\times\mathcal{P}(Y)$ the Monge optimal transport problem writes

$$\mathcal{W}_2(\mu,\nu) = \inf \left\{ \int_X |x - T(x)|^2 \,\mathrm{d}\mu(x) : T_\#\mu = \nu \right\}^{\frac{1}{2}}. \tag{25}$$

When $X = Y$, $\mathcal{W}_2$ defines a distance on the subset of probabilities on X with finite second-order moments.

From now on, we assume $X = Y = \mathbb{R}^d$ where d is a positive integer. Set

$$\mathcal{P}_2(\mathbb{R}^d) = \left\{ \mu \in \mathcal{P}(\mathbb{R}^d) : \int_{\mathbb{R}^d} |x|^2 d\mu < +\infty \right\} \tag{26}$$

The solution to (25) is called *an optimal transport*. Such a transport generally exists thanks to:

Theorem 3 (Brenier's theorem [9, 32])**.** *Consider $(\mu,\nu) \in \mathcal{P}_2(\mathbb{R}^d)^2$ and assume that μ is regular in the sense that each hyper-surface has a null measure. Then the Monge optimal transport problem* (25) *has a unique solution T called the Brenier map. Moreover $T = \nabla u$ μ-a.e. for some convex function $u : \mathbb{R}^d \to \mathbb{R}$ and ∇u is the unique (up to μ-a.e. equivalence) gradient of a convex function transporting μ onto ν.*

4.2. **Convexity and geodesics.**

Definition 1 (McCann's interpolant and geodesics).

(i) Consider $(\nu_0,\nu_1) \in \mathcal{P}_2(\mathbb{R}^d)^2$ and assume that there is a Borel map $T : \mathbb{R}^d \to \mathbb{R}^d$ such that $T_\#\nu_0 = \nu_1$. The *McCann's interpolant between ν_0 and ν_1*, see [25], is the curve of measures

$$t \in [0,1] \mapsto \nu_t := ((1-t)\mathrm{id} + tT)_\#\nu_0.$$

(ii) A curve $[0,1] \ni t \to \nu_t$ in $\mathcal{P}_2(\mathbb{R}^d)$ satisfying

$$\mathcal{W}_2(\nu_t,\nu_s) = |t-s|\mathcal{W}_2(\nu_0,\nu_1)$$

is called a *(constant speed) geodesic*.

We verify easily that a McCann's interpolant is a geodesic.

Definition 2 (Displacement convexity). The functional $\mathcal{J}$: $\mathcal{P}_2(\mathbb{R}^d) \to (-\infty,+\infty]$ is called *displacement convex* if for every pair $(\nu_0,\nu_1) \in \operatorname{dom}\mathcal{J} \times \operatorname{dom}\mathcal{J}$ and McCann interpolant $\{\nu_t\}_{t\in[0,1]}$ between ν_0 and ν_1, one has

$$\mathcal{J}[\nu_t] \le (1-t)\mathcal{J}[\nu_0] + t\mathcal{J}[\nu_1],$$

for every $t \in [0,1]$.

By [32, Proposition 5.29, p.161]:

Proposition 1 (Convex inequality)**.** *Let $\mathcal{J} : \mathcal{P}_2(\mathbb{R}^d) \to (-\infty,+\infty]$ be a displacement convex function, and $\nu_0 \in \operatorname{dom}\partial J$, $\nu_1 \in \operatorname{dom}\mathcal{J}$ be two probability measures. Then*

$$\mathcal{J}[\nu_1] \ge \mathcal{J}[\nu_0] + \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0}^{+}\mathcal{J}[\nu_t], \tag{27}$$

where $\{\nu_t\}_{t\in[0,1]}$ is the McCann's interpolant between ν_0 and ν_1.

4.3. **Riemannian aspects,** [26, 1]**.** Let ρ be in $\mathcal{P}_2(\mathbb{R}^d)$, the tangent space to $\mathcal{P}_2(\mathbb{R}^d)$ at ρ, written $T_\rho\mathcal{P}_2(\mathbb{R}^d)$, is identified to the subspace of distributions formed by the vectors $s = -\nabla\cdot(\rho\nabla u)$ where u ranges over $C^\infty(\mathbb{R}^d, \mathbb{R})$. The scalar product of two vectors $s_1 = -\nabla\cdot(\rho\nabla u_1)$, $s_2 = -\nabla\cdot(\rho\nabla u_2)$, is given by

$$\langle s_1, s_2\rangle_\rho = \int_{\mathbb{R}^d} \nabla u_1 \cdot \nabla u_2 \,\mathrm{d}\rho.$$

The associated norm is as usual $\|s\| = \|s\|_\rho := \langle s, s\rangle_\rho$.

Let $\mathcal{J} : \mathcal{P}_2(\mathbb{R}^d) \to (-\infty, +\infty]$ be a displacement convex function. Define the metric (or strong) slope of $\mathcal{J}$ at $\rho \in \operatorname{dom}\mathcal{J}$ by

$$|\nabla|\mathcal{J}[\rho] = \limsup_{\mu\to\rho} \frac{(\mathcal{J}[\rho] - \mathcal{J}[\mu])^+}{\mathcal{W}_2(\rho,\mu)} \in (-\infty, +\infty].$$

For the subdifferential of $\mathcal{J}$, we pertain to [1, Definition 10.1.1, p.229] which also admits the equivalent formulation:

Definition 3. ([1, Property B, p.231]) Take $\mu \in \operatorname{dom}|\nabla|\mathcal{J}, \nu \in \operatorname{dom}\mathcal{J}$, $\nu \in L^2_\mu(\mathbb{R}^d)$, and denote by T the optimal transport from μ to ν. Then

$$\nu \in \partial\mathcal{J}[u] \Leftrightarrow \mathcal{J}[\nu] \geq \mathcal{J}[\mu] + \int_{\mathbb{R}^d} \nu(x)\cdot(T(x) - x)\,\mathrm{d}\mu(x).$$

The set $\partial\mathcal{J}[\mu]$ is obviously closed and convex in $L^2_\mu(\mathbb{R}^d)$. When it is nonempty, one defines *the minimal norm subgradient*

$$\partial^0\mathcal{J}[\mu] = \operatorname{Argmin}\{\|\nu\|_{L^2_\mu(\mathbb{R}^d)} : \nu \in \partial\mathcal{J}[\mu]\}, \tag{28}$$

see [1, Lemma 10.1.5, p.233]. In the same lemma it is shown that $\operatorname{dom}\partial\mathcal{J} = \operatorname{dom}|\nabla|\mathcal{J}$ and

$$|\nabla|\mathcal{J}[\mu] = \|\partial^0\mathcal{J}[\mu]\|, \quad \forall\mu \in \mathcal{P}_2(\mathbb{R}^d). \tag{29}$$

Acknowledgements The second author thanks the Air Force Office of Scientific Research, Air Force Material Command, USAF, under grant number FA9550-14 -1-0056 & FA9550-14-1-0500 and the FMJH Program Gaspard Monge in optimization for supporting his research.

References

[1] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, Lectures in Mathematics, Birkhäuser, 2005.

[2] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, and G. Scheffer, *Sur les inégalités de Sobolev logarithmiques*, (2000).

[3] D. Bakry and M. Émery, *Diffusions hypercontractives*, in Séminaire de Probabilités XIX 1983/84, Springer, 1985, pp. 177–206.

[4] S. G. Bobkov, Y. Ding, et al., *Optimal transport and rényi informational divergence*, Preprint, (2014).

[5] S. G. Bobkov, I. Gentil, and M. Ledoux, *Hypercontractivity of Hamilton–Jacobi equations*, Journal de Mathématiques Pures et Appliquées, 80 (2001), pp. 669–696.

[6] J. Bolte, A. Daniilidis, A. Lewis, and M. Shiota, *Clarke subgradients of stratifiable functions*, SIAM Journal on Optimization, 18 (2007), pp. 556–572.

[7] J. Bolte, A. Daniilidis, O. Ley, and L. Mazet, *Characterizations of łojasiewicz inequalities: subgradient flows, talweg, convexity*, Transactions of the American Mathematical Society, 362 (2010), pp. 3319–3363.

[8] J. Bolte, T. P. Nguyen, J. Peypouquet, and B. Suter, *From error bounds to the complexity of first-order descent methods for convex functions*, arXiv preprint arXiv:1510.08234, to appear in Math. Prog. Ser. A, (2015).

[9] Y. Brenier, *Polar factorization and monotone rearrangement of vector-valued functions*, Communications on pure and applied mathematics, 44 (1991), pp. 375–417.

[10] L. A. Caffarelli, *Boundary regularity of maps with convex potentials*, Comm. Pure Appl. Math., 45 (1992), pp. 1141–1151.

[11] J. A. Carrillo, R. J. McCann, and C. Villani, *Contractions in the 2-wasserstein length space and thermalization of granular media*, Archive for Rational Mechanics and Analysis, 179 (2006), pp. 217–263.

[12] J. A. Carrillo, R. J. McCann, C. Villani, et al., *Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates*, Revista Matematica Iberoamericana, 19 (2003), pp. 971–1018.

[13] J. A. Carrillo and G. Toscani, *Asymptotic L1-decay of solutions of the porous medium equation to self-similarity*, Indiana University Mathematics Journal, 49 (2000), pp. 113–142.
[14] R. Chill, *On the Łojasiewicz–Simon gradient inequality*, Journal of Functional Analysis, 201 (2003), pp. 572–601.
[15] D. Cordero-Erausquin, W. Gangbo, and C. Houdré, *Inequalities for generalized entropy and optimal transportation*, Contemporary Mathematics, 353 (2004), pp. 73–94.
[16] M. Del Pino and J. Dolbeault, *Best constants for Gagliardo–Nirenberg inequalities and applications to nonlinear diffusions*, Journal de Mathématiques Pures et Appliquées, 81 (2002), pp. 847–875.
[17] H. Djellout, A. Guillin, L. Wu, et al., *Transportation cost-information inequalities and applications to random dynamical systems and diffusions*, The Annals of Probability, 32 (2004), pp. 2702–2732.
[18] W. Gangbo and R. J. McCann, *The geometry of optimal transportation*, Acta Mathematica, 177 (1996), pp. 113–161.
[19] A. Haraux and M. A. Jendoubi, *Convergence of bounded weak solutions of the wave equation with dissipation and analytic nonlinearity*, Calculus of Variations and Partial Differential Equations, 9 (1999), pp. 95–124.
[20] S. I. Ohta and A. Takatsu, *Displacement convexity of generalized relative entropies*, Advances in Mathematics, 228 (2011), pp. 1742 – 1787.
[21] R. Jordan, D. Kinderlehrer, and F. Otto, *The variational formulation of the Fokker–Planck equation*, SIAM journal on mathematical analysis, 29 (1998), pp. 1–17.
[22] K. Kurdyka, *On gradients of functions definable in o-minimal structures*, in Annales de l'institut Fourier, vol. 48, 1998, pp. 769–783.
[23] M. Ledoux, *Inégalités isopérimétriques en analyse et probabilités*, Séminaire Bourbaki, 35 (1993), pp. 343–375.
[24] ———, *The concentration of measure phenomenon*, no. 89, American Mathematical Soc., 2005.
[25] R. J. McCann, *A convexity principle for interacting cases*, Adv. Math., 128 (1997), pp. 153–179.
[26] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations, 26 (2001), pp. 101–174.
[27] F. Otto and C. Villani, *Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality*, Journal of Functional Analysis, 173 (2000), pp. 361 – 400.
[28] F. Santambrogio, *Optimal transport for applied mathematicians*, Progress in Nonlinear Differential Equations and their applications, 87 (2015).
[29] L. Simon, *Asymptotics for a class of non-linear evolution equations, with applications to geometric problems*, Annals of Mathematics, (1983), pp. 525–571.
[30] S. Stanislaw Łojasiewicz, *Ensembles semi-analytiques*, Inst. Hautes Études Sci., Bures-sur-Yvette, (1965).
[31] J. L. Vázquez, *The porous medium equation: mathematical theory*, Oxford University Press, 2007.
[32] C. Villani, *Topics in optimal transportation*, vol. 58 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 2003.

Toulouse School of Economics, Université Toulouse Capitole, Manufacture des Tabacs, 21 allée de Brienne, Toulouse, France.
E-mail address: adrien.blanchet@tse-fr.eu, jerome.bolte@tse-fr.eu