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. We also discuss the more general case of λ-convex functions and we provide a general convergence theorem for the corresponding gradient dynamics. Specialising our results to the Boltzmann entropy, we recover Otto-Villani's theorem asserting the equivalence between logarithmic Sobolev and Talagrand's inequalities. The choice of power-type entropies shows a new equivalence between Gagliardo-Nirenberg inequality and a nonlinear Talagrand inequality. Some nonconvex results and other types of equivalences are discussed.

1. INTRODUCTION

Łojasiewicz inequalities are known to be extremely powerful tools for studying the long-time behaviour of dissipative systems in an Euclidean or Hilbert space, see e.g., [33, 22, 15, 8] and references therein. Their connection with the asymptotics of gradient flows comes from the fact that one of this inequality asserts that the underlying energy can be rescaled near critical points into a sharp function[1]. A consequence of this inequality is that gradient curves can be shown to have finite length through the choice of an adequate Lyapunov function, see [29, 26, 8][2].

In parallel the study of large time asymptotics of various PDEs, also based on energy techniques, was developed in close conjunction with functional inequalities. A classical study protocol is to evidence Lyapounov functionals and use functional inequalities to derive quantitative contractive properties of the flow. The heat equation provides an elementary but illustrative example of this approach: the Boltzmann entropy gives a Lyapunov functional while the Logarithmic-Sobolev inequality ensures the exponential convergence of the solution curve to a self-similar profile. Numerous applications of these techniques, as well as their stochastic counterparts, can be found in e.g., [4, 12, 3, 16, 6, 18]. Standard references for functional inequalities are for instance [13, 28, 27, 21].

In this article we show that the joint use of metric gradient flows and Łojasiewicz inequalities allows for a systematic and transparent treatment of these evolution equations. In this regard the "Riemannian structure" of the set of probability measures endowed with the Monge-Kantorovich distance (see [30, 2, 19, 35, 25]) plays a fundamental role in our approach. It allows in particular to interpret some PDEs as gradient flows, like Fokker-Planck, porous medium, or fast diffusion equations, and it provides a setting sufficiently rich to formulate precisely Łojasiewicz inequalities.

In a first step we indeed introduce two types of Łojasiewicz inequalities in the probability space equipped with the Monge-Kantorovich distance. One of these inequalities is a growth measure of the energy functional with respect to the Monge-Kantorovich distance to stationary points, while the other provides a relationship between the values of the energy and its slope. The latter is called the *gradient Łojasiewicz inequality*. In this functional setting, both inequalities can be viewed as families of abstract functional inequalities. We prove their equivalence in the case of convex functionals. Specialising our results to Boltzmann's entropy we recover Otto-Villani's theorem [31] stating the equivalence between the logarithmic Sobolev and Talagrand inequalities. We also prove a new equivalence between Gagliardo-Sobolev inequality and a nonlinear Talagrand

[1]See Remark 1 for further explanations

[2]See also the proof of Theorem 1

type inequality (22). For general λ-convex functionals the gradient Łojasiewicz gradient inequality merely implies the growth around the minimisers set, the reverse implication is in general false. This difficulty can be felt through the fact that Talagrand's inequality is not known to imply logarithmic Sobolev inequality for a general non-convex[3] confinement potential, see [35, Section 9.3.1, p.292]. In Section 3 we discuss further this issue and provide a new proof that the reverse implication holds for convex potentials with L^∞ variations, [31, Corollary 2.2].

In Theorem 2, we use Łojasiewicz inequalities in a classical way to analyse the convergence and the rate of evolution equations. Our results do not subsume the uniqueness of a minimiser/critical point, neither the convexity of the potential. We recover many classical results in a unified way and we provide new insights into the stabilisation phenomena in the presence of a continuum of equilibria.

The structure of the paper is as follows. In Section 2, we state two fundamental types of Łojasiewicz inequalities and study their equivalence. General convergence rate results for subgradient systems are also provided. 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 postponed to Section 4.

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 function which is λ-convex along generalised geodesics with λ in $\mathbb{R}$.

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

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

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

where $\partial \mathcal{J}$ stands for the subdifferential of $\mathcal{J}$ in the probability space equipped with the Monge-Kantorovich metric and $\| \cdot \|_\rho$ is the norm in $L^2_\rho(\mathbb{R}^d)$, see Section 4.

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

$$\hat{\mathcal{J}} < \mathcal{J}[\rho] < r_0 \quad \Rightarrow \quad c_f \, \mathcal{W}_2(\rho, \mathrm{Argmin}\, \mathcal{J})^{1/\theta} \leq \mathcal{J}[\rho] - \hat{\mathcal{J}}. \tag{2}$$

Remark 1. (a) A formal geometric interpretation of (1) is as follows: assume $\mathcal{J}$ sufficiently smooth to rewrite the inequality (1) as

$$\|\nabla \left(\mathcal{J} - \hat{\mathcal{J}} \right)^\theta [\rho]\|_\rho \geq \theta c_g,$$

for any ρ such that $\hat{\mathcal{J}} < \mathcal{J}[\rho] < r_0$. With this reformulation, we see that the norm of the gradient of $(\mathcal{J} - \hat{\mathcal{J}})^\theta$ is bounded away from 0 for non-critical measures. The obtained reparameterization $(\mathcal{J} - \hat{\mathcal{J}})^\theta$ of $\mathcal{J}$ is called *sharp* in reference to the sharpness of its profile near its critical set. It provides a Lyapunov function for the corresponding gradient system with strong decrease rate properties reflecting the V-shape of the rescaled energy (see the proof of Theorem 1).

The second inequality (2) is a classical growth inequality around stationary points.

(b) Original inequalities for analytic and subanalytic functions can be found in the IHES lectures [29] by Łojasiewicz. Several generalisations of gradient inequalities followed, see in particular [33, 26, 7].

(c) In this infinite dimensional setting, the above Łojasiewicz inequalities should be thought as families of functional inequalities. Formal connections with the Talagrand, logarithmic Sobolev

---

[3]However it holds for potential whose Hessian is bounded from below, see [31, Corollary 3.1]

and Gagliardo-Nirenberg inequalities are provided in Section 3.

(d) When $r_0 = +\infty$ the above inequalities are *global inequalities* which will be the case of most examples provided in Section 3.

(e) Note that when $\mathcal{J}$ is not convex, Property (1) implies that there are no critical values between $\hat{\mathcal{J}}$ and r_0, while (2) does not preclude this possibility.

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 functional which has at least a global minimiser.*

    (i) *If $\mathcal{J}$ is λ-convex along generalised geodesics for some λ in $\mathbb{R}$, then the Łojasiewicz gradient property (1) implies the Łojasiewicz functional property (2).*

   (ii) *If $\mathcal{J}$ is convex along generalised geodesics, then the Łojasiewicz gradient property (1) and the Łojasiewicz functional property (2) are equivalent.*

Remark 2. (a) Note that the quantities r_0, θ are conserved in both cases. Theorem 1 provides a relationship between c_f and c_g but the optimality of the constant might be lost[4]. Indeed our proof gives $c_g = c_f^\theta$ when one establishes (2) $\Rightarrow$ (1), while $c_f = (\theta c_g)^{1/\theta}$ when (1) $\Rightarrow$ (2) is proved. Yet, when $\theta = 1$ and $\mathcal{J}$ is convex the equivalence between (1) and (2) holds with $c_f = c_g$.

(b) Assume for simplicity that $\hat{\mathcal{J}} = 0$. The extension of the gradient Łojasiewicz inequality by Kurdyka [26] would write in our setting:

$$0 < \mathcal{J}[\rho] < r_0 \quad \Rightarrow \quad \forall \nu \in \partial \mathcal{J}[\rho] \, , \, 1 \le \varphi'(\mathcal{J}[\rho]) \|\nu\|_\rho, \tag{3}$$

for some $\varphi \in C^0[0, r_0) \cap C^1(0, r_0)$ such that $\varphi(0) = 0$ and $\varphi' > 0$ on $(0, r_0)$. Note that (1) is nothing but inequality (3) with $\varphi(s) = s^\theta/(\theta c_g)$. Mimicking the proof of Theorem 1 one can establish that 3 implies:

$$0 < \mathcal{J}[\rho] < r_0 \quad \Rightarrow \quad \varphi^{-1}(\mathcal{W}_2(\rho, \mathrm{Argmin}\, \mathcal{J})) \le \mathcal{J}[\rho]. \tag{4}$$

Yet (3) and (4) are not in general equivalent. One can build a C^2 convex coercive function in $\mathbb{R}^2$ which satisfies (4) but not (3), whatever the choice of φ, see [9]. The construction is fairly complex, involves highly oscillatory behaviour of level sets and originates in [8]. This limitation helps to understand the discrepancy between constants mentioned in item (a) of this remark.

(c) An Hilbertian version of the above result was provided in [9] in view of studying the complexity of first-order methods.

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

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho(t) + \partial \mathcal{J}[\rho(t)] \ni 0, \text{ for almost every } t \text{ in } (0, +\infty), \tag{5}$$

with initial condition $\rho(0) = \rho_0$. Such a solution exists and is unique see [2, Theorem 11.3.2, p.305] and [2, Theorem 11.1.4, p.285]. Moreover by the energy identity [2, Theorem 11.3.2, p.305], one has:

$$\mathcal{J}[\rho(\cdot)] \text{ is non-increasing, absolutely continuous,} \tag{6}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{J}[\rho(t)] = -\left|\frac{\mathrm{d}\rho}{\mathrm{d}t}\right|^2 (t) \text{ a.e. on } (0, +\infty). \tag{7}$$

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[5].

---

[4]See Section 3.1 where optimality of the constant is preserved by (1) $\Rightarrow$ (2)

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

Theorem 2 (Global convergence rates and functional inequalities). *Consider $\mathcal{J}$ a proper lower semi-continuous λ-convex functional. Assume $\mathcal{J}$ has at least a minimiser and satisfies the gradient inequality (1). Let $\rho_0 \in [\hat{\mathcal{J}} \leq \mathcal{J} < r_0]$, and consider a trajectory of (5) starting at ρ_0. Then this trajectory has a finite length and converges for the Monge-Kantorovich metric to a minimiser ρ_∞ of $\mathcal{J}$.*

Moreover the following estimations hold:

(i) If $\theta \in (0, 1/2)$, then for all $t \geq 0$,

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

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

(ii) If $\theta = 1/2$, then for all $t \geq 0$,

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

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

(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 $[0, T]$

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

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

Remark 3. (a) A fundamental feature of this convergence result is that *it does not subsume the uniqueness of a minimiser* which is uncommon in the domain.

(b) 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 (5).

(c) By Theorem 1, when $\mathcal{J}$ is convex it is sufficient to assume (2) instead of (1). This fact is quite important in practice since (2) is a "zero-order" functional inequality[6] and is in general easier to derive than (1).

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

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

$$\frac{\mathcal{J}[\nu] - \mathcal{J}[\mu]}{\mathcal{W}_2(\mu, \nu)} \leq \|\partial^0 \mathcal{J}[\nu]\|_\nu. \tag{8}$$

Proof. Set $\alpha = \mathcal{W}_2(\mu, \nu)$. Let $[0, \alpha] \ni t \to \rho_t := \mu_{t/\alpha}$ where μ_t is a constant speed geodesic between μ and ν (indeed $\mathcal{P}_2(\mathbb{R}^d)$ is a geodesic space [1, Theorem 2.10, p.35]). Since $t \mapsto \mathcal{J}[\rho_t]$ is convex, we have

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

As $t \mapsto \mu_t$ is a constant speed geodesic, $\mathcal{W}_2(\rho_t, \rho_\tau) = |t/\alpha - \tau/\alpha|\mathcal{W}_2(\mu, \nu) = |t - \tau|$ for all time $(t, \tau) \in [0, \alpha]^2$, the above can be rewritten as

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

---

[6]It only involves the value of the function $\mathcal{J}$ and no higher order information

where the last inequality follows from (34). □

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

• Proof of (i). We thus prove (1) ⇒ (2) without using convexity. Take ρ_0 with $\mathcal{J}[\rho_0] \in (0, r_0)$ and consider the dynamics

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho(t) + \partial\mathcal{J}[\rho(t)] \ni 0 \quad \text{with } \rho(0) = \rho_0. \tag{9}$$

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 (6), one has $\mathcal{J}[\rho(t)] = 0$ for all $t \geq \bar{t}$. By integrating (7) over $(\bar{t}, \tau)$, with $\tau \geq \bar{t}$, one obtains

$$\mathcal{J}[\rho(\bar{t})] - \mathcal{J}[\rho(\tau)] = \int_{\bar{t}}^{\tau} \left|\frac{\mathrm{d}\rho}{\mathrm{d}s}\right|^2 (s) \, \mathrm{d}s = 0. \tag{10}$$

where the last equality comes from the "lazy selection" principle [2, Theorem 11.3.2 (ii)]. The latter indeed implies that the velocity coincides for almost all $t \in (0, +\infty)$ with the minimum norm subgradient which is 0 in this case. Whence $\rho(t) = \rho(\bar{t})$ for all $t \geq \bar{t}$.

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

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

As ρ follows the dynamics (9), we have by (7)

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

Using Łojasiewicz gradient property (1), we obtain

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

for any $\nu(t) \in \partial\mathcal{J}[\rho(t)]$. From [2, Theorem 8.3.1, p.183], we have for almost every t:

$$\left|\frac{\mathrm{d}\rho}{\mathrm{d}t}\right|(t) = \left\|\frac{\mathrm{d}}{\mathrm{d}t}\rho(t)\right\|_\rho.$$

Using the gradient dynamics one may take $\nu(t) = -\mathrm{d}\rho/\mathrm{d}t$ for almost all t, to obtain

$$-\frac{\mathrm{d}}{\mathrm{d}t}\left[\mathcal{J}[\rho(t)]^\theta\right] \geq c_g\theta\left|\frac{\mathrm{d}\rho}{\mathrm{d}t}\right|(t). \tag{11}$$

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

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

As a consequence of (10) this yields

$$\int_0^\infty \left|\frac{\mathrm{d}\rho}{\mathrm{d}s}\right|(s) \, \mathrm{d}s \leq \frac{1}{c_g\theta}\mathcal{J}^\theta[\rho_0], \tag{12}$$

which implies in particular that $|\mathrm{d}\rho/\mathrm{d}s|(s)$ 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*

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

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

Proof of Claim 1. Simply observe that the absolute continuity and the definition of the metric derivative implies that

$$\mathcal{W}_2(\mu_t, \mu_s) \quad \leq \quad \int_t^s \left| \frac{\mathrm{d}\mu}{\mathrm{d}\tau} \right| (\tau) \, \mathrm{d}\tau, \, \forall s \geq t, \tag{13}$$

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 [2, Proposition 7.1.5, p.154]) and $t \mapsto \mu(t)$ converges to some $\bar{\mu}$ as t goes to infinity.$\square$

If we did not have $\lim_{t \to +\infty} \mathcal{J}[\rho(t)] = 0$, property (1) 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 (12).

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

Combining (12) and (13), we obtain

$$\mathcal{W}_2(\rho_0, \text{Argmin } \mathcal{J}) \leq \mathcal{W}_2(\rho_0, \bar{\rho}) \leq \int_0^\infty \left| \frac{\mathrm{d}\rho}{\mathrm{d}t} \right| (t) \, \mathrm{d}t \leq \frac{1}{c_g \theta} \mathcal{J}^\theta[\rho_0] \tag{14}$$

which was the stated result with $c_f = (\theta c_g)^{1/\theta}$.

• Proof of (ii). To establish (2) $\Rightarrow$ (1), we further assume that, $\lambda = 0$, *i.e.*, $\mathcal{J}$ is convex. If $\rho \in$ 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

$$\mathcal{J}[\rho] \leq \mathcal{W}_2(\rho, \bar{\rho}) \left\| \nu \right\|_\rho,$$

and thus

$$\mathcal{J}[\rho] \leq \mathcal{W}_2(\rho, \text{Argmin } \mathcal{J}) \left\| \nu \right\|_\rho.$$

By Łojasiewicz functional property (2), we obtain

$$\mathcal{J}[\rho] \leq \left(\frac{1}{c_f} \mathcal{J}[\rho] \right)^\theta \left\| \nu \right\|_\rho,$$

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

This concludes the proof of Theorem 1. $\square$

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 (14) implies

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

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 ρ_∞ of $\mathcal{J}$. Using (11) and applying once more (1), we obtain

$$-\frac{\mathrm{d}}{\mathrm{d}t} \left\{ \mathcal{J}[\rho(t)]^\theta \right\} \geq c_g^2 \theta \, \mathcal{J}[\rho(t)]^{1-\theta}. \tag{16}$$

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 (15).

3. APPLICATIONS TO FUNCTIONAL INEQUALITIES

3.1. **Relative internal energies.** 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 $\rho^* := \exp(-V)$. Consider a lower semi-continuous convex function $f : [0, +\infty) \to [0, +\infty)$ with $f(1) = 0$ and such that the map

$$s \in (0, +\infty) \mapsto f(s^{-d})s^d \text{ is convex and non-increasing.} \tag{17}$$

Standard examples are $f(s) = s \log s$ or $f(s) = (s^m - s)/(m - 1)$ with $m \geq 1 - 1/d$. The relative internal energy is defined as

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

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

$$\operatorname{dom} \partial \mathcal{J} = \left\{ \rho \in \mathcal{P}_2(\mathbb{R}^d) : P_f(\rho) \in W_{\mathrm{loc}}^{1,1}(\mathcal{U}), \frac{\rho^*}{\rho} \nabla \left(P_f(\rho/\rho^*) \right) \in L_\rho^2(\mathbb{R}^d) \right\}$$

and

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

In this context Łojasiewicz gradient inequality (1) would write, for all $\rho \in \mathcal{P}_2(\mathbb{R}^d)$:

$$\mathcal{J}[\rho] < r_0 \Rightarrow \sqrt{\int_{\mathbb{R}^d} \left| \frac{\rho^*}{\rho} \nabla \left(P_f(\rho/\rho^*) \right) \right|^2 \mathrm{d}\rho} \geq c_g \left(\int_{\mathbb{R}^d} f(\rho/\rho^*) \, \mathrm{d}\rho^* \right)^{1-\theta},$$

for some $c_g > 0$, $r_0 \in (0, +\infty]$ and $\theta \in (0, 1]$. Theorem 1 asserts that this inequality is equivalent to the functional Łojasiewicz inequality (2), for all $\rho \in \mathcal{P}_2(\mathbb{R}^d)$:

$$\mathcal{J}[\rho] < r_0 \Rightarrow c_f \mathcal{W}_2(\rho, \operatorname{Argmin} \mathcal{J})^{1/\theta} \leq \int_{\mathbb{R}^d} f(\rho/\rho^*) \, \mathrm{d}\rho^*, \tag{19}$$

for some $c_f > 0$.

Whether such inequalities are satisfied for a general f is not clear. However in the case of the Boltzmann entropy *i.e.,* $f(s) = s \log s$ much more can be said.

The logarithmic Sobolev inequality is equivalent to Talagrand inequality. Consider the case when $f(s) = s \log s$, $V : \mathbb{R}^d \to \mathbb{R}$, a C^2-function with $\nabla^2 V \geq K I_d$ and $K > 0$. Hence for all ρ in the domain of $\mathcal{J}$,

$$\mathcal{J}[\rho] = \int_{\mathbb{R}^d} \log \left(\frac{\rho}{\rho^*} \right) \, \mathrm{d}\rho.$$

We have

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

In this case the Łojasiewicz gradient inequality takes the form of the following logarithmic Sobolev inequality, see e.g., [35, Formula (9.27), p.279]:

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

corresponding to $\theta = 1/2$, and $r_0 = +\infty$. The optimal constant c_g is given by $\sqrt{2K}$.

On the other hand the functional Łojasiewicz inequality (2) is exactly a Talagrand type inequality:

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

where the optimal constant is $c_f = K/2$.

Therefore Theorem 1 ensures that, up to a multiplicative constant, the Talagrand and logarithmic Sobolev inequalities are equivalent under a convexity assumption. This result is known and due to Otto-Villani, see [31]; it was obtained by completely different means[**?**]. Remark 2 shows further that (1), with the optimal constant $c_g = \sqrt{2K}$, implies (2) with the constant

$$c_f = \left(\frac{1}{2}\sqrt{2K}\right)^2 = \frac{K}{2},$$

which happens to be the optimal constant of the Talagrand inequality. For the reverse implication, the constant c_g is half the optimal constant.

The corresponding gradient system is the classical linear Fokker-Planck equation describing the evolution of the density within Ornstein-Uhlenbeck process:

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho = \Delta\rho + \mathrm{div}\,(\rho\nabla V), \; \rho(0) \in \mathrm{dom}\,\mathcal{J}.$$

Exponential stabilisation rates are of course recovered through Theorem 2 (ii).

3.2. **Sum of internal and potential energies.** Let $F : [0, +\infty) \to \mathbb{R}$ be a convex differentiable function with super-linear growth satisfying $F(0) = 0$, (17) and

$$\liminf_{s \to 0} s^{-\alpha} F(s) > -\infty, \text{ for some } \alpha > d/(d+2).$$

Consider $V : \mathbb{R}^d \to \mathbb{R}$ a differentiable, λ-convex function with $\lambda \in \mathbb{R}$. Set

$$\mathcal{J}[\rho] := \begin{cases} \displaystyle\int_{\mathbb{R}^d} F(\rho)\,\mathrm{d}x + \int_{\mathbb{R}^d} V\,\mathrm{d}\rho \text{ if } \rho \text{ is absolutely continuous w.r.t the Lebesgue measure,} \\ +\infty \text{ otherwise.} \end{cases}$$

The function $\mathcal{J}$ is lower semi-continuous, λ-convex in $\mathcal{P}_2(\mathbb{R}^d)$ – combine indeed [2, Theorem 9.3.9, p.212] and [2, Proposition 9.3.2, p.210]. Moreover

$$\mathrm{dom}\,\partial\mathcal{J} = \left\{\rho \in \mathcal{P}_2(\mathbb{R}^d) : P_F(\rho) \in W^{1,1}_{\mathrm{loc}}(\mathbb{R}^d), \nabla F'(\rho) + \nabla V \in L^2_\rho(\mathbb{R}^d)\right\},$$

see [2, Theorem 10.4.13 p.273].

In this context Łojasiewicz gradient inequality (1) would write, for all ρ in $\mathcal{P}_2(\mathbb{R}^d)$:

$$\hat{\mathcal{J}} < \mathcal{J}[\rho] < r_0 \Rightarrow \sqrt{\int_{\mathbb{R}^d} |\nabla F'(\rho) + \nabla V|^2 \, \mathrm{d}\rho} \geq c_g \left(\int_{\mathbb{R}^d} F(\rho)\,\mathrm{d}x + \int_{\mathbb{R}^d} V\,\mathrm{d}\rho\right)^{1-\theta},$$

for some $c_g > 0, r_0 \in (\hat{\mathcal{J}}, +\infty]$ and $\theta \in (0, 1]$. Our main Theorem 1 asserts that this inequality is equivalent to the functional Łojasiewicz inequality (2), for all $\rho \in \mathcal{P}_2(\mathbb{R}^d)$:

$$\hat{\mathcal{J}} < \mathcal{J}[\rho] < r_0 \Rightarrow c_f \mathcal{W}_2(\rho, \mathrm{Argmin}\,\mathcal{J})^{1/\theta} \leq \int_{\mathbb{R}^d} F(\rho)\,\mathrm{d}x + \int_{\mathbb{R}^d} V\,\mathrm{d}\rho,$$

for some $c_f > 0$.

General conditions for the validity of theses inequalities are, up to our knowledge, not known at this day. Yet they hold for two specific choices of functions: $F(s) = s \log s$ and $F(s) = s^m/(m-1)$ for $m \geq 1-1/d$, corresponding respectively to the Boltzmann and non-linear entropies. Łojasiewicz inequalities for the case involving the Boltzmann entropy boil down to logarithmic Sobolev and Talagrand inequalities (see Section 3.1).

The case of the "power-entropy" is developed below.

The Gagliardo-Nirenberg inequality is equivalent to a non-linear Talagrand type inequality. We assume further V is a C^2-function with $\nabla^2 V \geq K I_d$ and $K > 0$. Let $d \geq 1$ and consider $F(s) = s^m/(m-1)$, with $m \geq 1 - 1/d$. The unique minimiser of $\mathcal{J}$ is a Barenblatt profile, see [34]

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

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

$$\mathrm{dom}\, \partial \mathcal{J} = \left\{ \rho \in \mathcal{P}_2(\mathbb{R}^d) : \rho \in W^{1,m}_{\mathrm{loc}}(\mathbb{R}^d), \nabla \left(\frac{m}{m-1} \rho^{m-1} + V \right) \in L^2_\rho(\mathbb{R}^d) \right\}$$

and

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

When $r_0 = +\infty$ and $\theta = 1/2$ inequality (1) writes

$$\int_{\mathbb{R}^d} \left| \frac{m}{m-1} \nabla(\rho^{m-1}) + \nabla V \right|^2 \mathrm{d}\rho \geq c_g^2 \left(\int_{\mathbb{R}^d} \frac{\rho^m}{m-1} \mathrm{d}x + \int_{\mathbb{R}^d} V \mathrm{d}\rho \right), \tag{20}$$

while (2) is

$$\int_{\mathbb{R}^d} \frac{\rho^m}{m-1} \mathrm{d}x + \int_{\mathbb{R}^d} V \mathrm{d}\rho \geq c_f \mathcal{W}_2(\rho, \rho^*)^2. \tag{21}$$

By [17, 11] inequality (20) is an instance of Gagliardo-Nirenberg inequality and holds true for an optimal $c_g = \sqrt{2K}$. Therefore (21) holds true for $c_f = K/2$:

$$\int_{\mathbb{R}^d} \frac{\rho^m}{m-1} \mathrm{d}x + \int_{\mathbb{R}^d} V \mathrm{d}\rho \geq \frac{K}{2} \mathcal{W}_2(\rho, \rho^*)^2, \ \forall \rho \in W^{1,1}_{\mathrm{loc}}(\mathbb{R}^d). \tag{22}$$

This inequality was recently obtained by Ohta-Takatsu in [24] by completely different means. Our theorem also ensures that Ohta-Takatsu inequality implies Gagliardo-Nirenberg inequality up to a multiplicative constant.

Remark 4. Inequality (22) for the $\mathcal{W}_1$ distance is characterised in [5] through the moments of the invariant measure.

Observe finally that the application of Theorem 2 (ii) in this framework allows to recover convergence rate results of nonlinear Fokker-Planck/porous medium dynamics for $m \geq 1 - 1/d$:

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

see e.g., [17, 14, 12].

Logarithmic Sobolev and Talagrand's inequalities: the non-convex case. We provide here an important example of a non-convex functional $\mathcal{J}$. Consider $F(s) = s \log s$, and assume that $V : \mathbb{R}^d \to \mathbb{R}$ is of the form

$$V = V_1 + V_2,$$

where V_1 is C^2 with $\nabla^2 V_1 \geq K I_d$, $K > 0$, and V_2 is in $L^\infty(\mathbb{R}^n)$. Set $\mathrm{osc}\,(V_2) := \sup V_2 - \inf V_2 < +\infty$ and $\rho^* := \exp(-V)$. We assume that ρ^* is in $\mathcal{P}_2(\mathbb{R}^d)$. Then one has a logarithmic Sobolev inequality, see Holley-Stroock's article [23],

$$\int_{\mathbb{R}^d} |\nabla \log \rho + \nabla V|^2 \mathrm{d}\rho \geq c_g^2 \left(\int_{\mathbb{R}^d} \rho \log \rho \mathrm{d}x + \int_{\mathbb{R}^d} V \mathrm{d}\rho \right), \tag{23}$$

where $c_g^2 = 2K \exp(-\mathrm{osc}\,(V_2))$. Theorem 1 (i) shows that (23) implies

$$c_f \mathcal{W}_2(\rho, \rho^*)^2 \leq \int_{\mathbb{R}^d} \rho \log \rho \mathrm{d}x + \int_{\mathbb{R}^d} V \mathrm{d}\rho,$$

with $c_f = K \exp(-\mathrm{osc}\,(V_2))/2$. This implication was already proved in [31] for non-convex V, see also [20]. A famous family of energies for which (23) holds, includes the double-well potential: $x \mapsto ax^4 - bx^2$ where $a > 0$ (see [23]).

The application of Theorem 2 to this setting allows to recover classical convergence results for Fokker-Planck dynamics in the absence of convexity.

3.3. **Potential energies with λ-convex subanalytic functions.** Let $V : \mathbb{R}^d \to \mathbb{R}$ be a λ-convex function, $\lambda \in \mathbb{R}$. Assume that there exist a, b in $\mathbb{R}$ with $V(x) \geq -a|x|^2 + b$. The functional, called *potential energy* is given by

$$\mathcal{J}[\rho] := \int_{\mathbb{R}^d} V \, \mathrm{d}\rho.$$

$\mathcal{J}$ is well defined, lower semi-continuous, and λ-convex, see [2, Proposition 9.3.2 p.210].

Lifted convex functions. We first consider the case when V is convex, *i.e.*, $\lambda = 0$. We obviously have $\operatorname{Argmin} \mathcal{J} = \{\rho \in \mathcal{P}_2(\mathbb{R}^d) : \operatorname{supp} \rho \subset \operatorname{Argmin} V\}$ and $\hat{\mathcal{J}} = \hat{V}$. If $\mathcal{S}$ is a subset of $\mathbb{R}^d$, we set $\operatorname{dist}(x, \mathcal{S}) = \inf\{|x - y| : y \in \mathcal{S}\}$.

In that case Łojasiewicz inequalities can be lifted in the Monge-Kantorovich space. As a consequence, we will derive the following result:

Theorem 3 (Łojasiewicz inequality for a convex potential energy). *Let $V : \mathbb{R}^d \to \mathbb{R}$ be a convex function such that $\hat{V} := \inf V > -\infty$. If V is subanalytic with compact level sets then there exist $c_f > 0$, $c_g > 0$, and $\theta \in (0, 1]$ such that $\mathcal{J}$ satisfies the Łojasiewicz inequalities (1) and (2) with $r_0 = +\infty$.*

The proof of the above theorem ensues from the fact that V satisfies itself the Łojasiewicz inequality with $r_0 = +\infty$ (see e.g. [9]) and from the following equivalence result:

Theorem 4 (Łojasiewicz inequality in $\mathcal{P}_2(\mathbb{R}^d)$ is equivalent to Łojasiewicz inequality in $\mathbb{R}^d$). *We make the assumptions of Theorem 3. Fix $r_0 \in (\hat{V}, +\infty]$. The two following assertions are equivalent:*
(i) There exist $c > 0$, $\theta \in (0, 1]$ such that for each x in $\mathbb{R}^d$,

$$\hat{V} < V(x) < r_0 \quad \Rightarrow \quad c \operatorname{dist}(x, \operatorname{Argmin} V)^{1/\theta} \leq V(x) - \hat{V}.$$

(ii) There exist $c' > 0$, $\theta' \in (0, 1]$ such that for each ρ in $\mathcal{P}_2(\mathbb{R}^d)$,

$$\operatorname{Supp} \rho \subset V^{-1}(\hat{V}, r_0) \quad \Rightarrow \quad c' \, \mathcal{W}_2(\rho, \operatorname{Argmin} \mathcal{J})^{1/\theta'} \leq \mathcal{J}[\rho] - \hat{\mathcal{J}}. \tag{24}$$

Proof. $(ii) \Rightarrow (i)$. Take $x \in \mathbb{R}^d$ and write the inequality (24) for δ_x, the Dirac at x. The result follows from $\hat{\mathcal{J}} = \hat{V}$ and

$$\operatorname{dist}(x, \operatorname{Argmin} V) = \mathcal{W}_2(\delta_x, \operatorname{Argmin} \mathcal{J}).$$

Observe that this implication holds regardless of the value of θ.
$(i) \Rightarrow (ii)$. Assume for simplicity that $\hat{V} = 0$ and thus $\hat{\mathcal{J}} = 0$. Using the Łojasiewicz inequality for V, we have

$$V(x) \geq c \operatorname{dist}(x, \mathcal{S})^{1/\theta} \text{ for all } x \text{ such that } V(x) < r_0,$$

where $\mathcal{S} = \operatorname{Argmin} V$. As a consequence

$$c^{-1} \mathcal{J}[\rho] \geq \int_{\mathbb{R}^d} \operatorname{dist}(x, \mathcal{S})^{1/\theta} \, \mathrm{d}\rho = \int_{\mathbb{R}^d} |x - \operatorname{proj}_{\mathcal{S}}(x)|^{1/\theta} \, \mathrm{d}\rho$$

If $\theta \leq 1/2$, using Hölder's inequality, we obtain

$$c^{-1} \mathcal{J}[\rho] \geq \left(\int_{\mathbb{R}^d} |x - \operatorname{proj}_{\mathcal{S}}(x)|^2 \, \mathrm{d}\rho \right)^{1/(2\theta)}. \tag{25}$$

For $\theta > 1/2$, we need the following Gagliardo-Nirenberg inequality:

$$\|u\|_p \leq C \|u\|_q^{1-a} \|u\|_{W^{1,r}}^a \quad \text{with} \quad a := \frac{1/q - 1/p}{1/q + 1/d - 1/r},$$

$1 \leq q \leq p \leq \infty$, $r > d$ and where $C = C(p, q, d)$. Applying the above to $u = \operatorname{id} - \operatorname{proj}_{\mathcal{S}}$ with $p = 2$ and $q = 1/\theta$ entails

$$c^{-1} \mathcal{J}[\rho] \geq \gamma \left(\int_{\mathbb{R}^d} |x - \operatorname{proj}_{\mathcal{S}}(x)|^2 \, \mathrm{d}\rho \right)^{1/(2\theta(1-a))} \tag{26}$$

where $\gamma > 0$ is an adequate constant. For any $\theta \in (0,1]$, we can assemble (25) and (26) into

$$(c')^{-1} \mathcal{J}[\rho] \geq \left(\int_{\mathbb{R}^d} |x - \text{proj}_{\mathcal{S}}(x)|^2 \, \mathrm{d}\rho \right)^{\alpha} \tag{27}$$

where $\alpha := 1/2\theta$ and $c = c'$ if $\theta \leq 1/2$, while $\alpha := 1/(2\theta(1-a))$ and $c' = \gamma c$ otherwise. We finally obtain

$$
\begin{aligned}
(c')^{-1} \mathcal{J}[\rho] &\geq \inf \left\{ \int_{\mathbb{R}^d} |x - T(x)|^2 \, \mathrm{d}\rho : T \text{ Borel map on } \mathbb{R}^d \text{ with } T(\text{Supp}\,\rho) \subset \mathcal{S} \right\}^{\alpha} \\
&= \mathcal{W}_2(\rho, \text{Argmin}\,\mathcal{J})^{2\alpha}
\end{aligned}
$$

When $\theta \in (0, 1/2]$, one obtains the result with $\theta' = \theta$ and $c' = c$. Otherwise, for $\theta > 1/2$, we choose

$$\theta' = \theta \frac{1/2 + 1/d}{\theta + 1/d} \in (0,1), \tag{28}$$

which concludes the proof. $\qquad\square$

Remark 5. (a) With the same assumptions and for $\theta \in (0, 1/2]$, the equivalence between (i) and (ii) holds with $\theta = \theta'$ and $c = c'$.
(b) In the case when $\theta \in (1/2, 1]$, the curvature of the Monge-Kantorovich space deteriorates the lifted exponent θ' by flattening the profile $\mathcal{J}$. To see this, consider for $d = 1$, the sharp function $V(x) = |x|$ whose exponent is $\theta = 1$ and the path $(\rho_t)_{t \in [0,1]} = (1-t)\delta_0 + t\delta_1$ in $\mathcal{P}_2(\mathbb{R}^d)$, then $\mathcal{J}[\rho_t] = t$ while $\mathcal{W}_2(\rho_t, \text{Argmin}\,\mathcal{J}) = \mathcal{W}_2(\rho, \delta_0) = \sqrt{t}$ for all t in $[0,1]$. Hence the optimal exponent of $\mathcal{J}$ is lower than $1/2$ and thus $\mathcal{J}$ does not inherit of the sharpness of its kernel V.

Observe that, in the general case, the estimation (28) predicts indeed a lifted exponent with a value lower than θ:

$$\theta' = \theta \frac{1/2 + 1/d}{\theta + 1/d} < \theta.$$

Non-convex kernels. When V is non-convex but merely λ-convex for λ in $\mathbb{R}$, we observe a mixing of critical values of the energy. It is easily seen by considering two arbitrary critical points x, y of V and by considering the segment of critical measures $s \mapsto s\delta_x + (1-s)\delta_y$ whose image by $\mathcal{J}$ is the segment $[V(x), V(y)]$. This shows that the set of critical values of $\mathcal{J}$ is given by the convex envelope of the critical values of V, implying a quasi-systematic failure of Sard's theorem. Hence formulating Łojasiewicz inequalities on slice of level sets is no longer relevant.

Instead we proceed as follows:

Lemma 2 (Lifting the gradient inequality on $\mathcal{P}_2(\mathbb{R}^d)$). *Let* $\bar{V} \in \mathbb{R}$ *and* $X \subset \mathbb{R}^n$. *Consider* $\theta \in (0,1]$ *and* $c > 0$. *If*

$$\left| V(x) - \bar{V} \right|^{1-\theta} \leq c \, |\nabla V(x)|, \text{ for all } x \text{ in } X, \tag{29}$$

then

$$\left| \mathcal{J}[\rho] - \bar{V} \right|^{1-\theta} \leq c \, \|\partial \mathcal{J}[\rho]\|_{\rho},$$

for all ρ in $\mathcal{P}_2(\mathbb{R}^d)$ such that $\text{Supp}\,\rho \subset X$.

Proof. Take ρ such that $\text{Supp}\,(\rho) \subset X$. By Hölder's inequality and (29), we have

$$
\begin{aligned}
\left| \mathcal{J}[\rho] - \bar{V} \right| &= \left| \int_X \rho(x)(V(x) - \bar{V}) \, \mathrm{d}x \right| \\
&\leq c^{1/(1-\theta)} \int_X \rho(x) \, |\nabla V(x)|^{1/(1-\theta)} \, \mathrm{d}x \\
&\leq c^{1/(1-\theta)} \left[\int_X \rho(x) \, |\nabla V(x)|^2 \, \mathrm{d}x \right]^{1/(2(1-\theta))} \\
&= c^{1/(1-\theta)} \|\partial \mathcal{J}[\rho]\|_{\rho}^{1/(1-\theta)} .
\end{aligned}
$$

Which is the stated result. $\qquad\square$

As a consequence we obtain the following general Łojasiewicz gradient inequality for the class of potential energies.

Proposition 1 (Łojasiewicz gradient inequalities for potential energies). *Let $V : \mathbb{R}^d \to \mathbb{R}$ be a differentiable subanalytic function and $\mathcal{C}$ a connected component of $\nabla V^{-1}(0)$.*
  (i) $\mathcal{J}$ is constant on the set of measures having support in $\mathcal{C}$, we denote by $\bar{\mathcal{J}}$ this value.
  (ii) Fix $R > 0$. There exist $\epsilon > 0$, $c > 0$, and $\theta \in (0,1]$ such that

$$\left| \mathcal{J}[\rho] - \bar{\mathcal{J}} \right|^{1-\theta} \leq c \left\| \partial \mathcal{J}[\rho] \right\|_\rho,$$

for all ρ in $\mathcal{P}_2(\mathbb{R}^d)$ such that $\operatorname{Supp} \rho \subset \left\{ x \in \mathbb{R}^d : \operatorname{dist}(x, \mathcal{C}) < \epsilon, |x| \leq R \right\}$.

Proof. Standard results on subanalytic geometry can be found in [29]. Observe first that, by subanaliticity and continuity, V must be constant on $\mathcal{C}$. We then combine the classical Łojasiewicz inequality and a compactness argument to obtain a uniform inequality on

$$X := \left\{ x \in \mathbb{R}^d : \operatorname{dist}(x, \mathcal{C}) < \epsilon, |x| \leq R \right\}$$

for some $\epsilon > 0$. The conclusion follows by Lemma 2.   □

Potential energies provide straightforward examples of convergence of subgradient flows of the form (see [2, Example 11.2.2, p. 298]):

$$\frac{\mathrm{d}}{\mathrm{d}t} \rho(t) + \nabla \cdot (\rho(t) v(t)) = 0 \qquad \text{a.e. on } (0, +\infty), \tag{30}$$

$$v(t) \in \partial V(\rho(t)) \qquad \text{a.e. on } (0, +\infty), \tag{31}$$

covering all possible convergence rates since θ may range in $(0,1]$ (use Theorem 2). For instance when the norm of the subgradients of V are bounded away from zero at each point, save of course at minimisers, the functional $\mathcal{J}$ is sharp and convergence of the associated flow occurs in finite time.

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 [35, 2, 1, 32].

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

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

A transport map $T : X \to Y$ between μ and ν is a Borel map such that $T_\# \mu = \nu$.

Assume $X = Y$ and denote by d the distance on X. For $(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(X)$ the Monge optimal transport problem writes

$$\mathcal{W}_2(\mu, \nu) = \sqrt{\inf \left\{ \int_X d^2(x, T(x)) \, \mathrm{d}\mu(x) : T \text{ Borel map s.t. } T_\# \mu = \nu \right\}}. \tag{32}$$

$\mathcal{W}_2$ defines a distance on the subset of probabilities on X with finite second-order moments. Given $\mathcal{S} \subset \mathcal{P}(X)$ and μ in $\mathcal{P}(X)$, we set $\mathcal{W}_2(\mu, \mathcal{S}) = \inf \{ \mathcal{W}_2(\mu, \nu) : \nu \in \mathcal{S} \}$.

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

$$\mathcal{P}_2(\mathbb{R}^d) = \left\{ \mu \in \mathcal{P}(\mathbb{R}^d) : \int_{\mathbb{R}^d} |x|^2 \, \mathrm{d}\mu(x) < +\infty \right\}. \tag{33}$$

A solution to (32) is called *an optimal transport.* Such a transport generally exists thanks to:

Theorem 5 (Brenier's theorem [10]). *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 has a unique solution T to (32). 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.*

In the absence of regularity of the source measure, Monge's formulation can be relaxed into the so-called Kantorovich's formulation

$$\mathcal{W}_2^2(\mu, \nu) := \min_{\gamma \in \Pi(\mu, \nu)} \iint_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 \, \mathrm{d}\gamma(x, y)$$

where $\Pi(\mu, \nu)$ is the set of measures in $\mathbb{R}^d \times \mathbb{R}^d$ whose first and second marginals are respectively μ and ν. This problem always has a solution called an optimal transport plan; the set of such plans is denoted by $\Gamma_{\mathrm{opt}}(\mu, \nu)$.

4.2. **Convexity and geodesics.** Let us consider real-extended-valued functions $\mathcal{J} : \mathcal{P}_2(\mathbb{R}^d) \to (-\infty, +\infty]$ for which we set $\mathrm{dom}\,\mathcal{J} = \{\rho \in \mathcal{P}_2(\mathbb{R}^d) : \mathcal{J}[\rho] < +\infty\}$. These functionals are called *proper* when $\mathrm{dom}\,\mathcal{J} \neq \emptyset$.

For ρ, μ, ν in $\mathcal{P}_2(\mathbb{R}^d)$, we denote by $\Pi(\rho, \mu, \nu)$ the set of measures in $\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d$ whose first, second and third marginals are respectively ρ, μ, and ν.

From [2, Definition 9.2.4, p. 207] we recall:

Definition 1 (Generalised geodesics and λ-convexity). Set $\pi_1(x, y, z) = (x, y)$ and $\pi_2(x, y, z) = (x, z)$ for all x, y, z in $\mathbb{R}^d$.

(i) Let ρ, μ_1, μ_2 be in $\mathcal{P}_2(\mathbb{R}^d)$. Take $\Psi \in \Pi(\rho, \mu_1, \mu_2)$ such that $(\pi_1)_\# \Psi \in \Gamma_{\mathrm{opt}}(\rho, \mu_1)$, $(\pi_2)_\# \Psi \in \Gamma_{\mathrm{opt}}(\rho, \mu_2)$. Define

$$\pi^t = (1 - t)\pi_1 + t\pi_2, \ t \in [0, 1],$$

then the generalised geodesic joining μ_1 to μ_2 with base ρ induced by Ψ is given by:

$$[0, 1] \ni t \to \pi^t_\#(\Psi).$$

(ii) Let $\lambda \in \mathbb{R}$. A functional $\mathcal{J} \colon \mathcal{P}_2(\mathbb{R}^d) \to (-\infty, +\infty]$ is called λ-*convex along generalised geodesics* if for every pair $(\nu_0, \nu_1) \in \mathrm{dom}\,\mathcal{J} \times \mathrm{dom}\,\mathcal{J}$ and any generalised geodesic $\{\nu_t\}_{t \in [0,1]}$ with base $\rho \in \mathcal{P}_2(\mathbb{R}^d)$ induced by $\Psi \in \Pi(\rho, \nu_0, \nu_1)$, one has

$$\mathcal{J}[\nu_t] \leq (1 - t)\mathcal{J}[\nu_0] + t\mathcal{J}[\nu_1] - \frac{\lambda}{2}\mathcal{W}_{2,\Psi}^2(\nu_0, \nu_1)$$

for every $t \in [0, 1]$ and where

$$\mathcal{W}_{2,\Psi}^2(\nu_0, \nu_1) := \iiint_{\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d} |x_3 - x_2|^2 \, \mathrm{d}\Psi(x_1, x_2, x_3).$$

Remark 6. The above definition implies in particular that the domain of $\mathcal{J}$ is convex in the following sense: for any regular measure γ, and for any measures μ and ν in $\mathrm{dom}\,\mathcal{J}$, the generalised interpolant with base γ between μ and ν exists and lies in $\mathrm{dom}\,\mathcal{J}$.

When $\lambda = 0$, which is a major focus within this article, the definition simplifies into:

Definition 2 (Generalised convexity). A functional $\mathcal{J} \colon \mathcal{P}_2(\mathbb{R}^d) \to (-\infty, +\infty]$ is called *convex* (along generalised geodesics)[7] if for every pair $(\nu_0, \nu_1) \in \mathrm{dom}\,\mathcal{J} \times \mathrm{dom}\,\mathcal{J}$ and any generalised geodesic $\{\nu_t\}_{t \in [0,1]}$ between ν_0 and ν_1, one has

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

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

---

[7]Convexity in the Monge-Kantorovich spaces is sometimes referred to as *displacement convexity*.

4.3. **Metric/Riemannian aspects,** [30, 2]**.** We first recall a purely metric notion: a curve $\gamma : (a,b) \to \mathcal{P}_2(\mathbb{R}^d)$ is called absolutely continuous if

$$\mathcal{W}_2(\gamma(t), \gamma(s)) = \int_t^s v(\tau) \, \mathrm{d}\tau, \; \forall t, s \in (a,b)$$

where v is $L^1(\mathbb{R})$. The metric derivative of γ at $t \in (a,b)$ is given by

$$\left| \frac{\mathrm{d}\gamma}{\mathrm{d}t} \right| (t) = \lim_{h \to 0} \frac{\mathcal{W}_2(\gamma(t+h), \gamma(t))}{|h|}.$$

This quantity is well defined for almost every $t \in (a,b)$, see [2, Theorem 1.1.2, p.24].

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 $\|s\|_\rho := \sqrt{\langle s, s \rangle_\rho}$.

Let $\mathcal{J} : \mathcal{P}_2(\mathbb{R}^d) \to (-\infty, +\infty]$ be a 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 the reduced subdifferential [2, Definition 10.3.1, p.241 and Remark 10.3.3 p.243] which also admits the equivalent formulation in the case of geodesically convex function, and thus in particular for convex functions in the sense of the previous definition:

Definition 3. ([2, Item 10.3.13 p. 243 and Theorem 10.3.6 p. 244] Reduced subdifferential) Let $\mathcal{J} : \mathcal{P}_2(\mathbb{R}^d) \to (-\infty, +\infty]$ be a λ-convex lower semi-continuous function bounded from below with $\lambda \in \mathbb{R}$. Take $\mu \in \operatorname{dom} \mathcal{J}$ and $\gamma \in L^2_\mu(\mathbb{R}^d)$. Then

$$\gamma \in \partial \mathcal{J}[\mu] \qquad \Longleftrightarrow \qquad \forall \nu \in \mathcal{P}_2(\mathbb{R}^d), \exists \Psi \in \Gamma_{\mathrm{opt}}(\mu, \nu),$$

$$\mathcal{J}[\nu] \geq \mathcal{J}[\mu] + \int_{\mathbb{R}^d \times \mathbb{R}^d} \gamma(x) \cdot (y - x) \, \mathrm{d}\Psi(x,y) + \frac{\lambda}{2} \mathcal{W}_2^2(\mu, \nu).$$

When $\mu \notin \operatorname{dom} \mathcal{J}$, the set $\partial \mathcal{J}[\mu]$ is empty.

One defines *the minimal norm subgradient*, whenever it exists, by

$$\partial^0 \mathcal{J}[\mu] = \operatorname{Argmin} \{ \|\gamma\|_\mu : \gamma \in \partial \mathcal{J}[\mu] \}.$$

Using [2, Theorem 10.3.10, p.246], we have $\operatorname{dom} \partial \mathcal{J} \subset \operatorname{dom} |\nabla| \mathcal{J}$ and

$$|\nabla| \mathcal{J}[\mu] \leq \|\partial^0 \mathcal{J}[\mu]\|, \quad \forall \mu \in \mathcal{P}_2(\mathbb{R}^d). \tag{34}$$

References

[1] L. Ambrosio and N. Gigli, *A user's guide to optimal transport,* Modelling and optimisation of flows on networks. Springer Berlin Heidelberg, 2013. pp. 1–155.

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

[3] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, and G. Scheffer, *Sur les inégalités de Sobolev logarithmiques*, volume 10 of Panoramas et Syntheses [Panoramas and Synthèses]. Société Mathématique de France, Paris, 2000, vol. 5. (2000).

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

[5] S. G. Bobkov and Y. Ding, *Optimal transport and Rényi informational divergence*, Electronic Communications in Probability, 2015, vol. 20.

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

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

[8] 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.

[9] J. Bolte, T. P. Nguyen, J. Peypouquet, and B. Suter, *From error bounds to the complexity of first-order descent methods for convex functions*, Math. Prog. Ser. A, 165 (2017), pp. 471–507.

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

[11] J. A. Carrillo, Jüngel, A., Markowich, P. A., Toscani, G., and Unterreiter, A., *Entropy dissipation methods for degenerate parabolic problems and generalized Sobolev inequalities*, Monatshefte für Mathematik, 133(1), (2001), pp. 1–82.

[12] 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.

[13] ———, *Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates*, Revista Matematica Iberoamericana, 19 (2003), pp. 971–1018.

[14] 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.

[15] R. Chill, *On the Łojasiewicz–Simon gradient inequality*, Journal of Functional Analysis, 201 (2003), pp. 572–601.

[16] D. Cordero-Erausquin, W. Gangbo, and C. Houdré, *Inequalities for generalized entropy and optimal transportation*, Contemporary Mathematics, 353 (2004), pp. 73–94.

[17] 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.

[18] H. Djellout, A. Guillin, and L. Wu, *Transportation cost-information inequalities and applications to random dynamical systems and diffusions*, The Annals of Probability, 32 (2004), pp. 2702–2732.

[19] W. Gangbo and R. J. McCann, *The geometry of optimal transportation*, Acta Mathematica, 177 (1996), pp. 113–161.

[20] N. Gigli and M. Ledoux, *From log Sobolev to Talagrand: a quick proof*, Discrete and Continuous Dynamical Systems-Series A, (2013).

[21] N. Gozlan and C. Léonard, *Transport inequalities. A survey*, Markov Process. Related Fields, 16(4) (2010), pp. 635–736.

[22] 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.

[23] R. Holley and D. Stroock, *Logarithmic Sobolev inequalities and stochastic Ising models*, Journal of statistical physics 46 (5), (1987), pp. 1159–1194.

[24] S. I. Ohta and A. Takatsu, *Displacement convexity of generalized relative entropies*, Advances in Mathematics, 228 (2011), pp. 1742 – 1787.

[25] 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.

[26] K. Kurdyka, *On gradients of functions definable in o-minimal structures*, in Annales de l'institut Fourier, vol. 48, 1998, pp. 769–783.

[27] M. Ledoux, *Inégalités isopérimétriques en analyse et probabilités*, Séminaire Bourbaki, 35 (1993), pp. 343–375.

[28] ———, *The concentration of measure phenomenon*, no. 89, American Mathematical Soc., 2005.

[29] S. Łojasiewicz, *Ensembles semi-analytiques*, Inst. Hautes Études Sci., Bures-sur-Yvette, (1965).

[30] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Communications in Partial Differential Equations, 26 (2001), pp. 101–174.

[31] 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.

[32] F. Santambrogio, *Optimal transport for applied mathematicians*, Progress in Nonlinear Differential Equations and their applications, 87 (2015).

[33] L. Simon, *Asymptotics for a class of non-linear evolution equations, with applications to geometric problems*, Annals of Mathematics, (1983), pp. 525–571.

[34] J. L. Vázquez, *The porous medium equation: mathematical theory*, Oxford University Press, 2007.

[35] 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, 31015, Toulouse, France.

Email address: adrien.blanchet@tse-fr.eu, jerome.bolte@tse-fr.eu