



An approximation of the squared Wasserstein distance and an application to Hamilton-Jacobi equations

Charles Bertucci, Pierre Louis Lions

► To cite this version:

Charles Bertucci, Pierre Louis Lions. An approximation of the squared Wasserstein distance and an application to Hamilton-Jacobi equations. 2024. hal-04700766

HAL Id: hal-04700766

<https://hal.science/hal-04700766v1>

Preprint submitted on 17 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

AN APPROXIMATION OF THE SQUARED WASSERSTEIN DISTANCE AND AN APPLICATION TO HAMILTON-JACOBI EQUATIONS

CHARLES BERTUCCI ^A, PIERRE-LOUIS LIONS^{B,C}

ABSTRACT. We provide a simple $C^{1,1}$ approximation of the squared Wasserstein distance on $\mathbb{R}^d$ when one of the two measures is fixed. This approximation converges locally uniformly. More importantly, at points where the differential of the squared Wasserstein distance exists, it attracts the differentials of the approximations at nearby points. Our method relies on the Hilbertian lifting of PL Lions and on the regularization in Hilbert spaces of Lasry and Lions. We then provide an application of this result by using it to establish a comparison principle for an Hamilton-Jacobi equation on the set of probability measures.

CONTENTS

1. Introduction	1
2. Preliminaries	2
3. The approximation	3
4. Regularity of the approximation	5
5. Convergence of the gradients	6
6. Applications to Hamilton-Jacobi equations on the space of probability measures	7
Acknowledgments	14
References	14

1. INTRODUCTION

In recent years, the study of partial differential equations on the set of probability measures, sometimes restricted through some integrability condition, has gained a lot of interest. Several difficulties arise when using traditional analysis methods and one of them is that quite often, one would like to use the Wasserstein distance as the natural distance on this set. A main difficulty lies in the fact that, contrary to Euclidean cases, the squared of this function is not smooth. It is only super-differentiable everywhere. We provide here a tool that we believe allows to address this issue in several situations. Namely, we present an approximation of the squared Wasserstein distance which is quite regular. More importantly, this approximation is somehow coherent with the differentiability of the squared Wasserstein distance in the sense that, at points where the differential of the squared Wasserstein distance exists, it attracts the differentials of the approximations at nearby points.

Our method of proof relies on considering an approximation in the spirit of Lasry and Lions [18]. The proof is established by lifting the Wasserstein distance in the space of random variables (the standard lifting introduced by PL Lions in [19]).

We first present our approximation and its convergence. We end this paper by giving an application of this result in the study of an Hamilton-Jacobi equation on the set of probability measures.

2. PRELIMINARIES

2.1. The Wasserstein distance. We denote by $\mathcal{P}_2(\mathbb{R}^d)$ the space of Borel probability measures on $\mathbb{R}^d$ with a finite second order moment, i.e. probability measures μ such that $M_2(\mu) := \int_{\mathbb{R}^d} |x|^2 d\mu < \infty$. The set $\mathcal{P}_2(\mathbb{R}^d)$ is a complete metric space endowed with the Wasserstein distance W_2 defined by

$$W_2^2(\mu, \nu) = \inf_{\gamma \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 \gamma(dx, dy),$$

where $\Pi(\mu, \nu)$ is the set of all couplings between μ and ν , i.e. the set of probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with first marginal μ and second marginal ν .

In all that follows, we consider a standard atomeless probabilistic space $(\Omega, \mathcal{A}, \mathbb{P})$ whose expectation is denoted by $\mathbb{E}$. We denote by $H = \mathbb{L}^2(\Omega, \mathbb{R}^d)$, the space of squared integrable random variables on $\mathbb{R}^d$. Recall that H is a separable Hilbert space endowed with the scalar product $\langle X, Y \rangle = \mathbb{E}[X \cdot Y]$ where $x \cdot y$ is the $\mathbb{R}^d$ scalar product between x and y . The norm of $X \in H$ is noted $\|X\|$. We note $X \sim \mu$ when μ is the law $\mathcal{L}(X)$ of $X \in H$.

The Wasserstein distance W_2 satisfies for all $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$,

$$W_2^2(\mu, \nu) = \inf \{ \mathbb{E}[|X - Y|^2] | X \sim \mu, Y \sim \nu \}.$$

We have the following immediate result.

Lemma 2.1. *For any $X \sim \mu$, we have that*

$$W_2^2(\mu, \nu) = \inf \{ \mathbb{E}[|X - Y|^2] | Y \sim \nu \}.$$

Proof. Consider an optimal couple (X^*, Y^*) in $W_2^2(\mu, \nu)$. It is a classical result of probabilities, see for instance Lemma 5.23 in [7], that for any $\epsilon > 0$, there exists a measure preserving map $\tau : \Omega \rightarrow \Omega$ such that $\|X - X^* \circ \tau\|_\infty \leq \epsilon$. Observe now that

$$\mathbb{E}[|X - Y^*|^2] \leq W_2^2(\mu, \nu) + 2\epsilon W_2(\mu, \nu) + \epsilon^2.$$

Hence the result follows by taking the limit $\epsilon \rightarrow 0$ in the above expression. $\square$

2.2. Regularity of functions in $\mathcal{P}_2(\mathbb{R}^d)$. Consider $\Phi : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$. We say that Φ is differentiable at $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ if there exists $\phi \in L_\mu^2(\mathbb{R}^d, \mathbb{R}^d)$ such that for all $\mu' \in \mathcal{P}_2(\mathbb{R}^d)$, $\gamma \in \Pi(\mu, \mu')$,

$$\Phi(\mu') - \Phi(\mu) = \int_{\mathbb{R}^d \times \mathbb{R}^d} \phi(x) \cdot (y - x) \gamma(dx, dy) + o \left(\left(\int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 \gamma(dx, dy) \right)^{\frac{1}{2}} \right).$$

In this case, we note for all $x \in \mathbb{R}^d$, $\phi(x) = D_\mu \Phi(\mu, x)$. We say that Φ is continuously differentiable over $\mathcal{P}_2(\mathbb{R}^d)$ if it is differentiable at every point and if for any compact set

$K \subset \mathcal{P}_2(\mathbb{R}^d)$, there exists a continuous function $\omega : [0, \infty) \times [0, \infty)$ with $\omega(0) = 0$ such that for any $\mu, \mu' \in K$, for any coupling $\gamma \in \Pi(\mu, \mu')$

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} |D_\mu \Phi(\mu, x) - D_\mu \Phi(\mu', y)|^2 \gamma(dx, dy) \leq \omega \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 \gamma(dx, dy) \right).$$

Finally, we say that Φ is $C^{1,1}$ if the previous ω can be taken linear.

These notions of regularity might seem unusual stated in this way. Let us emphasize the fact that they are the simple translations of the associate regularity properties for the lift $\tilde{\Phi} : H \rightarrow \mathbb{R}$ defined by $\tilde{\Phi}(X) = \Phi(\mathcal{L}(X))$. In particular, Φ is $C^{1,1}$ if and only if $\tilde{\Phi}$ is $C^{1,1}$, in the usual sense in H .

Recalling Theorem 3.2 in [1], we know that as soon as there exists an optimal (for the quadratic cost) transport map $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$ from μ to ν , then $\Phi : \mu' \rightarrow W_2^2(\mu', \nu)$ is differentiable at μ and

$$D_\mu \Phi(\mu, x) = 2(x - T(x)).$$

In particular, at such points, we have the relation

$$(1) \quad \int_{\mathbb{R}^d} |D_\mu \Phi(\mu, x)|^2 \mu(dx) = 4\Phi(\mu).$$

This relation links the squared of the norm of $D_\mu \Phi$ with Φ itself.

3. THE APPROXIMATION

We fix $\nu \in \mathcal{P}_2(\mathbb{R}^d)$. The previous Lemma provides a natural lift of $\mu \rightarrow W_2^2(\mu, \nu)$ to a function of $U : H \rightarrow \mathbb{R}$ defined by $W_2^2(\mathcal{L}(X), \nu) =: U(X) = \inf_{Y \sim \nu} \mathbb{E}[|X - Y|^2]$.

3.1. Basic properties. Note that U is 1-semi-concave, more precisely, for any $Z \in H$, $X \rightarrow U(X) - \|X - Z\|^2$ is concave as the infimum of linear functions since

$$U(X) - \|X - Z\|^2 = \inf_{Y \sim \nu} \{ \mathbb{E}[(X - Z) \cdot (Z - Y)] + \|Z - Y\|^2 \}.$$

This property hints that a regularization of U by means of a sup-convolution is possible.

For $\delta \in (0, 1)$, we consider

$$(2) \quad \Phi_\delta(\mu) = \sup_{\mu' \in \mathcal{P}_2(\mathbb{R}^d)} \left\{ W_2^2(\mu', \nu) - \frac{1}{\delta} W_2^2(\mu', \mu) \right\}.$$

Thanks to Lemma 2.1, we can lift Φ_δ to H with the formula

$$\begin{aligned} \Phi_\delta(\mathcal{L}(X)) &= \sup_{\mu' \in \mathcal{P}_2(\mathbb{R}^d)} \sup_{X' \sim \mu'} \left\{ W_2^2(\mathcal{L}(X'), \nu) - \frac{1}{\delta} \mathbb{E}[|X - X'|^2] \right\} \\ &= \sup_{X' \in H} \inf_{Y \sim \nu} \left\{ \mathbb{E}[|X' - Y|^2] - \frac{1}{\delta} \mathbb{E}[|X - X'|^2] \right\}. \end{aligned}$$

We shall note $U_\delta(X) = \Phi_\delta(\mathcal{L}(X))$. Note that by definition we always have

$$W_2^2(\mu, \nu) \leq \Phi_\delta(\mu).$$

On the other hand

$$\Phi_\delta(\mathcal{L}(X)) \leq \inf_{Y \sim \nu} \sup_{X' \in H} \left\{ \mathbb{E}[|X' - Y|^2] - \frac{1}{\delta} \mathbb{E}[|X - X'|^2] \right\}.$$

For any Y , the supremum in X' is reached for X' such that $\delta(X' - Y) = X' - X$, which yields $X' = (1 - \delta)^{-1}(X - \delta Y)$. Using this information, we obtain

$$\Phi_\delta(\mathcal{L}(X)) \leq \frac{1}{1 - \delta} \inf_{Y \sim \nu} \mathbb{E}[|X - Y|^2].$$

Hence the convergence of $(\Phi_\delta)_{\delta > 0}$ as $\delta \rightarrow 0$ follows from the relation

$$(3) \quad W_2^2(\mu, \nu) \leq \Phi_\delta(\mu) \leq \frac{1}{1 - \delta} W_2^2(\mu, \nu).$$

3.2. Further description of the approximation. We give sufficient conditions under which the second inequality of (3) is actually an equality. We then present a remark to understand what happens outside the assumptions of the following result.

Proposition 3.1. *Assume that $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ are such that the optimal transport maps T and T' from μ toward ν and vice versa are Lipschitz continuous. Then,*

$$\delta \leq \max \left\{ \sigma^-, \frac{1}{\sigma^+} \right\} \Rightarrow \Phi_\delta(\mu) = \frac{1}{1 - \delta} W_2^2(\mu, \nu),$$

where σ^- is the essential infimum of the smallest eigenvalue¹ of DT' and σ^+ is the essential supremum of the largest eigenvalue of DT .

Before presenting the proof of this result, we insist upon the fact that thanks to the now quite exhaustive literature on Monge-Ampère equations, there are various types of situations in which we can guarantee that the conditions of the previous proposition are satisfied. For instance, it is the case if μ and ν are absolutely continuous with respect to the Lebesgue measure, that they are supported on respectively X and Y , two bounded open sets such that Y is convex, and the densities of μ and ν are bounded from above and away from 0 and that they are Hölder continuous on respectively X and Y . We refer to the survey [17] and the reference therein for more details on the regularity of the transport maps.

Proof. The proof basically consists in showing that the intermediate measure μ' which achieves the maximum in (2) is well behaved. Take (X, Y) an optimal coupling for (μ, ν) , in particular, almost surely, $X = T'(Y)$ and $Y = T(X)$. We want to show that the supremum in $U_\delta(X)$ is reached for

$$X' = \frac{1}{1 - \delta} X - \frac{\delta}{1 - \delta} Y = \frac{1}{1 - \delta} (T' - \delta Id)(Y).$$

Remark that if $\delta \leq \sigma^-$, then $T' - \delta Id$ is monotone. On the other hand, since, almost everywhere on the support of μ , $T'(T(x)) = x$, it follows that this condition also reduces to $\delta \leq \frac{1}{\sigma^+}$. Under this smallness assumption on δ , we have in fact more than the monotonicity of the transport map, we have that it is the gradient of some convex

¹Recall that T and T' are monotone maps.

function, since T' is. Hence, (X', Y) is an optimal coupling, see for instance Theorem 1.48 in [23]. This implies that

$$\begin{aligned}\Phi_\delta(\mu) &\geq \mathbb{E}[|X' - Y|^2] - \frac{1}{\delta} \mathbb{E}[|X - X'|^2] \\ &= \frac{1}{1 - \delta} W_2^2(\mu, \nu).\end{aligned}$$

The result is thus proved. $\square$

We now present a formal remark which we believe helps to understand the case in which δ is not small enough.

Remark 3.2. Assume that we are in the case in which μ and ν are as well behaved as in the previous Proposition and that they have a density which is positive on their support, but that the equality for Φ_δ is not reached, which implies thanks to our result that δ is not small enough. Assume that the intermediate measure μ' , which achieves the maximum in (2), is sufficiently well behaved. Then it must be that for some convex function $g : \mathbb{R}^d \rightarrow \mathbb{R}$, $\nabla g \# \nu = \mu'$. Since μ' and ν are sufficiently well behaved, it cannot be that $D^2 g = 0$ at some point in the support of ν . But on the other hand, if $D^2 g > 0$ uniformly, then the argument of the previous Proposition shows that the equality much be reached.

We believe this hints strongly at the fact that whenever $\Phi_\delta(\mu) < \frac{1}{1 - \delta} W_2^2(\mu, \nu)$, the intermediate measure μ' has some singularity.

4. REGULARITY OF THE APPROXIMATION

We now turn to the regularity of U_δ , which will directly imply some regularity for Φ_δ . Following the computations of Lasry and Lions [18], we obtain semi-concavity and semi-convexity estimates for U_δ .

Proposition 4.1. For any $\delta \in (0, 1)$, U_δ is $C^{1,1}$. More precisely, for any $Z \in H$,

$$X \rightarrow U_\delta(X) + \frac{1}{\delta} \|X - Z\|^2 \text{ is convex,}$$

$$X \rightarrow U_\delta(X) - \frac{1}{1 - \delta} \|X - Z\|^2 \text{ is concave.}$$

Proof. We simply follow the proof from [18]. For any $X \in H$,

$$\begin{aligned}U_\delta(X) &= \sup_{X' \in H} \inf_{Y \sim \nu} \mathbb{E}[|X' - Y|^2] - \frac{1}{\delta} \mathbb{E}[|X - X'|^2] \\ &= -\frac{1}{\delta} \|X\|^2 + \sup_{X' \in H} \left\{ \frac{2}{\delta} \mathbb{E}[X \cdot X'] - \frac{1}{\delta} \|X'\|^2 + \inf_{Y \sim \nu} \mathbb{E}[|X' - Y|^2] \right\}\end{aligned}$$

Hence, U_δ is δ^{-1} semi-convex.

We want to show that Ψ defined by

$$\Psi(X, X') = \inf_{Y \sim \nu} \mathbb{E}[|X' - Y|^2] - \frac{1}{\delta} \mathbb{E}[|X - X'|^2] - C \|X\|^2$$

is concave in (X, X') for a well chosen $C > 0$. Clearly

$$\begin{aligned}\Psi(X, X') &= (1 - \delta)^{-1} \|X'\|^2 - \delta^{-1} \|X\|^2 + \inf_{Y \sim \nu} \{\|Y\|^2 - 2\mathbb{E}[X' \cdot Y]\} \\ &\quad + 2\delta^{-1} \mathbb{E}[X' \cdot X] - C\|X\|^2.\end{aligned}$$

Since $X' \rightarrow \inf_{Y \sim \nu} \{\|Y\|^2 - 2\mathbb{E}[X' \cdot Y]\}$ is concave in X' , we easily deduce that Ψ is indeed concave if $C \geq (1 - \delta)^{-1}$. Recalling the Lemma in [18] for instance, we know that $X \rightarrow \sup_{X'} \Psi(X, X')$ is concave, which implies that U_δ is $(1 - \delta)^{-1}$ semi-concave. Note that the same computation holds if we replace $C\|X\|^2$ by $C\|X - Z\|^2$. $\square$

Furthermore, the gradient of U_δ can be computed as we now show.

Proposition 4.2. *For any $\delta \in (0, 1)$, $X \in H$, there exists a unique $X^* \in H$ such that $U_\delta(X) = U(X^*) - \delta^{-1} \|X - X^*\|^2$. Moreover,*

$$\nabla U_\delta(X) = \frac{2}{\delta}(X^* - X).$$

Proof. The existence of X^* is immediate. Consider any $\tilde{X} \in H$, we can compute

$$\begin{aligned}U_\delta(\tilde{X}) - U_\delta(X) &\geq U(X^*) - \delta^{-1} \|X^* - \tilde{X}\|^2 - U(X^*) + \delta^{-1} \|X^* - X\|^2 \\ &= -2\delta^{-1}(X^* - X)(X - \tilde{X}) - \delta^{-1} \|\tilde{X} - X\|^2.\end{aligned}$$

Hence it follows that $\frac{2}{\delta}(X^* - X) \in \partial^- U_\delta(X)$, since U_δ is C^1 , we deduce that for any $X \in H$, $\partial^- U_\delta(X) = \{\nabla U_\delta(X)\}$, from which the result follows. $\square$

5. CONVERGENCE OF THE GRADIENTS

We now turn to the property that we believe is the most interesting of this approximation: the fact that the gradient of U_δ converges nicely toward the gradient of U , whenever the latter is well defined.

Theorem 5.1. *Consider $X_0 \in H$ such that U is differentiable at X_0 . Consider a sequence $(X_\delta)_{\delta \in (0, 1)}$ such that $\|X_\delta - X_0\| \rightarrow 0$ as $\delta \rightarrow 0$. Then*

$$\lim_{\delta \rightarrow 0} \nabla U_\delta(X_\delta) = \nabla U(X).$$

Proof. For any δ , we note $X_\delta^* \in H$ such that $U_\delta(X_\delta) = U(X_\delta^*) - \delta^{-1} \|X_\delta - X_\delta^*\|^2$. We already saw that $2\delta^{-1}(X_\delta^* - X_\delta) = \nabla U_\delta(X_\delta)$.

Observe that for any $X \in H$, by definition of X_δ^* ,

$$\begin{aligned}U(X) - U(X_\delta^*) &\leq \delta^{-1} \mathbb{E}[\|X - X_\delta\|^2 - \|X_\delta^* - X_\delta\|^2] \\ &= 2\delta^{-1} \mathbb{E}[(X - X_\delta^*)(X_\delta^* - X_\delta)] + \delta^{-1} \|X - X_\delta^*\|^2.\end{aligned}$$

This implies that $2\delta^{-1}(X_\delta^* - X_\delta) \in \partial^+ U(X_\delta^*)$. We now use Moreau's argument of continuity of the super-differential of a concave function. Define $f : X \rightarrow U(X) - \|X - X_0\|^2$ and observe that it is concave. Remark that $2\delta^{-1}(X_\delta^* - X_\delta) - 2(X_\delta^* - X_0) \in \partial^+ f(X_\delta^*)$. Since $X_\delta^* \rightarrow X_0$ as $\delta \rightarrow 0$ and $\partial^+ f(X_0) = \{\nabla U(X_0)\}$, we obtain, using Moreau's result [21], that

$$\frac{2}{\delta}(X_\delta^* - X_\delta) \rightarrow_{\delta \rightarrow 0} \nabla U(X_0).$$

It now remains to show that this convergence is in fact strong. Recall that $2\delta^{-1}(X_\delta^* - X_\delta) \in \partial^+ U(X_\delta^*)$. We want to use the Lipschitz regularity of U to show that this implies a useful bound on the norm of $2\delta^{-1}(X_\delta^* - X_\delta)$. For all $X' \in H, \epsilon > 0$, there exists $Y \in H, Y \sim \nu$, such that $|U(X') - \mathbb{E}[|X' - Y|^2]| \leq \epsilon$. It then follows that

$$\begin{aligned} U(X_\delta^*) - U(X') &\leq \mathbb{E}[|X_\delta^* - Y|^2 - |X' - Y|^2] + \epsilon \\ &\leq \mathbb{E}[|X' - X_\delta^*|^2 + 2(X_\delta^* - X') \cdot (X' - Y)] + \epsilon. \end{aligned}$$

This implies in particular that

$$\limsup_{X' \rightarrow X_\delta^*} \frac{U(X_\delta^*) - U(X')}{|X_\delta^* - X'|} \leq 2W_2(\mathcal{L}(X_\delta^*), \nu).$$

Using this estimate in the translation of the fact that $2\delta^{-1}(X_\delta^* - X_\delta) \in \partial^+ U(X_\delta^*)$ yields

$$\|2\delta^{-1}(X_\delta^* - X_\delta)\| \leq 2W_2(\mathcal{L}(X_\delta^*), \nu).$$

This implies that

$$\limsup_{\delta \rightarrow 0} \|2\delta^{-1}(X_\delta^* - X_\delta)\| \leq 2W_2(\mathcal{L}(X_0), \nu) = \|\nabla U(X_0)\|,$$

where the last equality follows from (1). Hence we obtain that the convergence is in fact a strong convergence which proves the claim. $\square$

Remark 5.2. *One of the strength of this result is that the convergence holds whatever the speed of convergence of $(X_\delta)_{\delta \in (0,1)}$ toward X_0 . Note also that no assumption on ν is needed here.*

We now conclude by giving the following Corollary on the regularity of $\Phi : \mu \rightarrow W_2^2(\mu, \nu)$.

Corollary 5.3. *Consider $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ a point of differentiability of Φ and sequences $(\delta_n)_{n \geq 0}$ of positive number converging toward 0 and $(\mu_n)_{n \geq 0}$ converging toward μ . Then, for any sequence of couplings $(\gamma_n)_{n \geq 0}$ such that $\gamma_n \in \Pi(\mu, \mu_n)$ and*

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 \gamma_n(dx, dy) = 0,$$

it holds that

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^d \times \mathbb{R}^d} |D_\mu \Phi_{\delta_n}(\mu_n, y) - D_\mu \Phi(\mu, x)|^2 \gamma_n(dx, dy) = 0.$$

6. APPLICATIONS TO HAMILTON-JACOBI EQUATIONS ON THE SPACE OF PROBABILITY MEASURES

6.1. Setting. We consider the Hamilton-Jacobi (HJ) equation

$$(4) \quad \lambda U + \int_{\mathbb{R}^d} F(\mu, x, D_\mu U(\mu, x)) \mu(dx) - \int_{\mathbb{R}^d} \operatorname{div}_x(D_\mu U(\mu, x)) \mu(dx) = G(\mu) \text{ in } \mathcal{P}_2(\mathbb{R}^d),$$

where $\lambda > 0, F : \mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ and $G : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$ are given and div_x is the standard divergence operator on $\mathbb{R}^d$. More precise assumptions on F shall be made later

on, but we are only concerned here with subquadratic F , i.e., there exists $C > 0$ such that

$$(5) \quad \forall \mu, x, p, \quad |F(\mu, x, p)| \leq \frac{C_F}{2} (1 + |p|^2).$$

The equation (4), or similar counterparts, are namely studied in mean field optimal control problems, see [14, 15, 12, 2, 24, 10, 9, 11, 8] for recent developments and the reference therein for more details on this topic, large deviations, see for instance [16] or [5] for a slightly different setting or in differential games [6, 15] for instance. One of the main difficulty in (4) is that it involves the term with the divergence which is singular. In the absence of such a term, the study of this type of equation is more standard, we refer for instance to [4].

The standard candidate theory to study this kind of equations is the theory of viscosity solutions. For a presentation of this theory in finite dimensional spaces, we refer to [13]. In the space of probability measures, a general theory is lacking at the moment, as different notions are used depending on variations of (4) which are studied. In most cases, the central result needed to develop a proper study of such HJ equations is a comparison principle between a viscosity super-solution and a viscosity sub-solution, which states that the former is larger than the latter.

In finite dimension, viscosity solutions are usually equivalently defined through smooth test functions or through super/sub differentials. Such an equivalence is not obvious in the space of probability measures, see for instance [4], as they are elements of general super-differentials which are not necessary given by the differential of smooth test functions. Working with sub/super-differentials is somehow similar to allowing the use of the squared Wasserstein distance as a test function.

The main result of this section is to prove that, thanks to presence of the term involving the divergence in (4), we can indeed use the Wasserstein as if it was a smooth test function. The argument is twofold. On one hand we use the approximation introduced above and on the other hand we take advantage of the singular term to prove that point of interest are points at which the Wasserstein distance is differentiable. This strategy was also used in [14], but here we shall be able to ask for fewer regularity, namely because of the approximation we provided earlier in the paper. Another situation in which the exact same strategy is used is the one studied in [5], which is dimension $d = 1$ and where the divergence term is replaced by another term involving the Hilbert transform.

6.2. Definition of viscosity solutions. When working with HJ equations involving singular terms, it is natural to perturb the solution by an appropriate functional. This technique has proven quite useful in [14, 5] for instance. Let us insist upon the fact that this idea has been present in the literature on viscosity solutions for quite a long time. In the case of (4), this function is the entropy defined by

$$E(\mu) := \begin{cases} \int_{\mathbb{R}^d} \log(\mu(x)) \mu(x) dx & \text{if } \mu \ll \text{Leb}, \\ +\infty & \text{else,} \end{cases}$$

where Leb is the Lebesgue measure. Note that E is only defined on a dense subset with empty interior of $\mathcal{P}_2(\mathbb{R}^d)$, hence it is difficult to talk about a differential of E .

Nonetheless, formally it is given by

$$D_\mu E(\mu, x) = \nabla_x \log(\mu(x)).$$

Observe that the size of this quantity can be given by the Fisher information defined by

$$\mathcal{I}(\mu) = \int_{\mathbb{R}^d} |\nabla_x \log(\mu(x))|^2 \mu(dx).$$

The link between (4) and E is then made clear through the formal following computation. Given a smooth map $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$

$$(6) \quad \begin{aligned} \int_{\mathbb{R}^d} \operatorname{div}_x(f(x)) \mu(dx) &= - \int_{\mathbb{R}^d} f(x) \nabla_x \mu(x) dx \\ &= - \int_{\mathbb{R}^d} f(x) \nabla_x (\log(\mu(x))) \mu(dx). \end{aligned}$$

We shall work with the following definition of viscosity solutions.

Definition 6.1. An usc² (resp. lsc) function $U : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$ is a viscosity sub-solution (resp. super-solution) if for any $\kappa \in (0, C_F^{-1})$ (resp. $\kappa \in (-C_F^{-1}, 0)$), $C^{1,1}$ function $\phi : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$, for any μ^* point of maximum (resp. of minimum) of $\mu \rightarrow U(\mu) - \phi(\mu) - \kappa E(\mu)$, it holds that $\mathcal{I}(\mu^*) < \infty$ and

$$\begin{aligned} \lambda \phi(\mu^*) + \int_{\mathbb{R}^d} F(\mu^*, x, D_\mu \phi(\mu^*, x) + \kappa \nabla_x (\log(\mu(x)))) \mu^*(dx) \\ - \int_{\mathbb{R}^d} \operatorname{div}_x(D_\mu \phi(\mu^*, x) + \kappa \nabla_x (\log(\mu(x)))) \mu^*(dx) \leq G(\mu^*), \end{aligned}$$

(resp. the opposite inequality holds).

Note that the previous is well defined by combining $\mathcal{I}(\mu^*) < \infty$ and (5). In particular, because $\mathcal{I}(\mu^*) < \infty$, $\mu \rightarrow W_2^2(\mu, \nu)$ is differentiable at μ^* .

6.3. Using the Wasserstein distance as a test function. From now on, we assume that F satisfies

$$(7) \quad \begin{aligned} \forall \mu, \nu, x, y, p, q, \quad |F(\mu, x, p) - F(\mu, x, q)| &\leq C_F(1 + M_2(\mu) + |x| + |p| + |q|)|p - q|, \\ \forall \mu, \nu, x, y, p, q, \quad |F(\mu, x, p) - F(\nu, y, p)| &\leq C_F(1 + W_2(\bar{\mu}, \bar{\nu}) + |x - y|)(1 + |p|). \end{aligned}$$

For such a coupling F , we can indeed prove that we can use the squared Wasserstein distance as a test function. This is done in the next result, using slightly more involved function, as it will be helpful later on.

Proposition 6.2. Let U be a bounded viscosity sub-solution of (4) (resp. super-solution). Take $\nu \in \mathcal{P}_2(\mathbb{R}^d)$, $\alpha, \epsilon > 0, \kappa \in (0, C_F^{-1})$ (resp. $\alpha, \epsilon < 0$ and $\kappa \in (-C_F^{-1}, 0)$) and $\bar{\mu} \in \mathcal{P}_2(\mathbb{R}^d)$ a point of strict maximum (resp. of strict minimum) of $\mu \rightarrow U(\mu) - \kappa E(\mu) - \frac{\alpha}{2} \int_{\mathbb{R}^d} |x|^2 \mu(dx) - \frac{1}{2\epsilon} W_2^2(\mu, \nu)$. Then, $\mathcal{I}(\bar{\mu}) < \infty$ and

$$\begin{aligned} \lambda U(\bar{\mu}) + \int_{\mathbb{R}^d} F(\bar{\mu}, x, U(\bar{\mu}), p(x) + \alpha x) \bar{\mu}(dx) - \int_{\mathbb{R}^d} \operatorname{div}_x(p(x) + \alpha x + \kappa \nabla_x (\log(\bar{\mu}(x)))) \bar{\mu}(dx) \\ \leq G(\bar{\mu}) + C\kappa(\kappa \mathcal{I}(\mu_\delta) + \sqrt{\mathcal{I}(\bar{\mu})}(\epsilon^{-1} W_2(\bar{\mu}, \nu) + M_2(\bar{\mu})), \end{aligned}$$

²usc stands for upper semi-continuous, while lsc stands for lower semi-continuous.

where C depends only on F and p is given by $p(x) = \epsilon^{-1}(x - T(x))$ where T is the optimal transport map of $\bar{\mu}$ toward ν . (Respectively the opposite inequality holds.)

Proof. We only prove the result for sub-solutions, the other case being symmetric. Take $\delta \in (0, 1)$ and consider Φ_δ defined in (2). For any δ , consider $\mu_\delta \in \mathcal{P}_2(\mathbb{R}^d)$ a point of maximum of $\mu \rightarrow U(\mu) - \kappa E(\mu) - \frac{\alpha}{2} \int_{\mathbb{R}^d} |x|^2 \mu(dx) - \frac{1}{2\epsilon} \Phi_\delta(\mu)$. By assumption, we obtain that

$$(8) \quad \begin{aligned} & \lambda U(\mu_\delta) + \int_{\mathbb{R}^d} F(\mu_\delta, x, \epsilon^{-1} D_\mu \Phi_\delta(\mu_\delta, x) + \alpha x + \kappa \nabla_x(\log(\mu_\delta(x)))) \mu_\delta(dx) \\ & - \int_{\mathbb{R}^d} \operatorname{div}_x(\epsilon^{-1} D_\mu \Phi_\delta(\mu_\delta, x) + \alpha x + \kappa \nabla_x(\log(\mu_\delta(x)))) \mu_\delta(dx) \leq G(\mu_\delta). \end{aligned}$$

Note that $(\mu_\delta)_{\delta>0}$ converges weakly toward $\bar{\mu}$ as $\delta \rightarrow 0$. Moreover, by construction of μ_δ , we also obtain that $(M_2(\mu_\delta))_{\delta>0}$ is bounded. Using (5) and (6) with $f = \nabla_x(\log(\mu))$, as well as standard quadratic estimates, we deduce that

$$\begin{aligned} (\kappa - C_F \kappa^2) \mathcal{I}(\mu_\delta) & \leq \left(\alpha \sqrt{M_2(\mu_\delta)} + \epsilon^{-1} \|D_\mu \Phi_\delta(\mu_\delta)\|_{L^2(\mu_\delta)} \right) \sqrt{\mathcal{I}(\mu_\delta)} \\ & + 2C_F(\alpha^2 M_2(\mu_\delta) + \epsilon^{-2} \|D_\mu \Phi_\delta(\mu_\delta)\|_{L^2(\mu_\delta)}^2) + \|G\|_\infty + \lambda \|U\|_\infty. \end{aligned}$$

Hence, we obtain that there exists $C > 0$ depending only on $\|G\|_\infty, \|U\|_\infty, \lambda$ and F such that

$$\kappa \sqrt{\mathcal{I}(\mu_\delta)} \leq C \left(1 + \alpha \sqrt{M_2(\mu_\delta)} + \epsilon^{-1} \|D_\mu \Phi_\delta(\mu_\delta)\|_{L^2(\mu_\delta)} \right).$$

Since $\mu \rightarrow W_2^2(\mu, \nu)$ is differentiable at $\bar{\mu}$, recall that thanks to Theorem 5.1,

$$\lim_{\delta \rightarrow 0} \|D_\mu \Phi_\delta(\mu_\delta)\|_{L^2(\mu_\delta)} = 2W_2(\bar{\mu}, \nu).$$

Having these estimates in hand, it now remains to show that we can indeed pass to the limit into (8). The upper-semi continuity of U and the continuity of G yields

$$\lambda U(\mu_\delta) \rightarrow_{\delta \rightarrow 0} \lambda U(\bar{\mu}), \quad \lim_{\delta \rightarrow 0} G(\mu_\delta) = G(\bar{\mu}).$$

Observe now that thanks to (7),

$$\begin{aligned} & |F(\mu_\delta, x, \epsilon^{-1} D_\mu \Phi_\delta(\mu_\delta, x) + \alpha x + \kappa \nabla_x(\log(\mu_\delta(x)))) - F(\mu_\delta, x, \epsilon^{-1} D_\mu \Phi_\delta(\mu_\delta, x) + \alpha x)| \\ & \leq C_F \kappa (1 + |\epsilon^{-1} D_\mu \Phi_\delta(\mu_\delta, x)| + M_2(\mu_\delta) + \alpha |x| + \kappa |\nabla_x(\log(\mu_\delta(x)))|) |\nabla_x(\log(\mu_\delta(x)))|. \end{aligned}$$

Hence, using Cauchy-Schwarz inequality, we deduce that

$$(9) \quad \begin{aligned} & \int_{\mathbb{R}^d} F(\mu_\delta, x, \epsilon^{-1} D_\mu \Phi_\delta(\mu_\delta, x) + \alpha x + \kappa \nabla_x(\log(\mu_\delta(x)))) \mu_\delta(dx) \\ & \geq \int_{\mathbb{R}^d} F(\mu_\delta, x, \epsilon^{-1} D_\mu \Phi_\delta(\mu_\delta, x) + \alpha x) \mu_\delta(dx) \\ & - C_F \kappa \left(\kappa \mathcal{I}(\mu_\delta) + \sqrt{\mathcal{I}(\mu_\delta)} \left(\epsilon^{-1} \|D_\mu \Phi_\delta(\mu_\delta)\|_{L^2(\mu_\delta)} + \alpha \sqrt{M_2(\mu_\delta)} + M_2(\mu_\delta) \right) \right), \end{aligned}$$

for a possible different C_F , but which depends only on F . To pass to the limit in the right hand side, remark that $(\mathcal{I}(\mu_\delta))_\delta$ and $(M_2(\mu_\delta))_\delta$ are bounded (in δ). It now remains

to show the convergence of the first term of the right hand side of (9). In order to do so, consider $X_\delta \in H$, $X_\delta \sim \mu_\delta$ which converges toward $\bar{X} \in H$, $\bar{X} \sim \bar{\mu}$. Let us remark that

$$\begin{aligned} & \int_{\mathbb{R}^d} F(\mu_\delta, x, U(\mu_\delta), \epsilon^{-1} D_\mu \Phi_\delta(\mu_\delta, x) + \alpha x) \mu_\delta(dx) \\ &= \mathbb{E} [F(\mu_\delta, X_\delta, U(\mu_\delta), \epsilon^{-1} \nabla U_\delta(X_\delta) + \alpha X_\delta)]. \end{aligned}$$

Using Theorem 5.1 and the regularity of F , we deduce that

$$\begin{aligned} & \lim_{\delta \rightarrow 0} \int_{\mathbb{R}^d} F(\mu_\delta, x, U(\mu_\delta), \epsilon^{-1} D_\mu \Phi_\delta(\mu_\delta, x) + \alpha x) \mu_\delta(dx) \\ &= \int_{\mathbb{R}^d} F(\bar{\mu}, x, U(\bar{\mu}), \epsilon^{-1} p(x) + \alpha x) \bar{\mu}(dx). \end{aligned}$$

To complete the proof, it only remains to remark that the term $\int_{\mathbb{R}^d} \operatorname{div}_x(D_\mu \Phi_\delta(\mu_\delta, x)) \mu_\delta(dx)$ converges to the correct limit, which is proved in a separate Lemma. $\square$

Lemma 6.3. *Under the assumptions and notation of the proof of Proposition 6.2,*

$$\lim_{\delta \rightarrow 0} \int_{\mathbb{R}^d} \operatorname{div}_x(D_\mu \Phi_\delta(\mu_\delta, x)) \mu_\delta(dx) = \int_{\mathbb{R}^d} \operatorname{div}_x(p(x)) \bar{\mu}(dx).$$

Proof. Let us start by remarking that

$$\int_{\mathbb{R}^d} \operatorname{div}_x(D_\mu \Phi_\delta(\mu_\delta, x)) \mu_\delta(dx) = - \int_{\mathbb{R}^d} D_\mu \Phi_\delta(\mu_\delta, x) \nabla_x(\log(\mu_\delta(x))) \mu_\delta(dx).$$

Hence, denoting $\psi_\delta(x) = \nabla_x(\log(\mu_\delta(x)))$, we deduce that

$$\int_{\mathbb{R}^d} \operatorname{div}_x(D_\mu \Phi_\delta(\mu_\delta, x)) \mu_\delta(dx) = -\mathbb{E}[\nabla U_\delta(X_\delta) \psi_\delta(X_\delta)].$$

Since $(\mathcal{I}(\mu_\delta))_\delta$ is bounded, we deduce that $(\psi_\delta(X_\delta))_\delta$ is bounded in H , thus it converges weakly to some limit $Z \in H$. Observe that the proof is complete by using a weak-strong convergence result, thanks to Theorem 5.1, if we can show that $Z = \nabla_x(\log(\bar{\mu}(\bar{X})))$. We now show that this is the case, using a form of convexity of E . Recalling the results of McCann [20], we know that E is displacement convex. This implies in particular that for any smooth function $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$ with bounded derivatives, for $h \in (\frac{-\|D_x f\|_\infty}{2}, \frac{-\|D_x f\|_\infty}{2})$

$$h \rightarrow E((Id + hf)_\# \mu_\delta) \text{ is convex,}$$

since the optimal transport of μ_δ toward $(Id + hf)_\# \mu_\delta$ is obtained precisely through the map $Id + hf$. This yields

$$E((Id + hf)_\# \mu_\delta) \geq E(\mu_\delta) + h \int_{\mathbb{R}^d} \psi_\delta(x) f(x) \mu_\delta(dx).$$

Passing to the limit $\delta \rightarrow 0$ produces the required result. Indeed, in the Hilbert space for instance, the previous can be written

$$E((Id + hf)_\# \mu_\delta) \geq E(\mu_\delta) + h \mathbb{E}[\psi_\delta(X_\delta) f(X_\delta)],$$

and passing to the limit $\delta \rightarrow 0$, since we can always choose $(X_\delta)_{\delta > 0}$ converging strongly toward $\bar{X}$, implies

$$E((Id + hf)_\# \bar{\mu}) \geq E(\bar{\mu}) + h \mathbb{E}[Z f(\bar{X})].$$

Note that since $(\mathcal{I}(\mu_\delta))_\delta$ is bounded, we deduce that $E(\mu_\delta) \rightarrow E(\bar{\mu})$. Recall now that $\mathcal{I}(\bar{\mu}) < \infty$ and that the derivative of E at $\bar{\mu}$ in smooth variations can be computed explicitly to obtain that $Z = \nabla_x \log(\bar{\mu}(\bar{X}))$. $\square$

An immediate corollary of Proposition 6.2 is the following.

Corollary 6.4. *Under the same notation as in Proposition 6.2, if $\bar{\mu}$ is a point of strict maximum of $\mu \rightarrow U(\mu) - \kappa E(\mu) - \frac{\alpha}{2} \int_{\mathbb{R}^d} |x|^2 \mu(dx) - \int_{\mathbb{R}^d} g d\mu - \frac{1}{2\epsilon} W_2^2(\mu, \nu)$ for a smooth bounded g . Then, $\mathcal{I}(\bar{\mu}) < \infty$ and*

$$\begin{aligned} & \lambda U(\bar{\mu}) + \int_{\mathbb{R}^d} F(\bar{\mu}, x, U(\bar{\mu}), p(x) + \alpha x + \nabla g(x)) \bar{\mu}(dx) \\ & - \int_{\mathbb{R}^d} \operatorname{div}_x(p(x) + \alpha x + \nabla g(x) + \kappa \nabla_x(\log(\bar{\mu}(x)))) \bar{\mu}(dx) \\ & \leq G(\bar{\mu}) + C\kappa(\kappa \mathcal{I}(\mu_\delta) + \sqrt{\mathcal{I}(\bar{\mu})}(\epsilon^{-1} W_2(\bar{\mu}, \nu) + M_2(\bar{\mu})), \end{aligned}$$

for the same constant C as in Proposition 6.2. The similar also holds for super-solutions.

6.4. Comparison principle. We now end this paper by giving a comparison principle for (4). Note that, unlike most comparison principles found in the literature around (4) [14, 15, 2, 24], the following does not require any Lipschitz continuity on G or on the viscosity sub/super solutions. We are not asking for any additional regularity on the viscosity sub and super solutions and only on uniform continuity on G . Let us note that, for simpler Hamiltonians, [15] established a similar results by approximation of measures by combination of Dirac masses.

Theorem 6.5. *Assume that U and V are respectively bounded viscosity sub and super solutions of (4), with right hand side given by respectively G^1 and G^2 , two mappings from $\mathcal{P}_2(\mathbb{R}^d)$ into $\mathbb{R}$ such that one of them is uniformly continuous. Then,*

$$\lambda \sup_{\mathcal{P}_2(\mathbb{R}^d)} (U - V) \leq \sup_{\mathcal{P}_2(\mathbb{R}^d)} (G^1 - G^2).$$

Proof. Consider $\alpha \in (0, 1), \epsilon, \eta, \kappa > 0$. Using a variant of Stegall's Lemma [22], such as an immediate extension of Lemma 2.1 in [3] for instance, we can always choose $g_1, g_2 \in C^2(\mathbb{R}^d, \mathbb{R})$ with $\|g_1\|_{C^2}, \|g_2\|_{C^2} \leq \eta$, such that there exists $(\bar{\mu}, \bar{\nu})$ a point of strict maximum of

$$\begin{aligned} & (\mu, \nu) \rightarrow U(\mu) - V(\nu) - \frac{1}{2\epsilon} W_2^2(\mu, \nu) - \frac{\alpha}{2} (M_2(\mu) + M_2(\nu)) - \kappa (E(\mu) + E(\nu)) \\ & - \int_{\mathbb{R}^d} g_1(x) \mu(dx) - \int_{\mathbb{R}^d} g_2(x) \nu(dx). \end{aligned}$$

Using Proposition 6.2, we deduce that $\mathcal{I}(\bar{\mu}) + \mathcal{I}(\bar{\nu}) < \infty$ and that

$$\begin{aligned} & \lambda U(\bar{\mu}) - \int_{\mathbb{R}^d} \operatorname{div}_x(p(x) + \alpha x + \nabla_x g_1(x) + \kappa \nabla_x(\log(\bar{\mu}(x)))) \bar{\mu}(dx) \\ & + \int_{\mathbb{R}^d} F(\bar{\mu}, x, p(x) + \alpha x + \nabla_x g_1(x)) \bar{\mu}(dx) \leq G_1(\bar{\mu}) + C\kappa(\kappa \mathcal{I}(\bar{\mu}) + \sqrt{\mathcal{I}(\bar{\mu})}(\epsilon^{-1} W_2^2(\bar{\mu}, \bar{\nu}) + M_2(\bar{\mu})), \end{aligned}$$

as well as

(11)

$$\begin{aligned} \lambda V(\bar{\nu}) - \int_{\mathbb{R}^d} \operatorname{div}_x(q(x) + \alpha x + \nabla_x g_2(x) - \kappa \nabla_x(\log(\bar{\nu}(x)))) \bar{\nu}(dx) \\ + \int_{\mathbb{R}^d} F(\bar{\nu}, x, q(x) + \alpha x + \nabla_x g_2(x)) \bar{\nu}(dx) \geq G_2(\bar{\nu}) - C\kappa(\kappa \mathcal{I}(\bar{\nu}) + \sqrt{\mathcal{I}(\bar{\nu})}(\epsilon^{-1}W_2^2(\bar{\mu}, \bar{\nu}) + M_2(\bar{\nu})), \end{aligned}$$

where $p(x) = \epsilon^{-1}(x - T(x))$ and $q(x) = \epsilon^{-1}(x - T^{-1}(x))$ for T the optimal transport map from $\bar{\mu}$ to $\bar{\nu}$ for the quadratic cost. Similarly to what we have done in the proof of Proposition 6.2, using (7), we obtain that

$$\begin{aligned} \left| \int_{\mathbb{R}^d} F(\bar{\mu}, x, p(x) + \alpha x + \nabla_x g_1(x)) \bar{\mu}(dx) - \int_{\mathbb{R}^d} F(\bar{\mu}, x, p(x)) \bar{\mu}(dx) \right| \\ \leq C(1 + \epsilon^{-1}W_2(\bar{\mu}, \bar{\nu}))(\alpha \sqrt{M_2(\bar{\mu})} + \eta), \end{aligned}$$

with a similar inequality for V . Combining these informations with taking the difference of (10) and (11), we obtain

$$\begin{aligned} (12) \quad & \lambda(U(\bar{\mu}) - V(\bar{\nu})) + \int_{\mathbb{R}^d} F(\bar{\mu}, x, p(x)) \bar{\mu}(dx) + \kappa(\mathcal{I}(\bar{\mu}) + \mathcal{I}(\bar{\nu})) \\ & - \int_{\mathbb{R}^d} F(\bar{\nu}, x, q(x)) \bar{\nu}(dx) - \int_{\mathbb{R}^d} \operatorname{div}_x(p(x)) \bar{\mu}(dx) + \int_{\mathbb{R}^d} \operatorname{div}_x(q(x)) \bar{\nu}(dx) \\ & \leq G_1(\bar{\mu}) - G_2(\bar{\nu}) + C(1 + \epsilon^{-1}W_2(\bar{\mu}, \bar{\nu}))(\alpha(\sqrt{M_2(\bar{\mu})} + \sqrt{M_2(\bar{\nu})}) + \eta) \\ & + C\kappa(\kappa(\mathcal{I}(\bar{\mu}) + \mathcal{I}(\bar{\nu})) + (\sqrt{\mathcal{I}(\bar{\mu})} + \sqrt{\mathcal{I}(\bar{\nu})})(\epsilon^{-1}W_2^2(\bar{\mu}, \bar{\nu}) + M_2(\bar{\mu}) + M_2(\bar{\nu})). \end{aligned}$$

We can estimate

$$\begin{aligned} \int_{\mathbb{R}^d} F(\bar{\nu}, x, q(x)) \bar{\nu}(dx) - \int_{\mathbb{R}^d} F(\bar{\mu}, x, p(x)) \bar{\mu}(dx) &= \int_{\mathbb{R}^d} F(\bar{\nu}, T(x), p(x)) - F(\bar{\mu}, x, p(x)) \bar{\mu}(dx) \\ &\leq C \int_{\mathbb{R}^d} (1 + |p(x)|)(W_2(\bar{\mu}, \bar{\nu}) + |T(x) - x|) \bar{\mu}(dx) \\ &\leq CW_2(\bar{\mu}, \bar{\nu})(1 + \epsilon^{-1}W_2(\bar{\mu}, \bar{\nu})), \end{aligned}$$

where we used $T_{\#}\mu = \nu$ in the equality, (7) for the first inequality and Cauchy-Schwarz inequality for the second inequality.

Let us now note π the optimal coupling between $\bar{\mu}$ and $\bar{\nu}$. We can compute

$$\begin{aligned} I &:= \int_{\mathbb{R}^d} \operatorname{div}_x(p(x)) \bar{\mu}(dx) - \int_{\mathbb{R}^d} \operatorname{div}_x(q(x)) \bar{\nu}(dx) \\ &= - \int_{\mathbb{R}^{2d}} (x - y) \cdot (\nabla \log(\bar{\mu}(x)) - \nabla \log(\bar{\nu}(y))) \pi(dx, dy). \end{aligned}$$

Using, as in Lemma 6.3 the convexity of E , we deduce that

$$\begin{aligned} E(\bar{\mu}) + \int_{\mathbb{R}^{2d}} (y - x) \cdot \nabla \log(\bar{\mu}(x)) \pi(dx, dy) &\leq E(\bar{\nu}) \\ E(\bar{\nu}) + \int_{\mathbb{R}^{2d}} (x - y) \cdot \nabla \log(\bar{\nu}(y)) \pi(dx, dy) &\leq E(\bar{\mu}). \end{aligned}$$

Hence, we obtain that $I \leq 0$. Coming back to (12), we deduce that
(13)

$$\begin{aligned} \lambda(U(\bar{\mu}) - V(\bar{\nu})) + \kappa(\mathcal{I}(\bar{\mu}) + \mathcal{I}(\bar{\nu})) &\leq G_1(\bar{\mu}) - G_2(\bar{\nu}) + C(1 + \epsilon^{-1}W_2(\bar{\mu}, \bar{\nu}))(\alpha\sqrt{M_2(\bar{\mu})} + \eta) \\ &\quad + C\kappa(\kappa(\mathcal{I}(\bar{\mu}) + \mathcal{I}(\bar{\nu})) + (\sqrt{\mathcal{I}(\bar{\mu})} + \sqrt{\mathcal{I}(\bar{\nu})})(\epsilon^{-1}W_2^2(\bar{\mu}, \bar{\nu}) + M_2(\bar{\mu}) + M_2(\bar{\nu})) \\ &\quad + CW_2(\bar{\mu}, \bar{\nu})(1 + \epsilon^{-1}W_2(\bar{\mu}, \bar{\nu})). \end{aligned}$$

Using the continuity of either G_1 or G_2 , we deduce that for a modulus of continuity ω , $G_1(\bar{\mu}) - G_2(\bar{\nu}) \leq \sup(G_1 - G_2) + \omega(W_2(\bar{\mu}, \bar{\nu}))$. The result shall follow by taking all the parameters to 0, in some precise order, which we now explain.

Recall first that from standard estimates, there exists a modulus of continuity ω such that $\epsilon^{-1}W_2^2(\bar{\mu}, \bar{\nu}) \leq \omega(\epsilon)$ and $\alpha(M_2(\bar{\mu}) + M_2(\bar{\nu})) \leq \omega(\alpha)$. Moreover, we clearly have $\lambda \sup(U - V) \leq \liminf \lambda U(\bar{\mu}) - V(\bar{\nu})$. Concerning the terms in κ we make the following argument.

From the proof of Proposition 6.2 for instance, $(u_\kappa)_{\kappa>0} := (\kappa(\sqrt{\mathcal{I}(\bar{\mu})} + \sqrt{\mathcal{I}(\bar{\nu})}))_{\kappa>0}$ is bounded, with a bound which depends on all the other parameters. Hence, we deduce from (13) that $(v_\kappa)_{\kappa>0} := (\kappa(\mathcal{I}(\bar{\mu}) + \mathcal{I}(\bar{\nu})))_{\kappa>0}$ is also bounded. It then implies that $\lim_{\kappa \rightarrow 0} u_\kappa = 0$, uniformly in the other arguments, i.e., up to changing ω , we can assume that $u_\kappa \leq \omega(\kappa)$. Hence, combining all the arguments above leads to

$$\begin{aligned} \lambda \sup(U - V) &\leq \sup(G_1 - G_2) + \omega(\epsilon) + C(1 + \epsilon^{-\frac{1}{2}}\omega(\epsilon))(\omega(\alpha) + \eta) \\ &\quad + C\omega(\kappa)(\omega(\epsilon) + \alpha^{-1}\omega(\alpha) + \omega(\epsilon)(1 + \epsilon^{-\frac{1}{2}}\omega(\epsilon))) + C\omega(\epsilon). \end{aligned}$$

Hence the result follows by taking the limit, in this order, $\kappa \rightarrow 0, \alpha \rightarrow 0, \eta \rightarrow 0$ and then $\epsilon \rightarrow 0$. $\square$

ACKNOWLEDGMENTS

The authors have been partially supported by the Chair FDD/FIME (Institut Louis Bachelier).

REFERENCES

- [1] Aurélien Alfonsi and Benjamin Jourdain. Squared quadratic wasserstein distance: optimal couplings and lions differentiability. *ESAIM: Probability and Statistics*, 24: 703–717, 2020.
- [2] Erhan Bayraktar, Ibrahim Ekren, and Xin Zhang. Comparison of viscosity solutions for a class of second order pdes on the wasserstein space. *arXiv preprint arXiv:2309.05040*, 2023.
- [3] Charles Bertucci. Monotone solutions for mean field games master equations: continuous state space and common noise. *Communications in Partial Differential Equations*, 48(10-12):1245–1285, 2023.
- [4] Charles Bertucci. Stochastic optimal transport and hamilton-jacobi-bellman equations on the set of probability measures. *arXiv preprint arXiv:2306.04283*, 2023.
- [5] Charles Bertucci, Pierre-Louis Lions, and Panagiotis Souganidis. Optimal control of the dyson equation and large deviations for random matrices. *forthcoming*, 2024.

- [6] Pierre Cardaliaguet and Marc Quincampoix. Deterministic differential games under probability knowledge of initial condition. *International Game Theory Review*, 10(01):1–16, 2008.
- [7] René Carmona and Francois Delarue. Probabilistic theory of mean field games: vol. i, mean field fbsdes, control, and games. *Stochastic Analysis and Applications*. Springer Verlag, 2017.
- [8] Alekos Cecchin and François Delarue. Weak solutions to the master equation of potential mean field games. *arXiv preprint arXiv:2204.04315*, 2022.
- [9] Giovanni Conforti, Richard C Kraaij, and Daniela Tonon. Hamilton–jacobi equations for controlled gradient flows: cylindrical test functions. *arXiv preprint arXiv:2302.06571*, 2023.
- [10] Giovanni Conforti, RICHARD CLEMENS Kraaij, and Daniela Tonon. Hamilton–jacobi equations for controlled gradient flows: the comparison principle. *Journal of Functional Analysis*, 284(9):109853, 2023.
- [11] Giovanni Conforti, Richard C Kraaij, Luca Tamanini, and Daniela Tonon. Hamilton–jacobi equations for wasserstein controlled gradient flows: existence of viscosity solutions. *arXiv preprint arXiv:2401.02240*, 2024.
- [12] Andrea Cosso, Fausto Gozzi, Idris Kharroubi, Huyên Pham, and Mauro Rosestolato. Master bellman equation in the wasserstein space: Uniqueness of viscosity solutions. *Transactions of the American Mathematical Society*, 377(01):31–83, 2024.
- [13] Michael G Crandall, Hitoshi Ishii, and Pierre-Louis Lions. User’s guide to viscosity solutions of second order partial differential equations. *Bulletin of the American mathematical society*, 27(1):1–67, 1992.
- [14] Samuel Daudin and Benjamin Seeger. A comparison principle for semilinear hamilton–jacobi–bellman equations in the wasserstein space. *Calculus of Variations and Partial Differential Equations*, 63(4):1–36, 2024.
- [15] Samuel Daudin, Joe Jackson, and Benjamin Seeger. Well-posedness of hamilton–jacobi equations in the wasserstein space: non-convex hamiltonians and common noise. *arXiv preprint arXiv:2312.02324*, 2023.
- [16] Jin Feng, Toshio Mikami, and Johannes Zimmer. A hamilton–jacobi pde associated with hydrodynamic fluctuations from a nonlinear diffusion equation. *Communications in Mathematical Physics*, 385(1):1–54, 2021.
- [17] Alessio Figalli. On the monge-amp\ere equation. *arXiv preprint arXiv:1805.05028*, 2018.
- [18] Jean-Michel Lasry and Pierre-Louis Lions. A remark on regularization in hilbert spaces. *Israel Journal of Mathematics*, 55:257–266, 1986.
- [19] Pierre-Louis Lions. Cours au college de france. *www.college-de-france.fr*, 2008-2009.
- [20] Robert J McCann. A convexity principle for interacting gases. *Advances in mathematics*, 128(1):153–179, 1997.
- [21] Jean-Jacques Moreau. Semi-continuité du sous-gradient d’une fonctionnelle. *Comptes rendus hebdomadaires des séances de l’Académie des sciences*, 260:1067–1070, 1965.
- [22] Robert R Phelps. *Convex functions, monotone operators and differentiability*, volume 1364. Springer, 2009.

- [23] Filippo Santambrogio. Optimal transport for applied mathematicians. *Birkhäuser, NY*, 55(58-63):94, 2015.
- [24] H Mete Soner and Qinxin Yan. Viscosity solutions of the eikonal equation on the wasserstein space. *Applied Mathematics & Optimization*, 90(1):1, 2024.

^A CMAP, ECOLE POLYTECHNIQUE, UMR 7641, 91120 PALAISEAU, FRANCE,, ^B UNIVERSITÉ PARIS-DAUPHINE, PSL RESEARCH UNIVERSITY, UMR 7534, CEREMADE, 75016 PARIS, FRANCE,, ^C COLLÈGE DE FRANCE, 3 RUE D'ULM, 75005, PARIS, FRANCE.