



The five gradients inequality for non quadratic costs

Thibault Caillet

► To cite this version:

Thibault Caillet. The five gradients inequality for non quadratic costs. Comptes Rendus. Mathématique, 2023, 361, pp. 715-721. $\langle 10.5802/crmath.444 \rangle$. $\langle \text{hal-03812742} \rangle$

HAL Id: hal-03812742

<https://hal.science/hal-03812742v1>

Submitted on 12 Oct 2022

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.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

The five gradients inequality for non quadratic costs

Thibault Caillet*

October 12, 2022

Abstract

We give a proof of the "five gradients inequality" of Optimal Transportation Theory for general costs of the form $c(x, y) = h(x - y)$ where h is a C^1 strictly convex radially symmetric function.

1 Introduction

While its name was popularized later on, the five gradients inequality was introduced in [5] as a way to derive estimates on the gradient of some recurring variational problems in Optimal Transport involving the Wasserstein distance W_2 . In particular the authors showed that the inequality can be used in the celebrated JKO scheme

$$\varrho_{k+1}^\tau \in \operatorname{argmin}_\varrho \int f(\varrho) \, dx + \frac{W_2^2(\varrho, \varrho_k^\tau)}{2\tau},$$

to derive BV estimates that can be iterated along the scheme uniformly in τ , and therefore pass to the limit PDE $\partial_t \varrho - \nabla \cdot (\varrho f''(\varrho) \nabla \varrho) = 0$, yielding that the BV norm of the solution is nonincreasing in time. In the same paper, the authors also use the inequality to prove BV estimates for the Wasserstein projection of a measure with BV density on the set of measures with density bounded by another given BV function. This result is then used in [4] to find bounds on the perimeters of solutions of some variational problems involving mutually singular measures. The five gradients inequality has also been used in [6] to derive Sobolev estimates for the solutions of the JKO scheme for the Fokker-Planck equation. One only needs to prove that the five gradients inequality holds in a more general setting to generalize most of these results beyond the W_2 case. It is now folk-lore that the inequality is also true for the distance W_p for $p > 1$ and more general cost functions, yet a full proof has not been available until now. In this paper we generalize the proof given in [5] to the case where the cost c is of the form $c(x, y) = h(x - y)$ where h is a strictly convex radially symmetric C^1 function. The inequality reads as follows:

Theorem 1.1. *Let $\Omega \subset \mathbb{R}^d$ be bounded and convex with non-empty interior, $\varrho, g \in W^{1,1}(\Omega)$ be two probability densities, $h \in C^1(\mathbb{R}^d)$ a radially symmetric strictly convex function and $H \in C^1(\mathbb{R}^d \setminus \{0\})$*

*Institut Camille Jordan, Université Claude Bernard - Lyon 1; 43 boulevard du 11 novembre 1918, 69622 Villeurbanne cedex
caillet@math.univ-lyon1.fr .

be a radially symmetric convex function, then the following inequality holds

$$\int_{\Omega} (\nabla \varrho \cdot \nabla H(\nabla \varphi) + \nabla g \cdot \nabla H(\nabla \psi)) \, dx \geq 0, \quad (1.1)$$

where (φ, ψ) is a choice of Kantorovich potentials for the optimal transport problem between ϱ and g for the transport cost given by h , with the convention that $\nabla H(0) = 0$.

A particular case is the one where $h(z) = |z|^p$ for $p > 1$, generalizing the inequality to the W_p case. Following the strategy established in [5], the generalized inequality for example implies that the BV norm of the solution of $\partial_t \varrho - \Delta_q(g(\varrho)) = 0$ (with $q = \frac{p}{p-1}$, and g nondecreasing) decreases in time. Indeed, this nonlinear PDE can be seen as the limit of a JKO-like scheme

$$\varrho_{k+1} \in \operatorname{argmin}_{\varrho} \frac{W_p^p(\varrho, \varrho_k)}{p\tau^{p-1}} + \int_{\Omega} f(\varrho_k) \, dx.$$

by choosing $g'(\varrho) = \varrho^{p-1} f''(\varrho)$ with suitable assumptions on f so that the JKO scheme indeed converges (see e.g. [1], [3], [10]).

To justify computations that involve second derivatives of Kantorovich potentials, the original proof made use of the well known Caffarelli regularity theory available for the cost $h(x-y) = |x-y|^2$. Since the works of Ma, Trudinger, Wang [9] and Loeper [8], sufficient and necessary conditions on the cost c to guarantee the existence of smooth potentials are known, and in these cases one could reproduce the proof given in [5]. Unfortunately, these conditions for regularity do not cover the cases where, for example, $h(z) = |z|^p$ for $p \neq 2$. Therefore in the sequel we shall instead approximate the cost with semiconcave cost functions, and use the fact that Kantorovich potentials inherit this semiconcavity, along with Alexandroff's theorem (see e.g. [7]) :

Theorem 1.2 (Alexandroff's theorem). *Let u be a semiconcave function on an open bounded set $A \subset \mathbb{R}^d$; then u is twice differentiable a.e., meaning for a.e. $x_0 \in A$, there exists $p_{x_0} \in \mathbb{R}^d$ and a symmetric matrix B_{x_0} such that*

$$\lim_{x \rightarrow x_0} \frac{u(x) - u(x_0) - p_{x_0} \cdot (x - x_0) + B_{x_0}(x - x_0) \cdot (x - x_0)}{|x - x_0|^2} = 0.$$

Moreover, the gradient of u , defined for almost every x_0 in A is BV and the absolutely continuous part of the second derivative $D_{ac}^2 u(x_0)$ is given by B_{x_0} .

Usual references for the theory of optimal transport that we will use throughout this paper include [12], [13] as well as [11] which also features a chapter dedicated to the JKO scheme for the Fokker-Planck equation.

2 Proof of the inequality

In the sequel, unless otherwise indicated, Ω will denote a bounded convex subset of \mathbb{R}^d with non-empty interior. The weak convergence of measures will be in duality with $C(\bar{\Omega})$, however, we will work with probability measures that have densities and that therefore cannot be concentrated on $\partial\Omega$ which is negligible for the Lebesgue measure because Ω is convex. We take $R > 0$ to be such that $\Omega \subset \bar{B}(R/2)$.

Definition 2.1. We say that a cost $c(x, y) = h(x - y)$ satisfies (H1) if :

1. $h \in C^2(\bar{B}(R))$
2. h is strictly convex
3. h is radially symmetric

Remark 2.2. In particular, a cost satisfying (H1) is semiconcave i.e. there exists $C > 0$ such that $x \mapsto h(x) - C|x|^2$ is concave on $B(R)$. Since it is known that Kantorovich potentials can be taken to be c -concave, i.e. of the form

$$\varphi(x) = \inf_{y \in \Omega} h(x - y) - \psi(y)$$

for some function $\psi : \Omega \rightarrow \bar{\mathbb{R}}$, they can also be asumed to be semiconcave, with the same semiconcavity constant C as h .

In order to deal with regularity issues, we will approximate the cost function with costs satisfying (H1).

Lemma 2.3. *Let $h \in C^1(\mathbb{R}^d)$ be a radially symmetric strictly convex function and η_ε be a radially symmetric mollifier. Then $h_\varepsilon = \eta_\varepsilon * h$ satisfies (H1) and $h_\varepsilon \xrightarrow{C^1(\bar{B}(R))} h$.*

To prove the convergence back to the original problem we will need a few lemmas. We recall that if (X, Σ, μ) is a measure space and f_n, f are measurable functions, we say that f_n converges in μ -measure to f if for every $\varepsilon > 0$

$$\lim_{n \rightarrow +\infty} \mu(\{|f_n - f| > \varepsilon\}) = 0.$$

and we will denote this convergence by $f_n \xrightarrow{\mu} f$.

Lemma 2.4. *Let $\Omega \subset \mathbb{R}^d$ and $T, T_n : \Omega \rightarrow \Omega$ be measurable functions. Let $\varrho \in \mathcal{P}(\Omega)$ then we have*

$$T_n \xrightarrow{\varrho} T \iff (\text{id} \times T_n)_\# \varrho \rightarrow (\text{id} \times T)_\# \varrho \text{ weakly,}$$

where $(\text{id} \times T) : \Omega \rightarrow \Omega$ is the function defined as $(\text{id} \times T)(x) = (x, T(x))$

Proof. First if $T_n \xrightarrow{\varrho} T$, assuming by contradiction that the weak convergence does not hold, there exists $f \in C_b(\Omega \times \Omega)$, $\varepsilon > 0$ and a subsequence $(T_{n_k})_k$ such that

$$\left| \int_{\Omega} f(x, T(x)) - f(x, T_{n_k}(x)) d\varrho(x) \right| \geq \varepsilon.$$

Extracting a further subsequence such that $T_{n_{k_j}}$ converges to T ϱ -a.e. and using dominated convergences gives a contradiction.

Let $\varepsilon > 0$, by Lusin's theorem, for $\delta > 0$, there exists a compact set $K \subset \Omega$ such that $T|_K$ is continuous

and $\varrho(\Omega \setminus K) \leq \delta$. Therefore $A = \{(x, y) \in K \times \Omega, |T(x) - y| \geq \varepsilon\}$ is a closed set in $\Omega \times \Omega$. Since the evaluation on closed sets is upper semi continuous for the weak convergence of measures,

$$\begin{aligned} 0 &= (\text{id} \times T)_\# \varrho(A) \geq \limsup (\text{id} \times T_n)_\# \varrho(A) \\ &= \limsup \varrho(\{x \in K, |T(x) - T_n(x)| \geq \varepsilon\}) \\ &\geq \limsup \varrho(\{x \in \Omega, |T(x) - T_n(x)| \geq \varepsilon\}) - \delta. \end{aligned}$$

Letting $\delta \rightarrow 0$ gives $T_n \xrightarrow{\varrho} T$. \square

Lemma 2.5. *Given metric spaces X and Y , let $f, f_n : X \rightarrow Y$ be bijective functions such that f_n converges uniformly to f . Then $g_n = f_n^{-1}$ converges uniformly to $g = f^{-1}$ if g is uniformly continuous. Furthermore, if $(x_n)_n \in X$ is a sequence such that $f_n(x_n) \rightarrow f(x)$ then $x_n \rightarrow x$.*

Proof. It is straightforward to check that if $w : Y \rightarrow X$ is uniformly continuous and if $h_n : X \rightarrow Y$ converges uniformly to h , then $w \circ h_n$ converges uniformly to $w \circ h$. Therefore to prove that $g_n = g \circ f \circ g_n$ uniformly converges to $g \circ f \circ g$ it is enough to check that $f \circ g_n$ uniformly converges to the identity function. For $y \in Y$, we have that $d(f \circ g_n(y), y) = d(f \circ g_n(y), f_n \circ g_n(y))$ and the uniform convergence of f_n to f concludes.

The second claim follows from the fact that uniform convergence of functions preserves the convergence of sequences of points. \square

Lemma 2.6. *Let $T_n, T : \Omega \rightarrow \mathbb{R}^d$ be uniformly bounded, $1 < p < \infty$ and $\varrho_n, \varrho \in L^1(\Omega)$ such that $\varrho_n \xrightarrow{L^1(\Omega)} \varrho$, $\varrho_n T_n \rightarrow \varrho T$ as measures in $\mathcal{M}(\Omega)$ and $\int_\Omega |T_n|^p \varrho_n \rightarrow \int_\Omega |T|^p \varrho$. Then $T_n \xrightarrow{L^p(\Omega, \varrho)} T$.*

Remark 2.7. The assumptions are satisfied if Ω is bounded and T_n, T take values in Ω , $\varrho_n \xrightarrow{L^1(\Omega)} \varrho$ and $(\text{id} \times T_n)_\# \varrho_n \rightarrow (\text{id} \times T)_\# \varrho$

Proof. Since the T_n are uniformly bounded in L^∞ , up to a subsequence we have weak-* convergence in $L^\infty : T_n \xrightarrow{*} \tilde{T} \in L^\infty(\Omega)$. Since $\varrho_n \rightarrow \varrho$ strongly in L^1 , strong-weak convergence yields convergence in distribution of the product $\varrho_n T_n \rightarrow \varrho \tilde{T}$. By uniqueness of the limit in the space of distributions, the convergence $\varrho_n T_n \rightarrow \varrho T$ therefore gives that $\varrho T = \varrho \tilde{T}$. Then we have :

$$\left| \int_\Omega |T_n|^p \varrho - \int_\Omega |T|^p \varrho \right| \leq \left| \int_\Omega |T_n|^p \varrho - \int_\Omega |T_n|^p \varrho_n \right| + \left| \int_\Omega |T_n|^p \varrho_n - \int_\Omega |T|^p \varrho \right|.$$

The second term goes to 0 by assumption and the first term is bounded by $C^p \|\varrho - \varrho_n\|_1$ which also goes to 0. Now it is enough to prove weak convergence in L^p , so let $f \in L^{p'}(\Omega, \varrho)$. Since by Hölder $f \varrho \in L^1(\Omega)$,

$$\left| \int_\Omega f T \varrho \, dx - \int_\Omega f T_n \varrho \, dx \right| = \left| \int_\Omega (T - T_n) f \varrho \, dx \right| \xrightarrow{n \rightarrow \infty} \left| \int_\Omega (T - \tilde{T}) f \varrho \, dx \right| = 0.$$

Hence $T_n \xrightarrow{L^p(\Omega, \varrho)} T$, and by uniqueness of this limit the whole sequence actually converges. \square

We are now ready to begin the proof of the five gradients inequality with smooth densities and a cost satisfying (H1)

Lemma 2.8. *Given h satisfying (H1), ϱ and g smooth probability densities, let $H \in C^2(\mathbb{R}^d)$ be a convex function, then the following inequality holds :*

$$\int_{\Omega} (\varrho \nabla \cdot [\nabla H(\nabla \varphi)] - g \nabla \cdot [\nabla H(-\nabla \psi)]) dx \leq 0, \quad (2.1)$$

where (φ, ψ) is a choice of Kantorovich potentials for the optimal transport problem between ϱ and g for the transport cost given by h , and $\nabla \cdot$ denotes the distributional divergence.

Proof. Let (φ, ψ) be a choice of c-concave potentials for the transport problem. From c-concavity we deduce that φ and ψ are semiconcave and hence by Alexandroff's theorem (1.2), φ and ψ are twice differentiable almost everywhere and $\nabla \varphi, \nabla \psi$ are functions of bounded variation on Ω .

We will denote by J_{φ} the set of approximate jump points of $\nabla \varphi$, $\nabla \varphi^+$ and $\nabla \varphi^-$ the left and right approximations of $\nabla \varphi$ and ν_{φ} the approximate normal to J_{φ} . We will denote by $D_{ac}^2 \varphi$ and $D_c^2 \varphi$ respectively the absolutely continuous part and the Cantor part of $D^2 \varphi$. We will now use the chain rule for BV functions (see [2] for more details on notation and precise statements).

$$D[\nabla H(\nabla \varphi)] = D^2 H(\nabla \varphi) D_{ac}^2 \varphi \mathcal{L}^d + D^2 H(\nabla \varphi) D_c^2 \varphi + (\nabla H(\nabla \varphi^+) - \nabla H(\nabla \varphi^-)) \otimes \nu_{\varphi} \mathcal{H}^{d-1} \llcorner J_{\varphi}.$$

Using the semiconcavity of φ , we deduce that there exists some $C > 0$ such that, as a measure, $D^2 \varphi \leq C \mathcal{L}^d$, and we have

$$\begin{aligned} D_{ac}^2 \varphi &\leq C \mathcal{L}^d \text{ a.e.}, \\ D_c^2 \varphi &\leq 0, \\ (\nabla \varphi^+ - \nabla \varphi^-) \otimes \nu_{\varphi} \mathcal{H}^{d-1} \llcorner J_{\varphi} &\leq 0. \end{aligned}$$

In particular since $a \otimes b \leq 0$ implies the existence of $\lambda \leq 0$ such that $a = \lambda b$, using the convexity of H we get

$$(\nabla H(\nabla \varphi^+) - \nabla H(\nabla \varphi^-)) \cdot (\nabla \varphi^+ - \nabla \varphi^-) \geq 0,$$

and therefore

$$(\nabla H(\nabla \varphi^+) - \nabla H(\nabla \varphi^-)) \cdot \nu_{\varphi} \leq 0 \quad \mathcal{H}^{d-1} \llcorner J_{\varphi} \text{ ae.}$$

Using the nonnegativity of $D^2 H$ we therefore have

$$\begin{aligned} \nabla \cdot [\nabla H(\nabla \varphi)] &= \text{Tr}(D^2 H(\nabla \varphi) D_{ac}^2 \varphi) \\ &\quad + \text{Tr}(D^2 H(\nabla \varphi) D_c^2 \varphi) \\ &\quad + (\nabla H(\nabla \varphi^+) - \nabla H(\nabla \varphi^-)) \cdot \nu_{\varphi} \mathcal{H}^{d-1} \llcorner J_{\varphi} \\ &\leq \text{Tr}(D^2 H(\nabla \varphi) D_{ac}^2 \varphi). \end{aligned}$$

Using the same arguments we also get

$$\nabla \cdot [\nabla H(-\nabla \psi)] \geq \text{Tr}(D^2 H(-\nabla \psi) D_{ac}^2 \psi).$$

Integrating with respect to non negative densities ϱ, g , we get

$$\begin{aligned}
& \int_{\Omega} (\varrho \nabla \cdot [\nabla H(\nabla \varphi)] - g \nabla \cdot [\nabla H(-\nabla \psi)]) \, dx \\
& \leq \int_{\Omega} (\varrho \operatorname{Tr}(D^2 H(\nabla \varphi) D_{ac}^2 \varphi) - g \operatorname{Tr}(D^2 H(-\nabla \psi) D_{ac}^2 \psi)) \, dx \\
& = \int_{\Omega} (\varrho \operatorname{Tr}(D^2 H(\nabla \varphi) D_{ac}^2 \varphi) - \varrho [\operatorname{Tr}(D^2 H(-\nabla \psi) D_{ac}^2 \psi) \circ T]) \, dx,
\end{aligned}$$

where T is the optimal transport map between ϱ and g . Let S be the optimal transport between g and ϱ (for the cost $\tilde{c}(x, y) = c(y, x)$), then $S \circ T = \operatorname{id}$ ϱ -a.e., and therefore for ϱ -a.e. x we have

$$\begin{aligned}
-\nabla \psi(T(x)) &= \nabla h(S \circ T(x) - T(x)) \\
&= \nabla h(x - T(x)) \\
&= \nabla \varphi(x).
\end{aligned}$$

We know that φ, ψ are twice differentiable \mathcal{L}^d -a.e. hence

$$\begin{aligned}
1 &= g(\{y, \psi \text{ is twice differentiable at } y\}) \\
&= \varrho(\{x, \psi \text{ is twice differentiable at } T(x)\}).
\end{aligned}$$

and since the function $v \mapsto \varphi(x + v) + \psi(T(x) - v) \leq h(x - y)$ is maximal at $v = 0$ for ϱ -a.e. x , using that, for BV functions, the absolutely continuous part of the derivative coincides with the pointwise derivative, we have $D_{ac}^2 \varphi(x) + D_{ac}^2 \psi(T(x)) \leq 0$ ϱ -a.e. Therefore

$$\begin{aligned}
& \int_{\Omega} (\varrho \nabla \cdot [\nabla H(\nabla \varphi)] - g \nabla \cdot [\nabla H(-\nabla \psi)]) \, dx \\
& \leq \int_{\Omega} (\varrho \operatorname{Tr}(D^2 H(\nabla \varphi) [D_{ac}^2 \varphi + D_{ac}^2 \psi \circ T]) \, dx \leq 0. \quad \square
\end{aligned}$$

Theorem 2.9. *Given h satisfying (H1), suppose that ϱ, g are smooth probability densities, and let $H \in C^2(\mathbb{R}^d)$ be a convex radially symmetric function, then the following inequality holds*

$$\int_{\Omega} (\nabla \varrho \cdot \nabla H(\nabla \varphi) + \nabla g \cdot \nabla H(\nabla \psi)) \, dx \geq 0,$$

where (φ, ψ) is a choice of Kantorovich potentials for the optimal transport problem between ϱ and g for the transport cost given by h .

Proof. First since H is radially symmetric we have $\nabla H(-\nabla \psi) = -\nabla H(\nabla \psi)$ Using Lemma 2.8 and

the trace theorem for BV functions we get

$$\begin{aligned}
& \int_{\Omega} (\nabla \varrho \cdot \nabla H(\nabla \varphi) + \nabla g \cdot \nabla H(\nabla \psi)) \, dx \\
&= - \int_{\Omega} (\varrho \nabla \cdot [\nabla H(\nabla \varphi)] + g \nabla \cdot [\nabla H(\nabla \psi)]) \, dx \\
&+ \int_{\partial \Omega} (\varrho \nabla H(\nabla \varphi) \cdot n + g \nabla H(\nabla \psi) \cdot n) \, d\mathcal{H}^{d-1} \\
&\geq \int_{\partial \Omega} (\varrho \nabla H(\nabla \varphi) \cdot n + g \nabla H(\nabla \psi) \cdot n) \, d\mathcal{H}^{d-1}.
\end{aligned}$$

Let $x_0 \in \partial \Omega$ be such that $\varrho(x_0) > 0$. By continuity of ϱ , for small $\varepsilon > 0$, there exists $r_0 > 0$ such that if $x \in B(x_0, r_0) \cap \Omega$, then $\varrho(x) > \varepsilon$. For a.e. $x \in B(x_0, r_0) \cap \Omega$, since $T(x) \in \Omega$ we have $(T(x) - x_0) \cdot n(x_0) \leq 0$ and therefore $(x - \nabla h^*(\nabla \varphi)) \cdot n(x_0) \leq x_0 \cdot n(x_0)$. For $r \leq r_0$, integrating on $B(x_0, r)$ we get

$$\int_{B(x_0, r) \cap \Omega} (x - \nabla h^*(\nabla \varphi)) \cdot n(x_0) \, dx \leq \int_{B(x_0, r) \cap \Omega} x_0 \cdot n(x_0) \, dx.$$

Taking $r \rightarrow 0$ we get $\nabla h^*(\nabla \varphi(x_0)) \cdot n(x_0) \geq 0$ for \mathcal{H}^{d-1} -a.e. $x_0 \in \partial \Omega$ such that $\varrho(x_0) > 0$.

Since H and h are convex and radially symmetric we have $\nabla H(z) = \alpha(z)z$ with $\alpha(z) \geq 0$ and since $\nabla h^* = (\nabla h)^{-1}$ we also have $\nabla h^*(z) = \alpha'(z)z$ hence for \mathcal{H}^{d-1} -a.e. $x_0 \in \partial \Omega$ such that $\varrho(x_0) > 0$ we have $\nabla H(\nabla \varphi(x_0)) \cdot n(x_0) \geq 0$. By the same arguments for g and $\nabla \psi$ we have

$$\int_{\partial \Omega} (\varrho \nabla H(\nabla \varphi) \cdot n + g \nabla H(\nabla \psi) \cdot n) \, d\mathcal{H}^{d-1} \geq 0,$$

which concludes the proof. \square

Remark 2.10. In the above proof the hypothesis that h and H are radially symmetric is only used in proving that ∇H and ∇h^* point in the same direction, to deal with the boundary integral. One therefore should be able to only assume that $H(z) = \alpha(h^*(z))$ for some increasing convex function α to get $\nabla H(z) = \alpha'(h^*(z))\nabla h^*(z)$. This however adds the difficulty of finding an adequate replacement for Lemma 2.3 and proving the ensuing convergence.

Proof of Theorem 1.1.

Proof. We begin by taking $H \in C^2(\mathbb{R}^d)$, smooth densities ϱ, g , and approximating the cost h using Lemma 2.3 and applying Theorem 2.9. Since $h_\varepsilon \xrightarrow{C^1} h$, by compactness of $\mathcal{P}(\Omega)$ the sequence $(\gamma_\varepsilon)_\varepsilon = ((\text{id} \times T_\varepsilon)_\# \varrho)_\varepsilon$ of optimal plans weakly converges up to a subsequence to $\gamma = (\text{id} \times T)_\# \varrho \in \mathcal{P}(\Omega)$ which is optimal for h . By uniqueness of such transport maps in fact the whole sequence converges and by Lemma 2.4 we have $T_\varepsilon \xrightarrow{\mathcal{L}} T$. Extracting a subsequence we therefore have $T_\varepsilon \rightarrow T$ ϱ -a.e, using that we have $T_\varepsilon(x) = x - \nabla h_\varepsilon^*(\nabla \varphi_\varepsilon(x))$ and $T(x) = x - \nabla h^*(\nabla \varphi(x))$ and Lemma 2.5 we get that $\nabla \varphi_\varepsilon(x) \rightarrow \nabla \varphi(x)$ ϱ -a.e. Doing the same for g and ψ and using that $\nabla H \in L^\infty$, the result is proved

by dominated convergence.

Next we take $\varrho, g \in W^{1,1}(\Omega)$, and for example by convolution we can find smooth functions $\varrho_\varepsilon, g_\varepsilon$ such that $(\varrho_\varepsilon, g_\varepsilon) \xrightarrow{W^{1,1}(\Omega)} (\varrho, g)$. Extracting a subsequence if necessary, we can assume the convergence to hold a.e. and to have a uniform L^1 domination. Normalizing $(\varrho_\varepsilon, g_\varepsilon)$ if necessary we can assume they are probability densities and apply the previous result to get

$$\int_{\Omega} (\nabla \varrho_\varepsilon \cdot \nabla H(\nabla \varphi_\varepsilon) + \nabla g_\varepsilon \cdot \nabla H(\nabla \psi_\varepsilon)) \, dx \geq 0. \quad (2.2)$$

Using Lemma 2.6 and up to a subsequence, we have $T_\varepsilon \rightarrow T$ ϱ -a.e. which implies just as before that $\nabla \varphi_\varepsilon \rightarrow \nabla \varphi$ ϱ -a.e. Therefore for a.e. $x \in \{\varrho > 0\}$, we have $\nabla \varphi_\varepsilon(x) \rightarrow \nabla \varphi(x)$ and, using $\nabla \varrho = 0$ a.e. on $\{\varrho = 0\}$, for a.e. $x \in \{\varrho = 0\}$, we have $\nabla \varrho_\varepsilon(x) \rightarrow \nabla \varrho(x)$. In any case, since $\nabla H \in L^\infty$, we have that $\nabla \varrho_\varepsilon(x) \cdot \nabla H(\nabla \varphi_\varepsilon(x)) \rightarrow \nabla \varrho(x) \cdot \nabla H(\nabla \varphi(x))$ for a.e. $x \in \Omega$ and dominated convergence gives

$$\int_{\Omega} \nabla \varrho_\varepsilon \cdot \nabla H(\nabla \varphi_\varepsilon) \, dx \xrightarrow{\varepsilon \rightarrow 0} \int_{\Omega} \nabla \varrho \cdot \nabla H(\nabla \varphi) \, dx.$$

Doing the same for g and ψ and adding the integrals gives the result.

Finally, approximating a radially symmetric convex function $H \in C^1(\mathbb{R}^d \setminus \{0\})$ with C^2 functions with the same properties proves that the result stays true in this case, when using the natural convention $\nabla H(0) = 0$. \square

References

- [1] M. AGUEH, Existence of solutions to degenerate parabolic equations via the Monge-Kantorovich theory. *Advances in Differential Equations* 10.3, p309-360 (2005).
- [2] L. AMBROSIO, N. FUSCO, D. PALLARA, *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford University Press, Oxford (2000).
- [3] L. AMBROSIO, N. GIGLI, G. SAVARÉ, *Gradient flows in metric spaces and in the spaces of probability measures*. Lectures in Mathematics, ETH Zurich, Birkhäuser (2005).
- [4] G. BUTTAZZO, G. CARLIER, M. LABORDE, On the Wasserstein distance between mutually singular measures. *Advances in Calculus of Variations*, vol. 13, no. 2, pp. 141-154 (2020).
- [5] G. DE PHILIPPIS, A.R. MÉSZÁROS, F. SANTAMBROGIO, B. VELICHKOV, BV estimates in Optimal Transportation and Applications. *Archive for Rational Mechanics and Analysis* 219.2, p. 829-860 (2020).
- [6] S. DI MARINO, F. SANTAMBROGIO, JKO estimates in linear and non-linear Fokker-Planck equations, and Keller-Segel: L^p and Sobolev bounds. *Annales de l'Institut Henri Poincaré C, Analyse Non Linéaire* (2021).

- [7] L.C. EVANS, R.F. GARIEPY, *Measure Theory and Fine Properties of Functions*. Studies in Advanced Mathematics CRC, Boca Raton (1992).
- [8] G. LOEPER, On the regularity of solutions of optimal transportation problems. *Acta Math.* 202, p.241-283 (2009).
- [9] X.-N. MA, N.S. TRUDINGER, X.-J. WANG, Regularity of potentials functions of the optimal transportation problem. *Arch. Ration. Mech. Anal.* 117(2), P.151-183 (2005)
- [10] F. OTTO, Doubly Degenerate Diffusion Equations as Steepest Descent. *Unpublished* (2000)
- [11] F. SANTAMBROGIO, *Optimal Transport for Applied Mathematicians*, Progress in Nonlinear Differential Equations and Their Applications no 87, Birkhäuser Basel (2015).
- [12] C. VILLANI, *Topics in Optimal Transportation*. Graduate Studies in Mathematics (American Mathematical Society, Providence, 2003).
- [13] C. VILLANI, *Optimal Transport, Old and New*, Grundlehren der mathematischen Wissenschaften, Vol. 338, Springer, 2009.