Non-smooth differential properties of optimal transport

Dario Cordero-Erausquin

1. Introduction

The usual optimal transport (for the quadratic cost $c(z) = |z|^2/2$) characterized by Brenier [**2**] as the gradient of a convex function is a very useful tool in geometric analysis for proving functional inequalities (see for instance [**10, 4**]). During the Azorean meeting, Luis Caffarelli raised the following natural question: how and where can we use optimal transport for other strictly convex costs? The first problem one encounters is the lack of regularity results when the cost is not quadratic. On the other hand, almost everywhere results are often sufficient for proving inequalities. The aim of the present note is only to explain how, under some (still restrictive) assumptions on the cost c, such non-smooth almost everywhere results hold. Most of (if not all) the results mentioned here are maybe known to mass transportation specialists since the method has its origin in the monotone change of variable theory of McCann [**10**]. More precisely, we closely follow the proofs given in the work by Cordero-Erausquin, McCann and Schmuckenschläger [**6**] where the monotone change of variable theory is extended to the Riemannian setting. And for economy and simplicity, those parts of the proof which carry over directly from [**6**] have been reproduced more or less verbatim where possible.

Let $T(x) = x - \nabla c^*(\nabla\phi(x))$ be the optimal transport for some strictly convex cost c, pushing $f(x)\,d\mathrm{vol}(x)$ forward to $g(y)\,d\mathrm{vol}(y)$, where vol denotes the Lebesgue measure on $\mathbf{R}^n$ (see definitions below). The goal of this note is to give a precise meaning and to prove the following Jacobian equation

$$f(x) = g(T(x))\,\det(I - \mathrm{Hess}_{\nabla\phi(x)}c^*\,\mathrm{Hess}_x\phi).$$

We also show that the term which plays the role of the differential of T in this equation satisfies an integration by parts inequality. This is due to the fact that the Hessian in the sense of Aleksandrov of a convex function is dominated by the distributional second derivative. Such an inequality is often sufficient for applications. Independently, we present some properties of the interpolating map $T_t(x) := x - t\nabla c^*(\nabla\phi(x))$ and of the interpolated density, which are relevant for the study of the displacement convexity phenomenon introduced by McCann [**10**].

In the next section we recall some standard definitions and properties of set-valued maps. Indeed, it will be more convenient to see the almost everywhere defined optimal map T as an everywhere defined set valued map. In section §3 we recall the relevant facts on c-concavity and on mass transportation theory. Section §4 contains the main result on the Jacobian equation and the integration by parts inequality. The final section §5 is devoted to the study of the interpolation along mass transport. In particular we show that the interpolated measure between two densities is itself absolutely continuous with respect to the Lebesgue measure.

We would like to thank the hospitality of the Georgia Institute of Technology, and our hosts Wilfrid Gangbo and Robert J. McCann for stimulating discussions on the topic.

2. Set-valued maps

We start by recalling some vocabulary:

(1) A *set-valued* map $T : \mathbf{R}^n \longrightarrow \mathbf{R}^n$ is an application from $\mathbf{R}^n$ to the subsets of $\mathbf{R}^n$.
(2) The *definition domain* of a set-valued map T is the set
$$\mathcal{D}(T) := \{x \in \mathbf{R}^n \,;\ T(x) \neq \emptyset\}.$$
(we may also write $T : \mathcal{D}(T) =: X \longrightarrow Y := T(X) = \bigcup_{x\in X} T(x)$).
(3) If $T : \mathcal{D}(T) =: X \longrightarrow Y := T(X)$ is a set-valued map, the inverse of T is the set-valued map $T^{-1} : \mathcal{D}(T^{-1}) = T(X) := Y \longrightarrow X = \mathcal{D}(T)$ defined, for every $y \in Y$, by
$$T^{-1}(y) = \{x \in X \,;\ y \in T(x)\}.$$
(4) The *unicity domain* of a set-valued map T is the subset of the definition domain given by
$$dom(T) := \{x \in \mathcal{D}(T) \,;\ T(x) \text{ is a single point}\}.$$
Thus $T : dom(T) \longrightarrow \mathbf{R}^n$ is a classical map.

Definition 2.1 (Differential). *Let T be a set-valued map and $x \in int(\mathcal{D}(T))$ a point in the interior of $\mathcal{D}(T)$. We say that T is differentiable at x if $\underline{x \in dom(T)}$ and if there exists a linear map L such that*
$$\sup_{y\in T(x+u)} |y - T(x) - Lu| = o(|u|). \tag{1}$$
The linear map L is called the differential of T at $x \in dom(T)$ and is denoted by dT_x.

For consistency, it is worth noticing that, when existing, the differential L is unique, since for any $y_n \in T(x + u/n)$ one has $y_n = T(x) + \frac{1}{n}Lu + o\left(\frac{1}{n}\right)$. In the same vein one can observe that:

Remark 2.2 (Differentiability implies "continuity"). *With the same notations as in the previous definition, if $x_n \to x$ and $y_n \in T(x_n)$, then $y_n \to T(x)$ as soon as T is differentiable at x.*

Lemma 2.3 (Inclusion and chain rule). *Let T, F be two set-valued maps on $\mathbf{R}^n$.*
1) Assume that for every $x \in \mathbf{R}^n$ one has $T(x) \subset F(x)$. If $x \in int(\mathcal{D}(T))$ is a point where F is differentiable, then T is differentiable at x and $dT_x = dF_x$.
2) Let $x \in int(\mathcal{D}(T))$ such that T is differentiable at x and F is differentiable at $T(x) \in int(\mathcal{D}(F))$. Then $F \circ T$ is differentiable at x and $d(F \circ T)_x = dF_{T(x)} \circ dT_x$.

The proof of the previous lemma is straightforward.

Lemma 2.4 (Inverse map property). *Let $T : \mathbf{R}^n \to \mathbf{R}^n$ be a set-valued map. Fix $x \in int(\mathcal{D}(T))$ a point where T has a differential and such that T^{-1} has a differential at $T(x) \in int(\mathcal{D}(T^{-1}))$. Then dT_x is invertible and $(dT_x)^{-1} = d(T^{-1})_{T(x)}$.*

Proof. Set $y := T(x)$ and fix $u \in \mathbf{R}^n$. Differentiability of T at x implies, for $y_n \in T(x + u/n)$ and $n \to \infty$: $y_n - y = dT_x(u)/n + o(1/n)$. Since $x + u/n \in T^{-1}(y_n)$ one has by differentiability of T^{-1} at y: $x + u/n = x - d(T^{-1})_{T(x)}(y - y_n) + o(|y - y_n|)$. Combining the two equalities we get the assertions of the lemma. □

Proposition 2.5 (Equivalence of geometric and algebraic Jacobians). *Let $T : \mathbf{R}^n \to \mathbf{R}^n$ be a set-valued map. Fix $x \in int(\mathcal{D}(T))$ a point where T has a differential and such that T^{-1} has a differential at $T(x) \in int(\mathcal{D}(T^{-1}))$. Then, when $r \to 0$, the set $T(B_r(x))$ shrinks nicely to $\{T(x)\}$ and*
$$\lim_{r\to 0} \frac{\mathrm{vol}[T(B_r(x))]}{\mathrm{vol}[B_r(x)]} = \det dT_x \neq 0.$$

Proof. As in [**10**, Proposition A2], set $c_1 := \|dT_x\|$ and $c_2 := \|(dT_x)^{-1}\| = \|d(T^{-1})_{T(x)}\|$. It is enough to prove that for every $\varepsilon > 0$ there exists $\delta > 0$ such that for every $r < \delta$ one has
$$T(x) + (1 + \varepsilon c_1)^{-1} dT_x(B_r(x)) \subset T\left(B_r(x)\right) \subset T(x) + (1 + \varepsilon c_2) dT_x(B_r(x)). \tag{2}$$
This follows from the definition (1) of the differential of T at x and of T^{-1} at $T(x)$. □

3. Background on optimal transport

We recall the results of Gangbo and McCann [**8, 9**] and Caffarelli [**3**] on optimal transport. Notations are taken from [**6**]. We will deal with compactly supported Borel probability measures μ, ν on $\mathbf{R}^n$. In this case, one can drop the technical assumptions on the growth at infinity of the cost c.

DEFINITION 3.1 (c-transforms and the subset $\mathcal{I}^c(X,Y)$ of c-concave functions). *Fix a convex function $c : \mathbf{R}^n \to \mathbf{R}_+$. Let X and Y be two compact subsets of $\mathbf{R}^n$. The set $\mathcal{I}^c(X,Y)$ of c-concave functions (relative to X and Y) is the set of functions $\phi : X \to \mathbf{R} \cup \{-\infty\}$ not identically $-\infty$, for which there exists a function $\psi : Y \to \mathbf{R} \cup \{-\infty\}$ such that*

$$\phi(x) = \inf_{y \in Y} c(x-y) - \psi(y) \qquad \forall x \in X. \tag{3}$$

We refer to ϕ as the c-transform *of ψ and abbreviate (3) by writing $\phi = \psi^c$. Similarly, given $\phi \in \mathcal{I}^c(X,Y)$, we define its* c-transform *$\phi^c \in \mathcal{I}^c(Y,X)$ by*

$$\phi^c(y) := \inf_{x \in X} c(x-y) - \phi(x) \qquad \forall y \in Y. \tag{4}$$

For $\phi \in \mathcal{I}^c(X,Y)$, it follows easily from (4) that

$$\phi(x) = \inf_{y \in Y} c(x-y) - \phi^c(y) \qquad \forall x \in X, \tag{5}$$

which we abbreviate by writing $\phi^{cc} = \phi$. Lipschitz continuity of ϕ^c follows from compactness of X and from the locally Lipschitz behavior of $c(x-y)$, whether or not $\phi : X \longrightarrow \mathbf{R} \cup \{-\infty\}$ is continuous. Thus it costs no generality to assume ψ and ϕ are continuous and real-valued in definition (3), in view of (5).

For two Borel measures μ and ν on $\mathbf{R}^n$, we say a map $T : \mathbf{R}^n \longrightarrow \mathbf{R}^n$ defined μ-a.e. *pushes μ forward to ν* (or *transports* μ onto ν) if ν is the image measure of μ under T, denoted $\nu = T_{\#}\mu$ where $(T_{\#}\mu)(B) := \mu(T^{-1}(B))$ for all Borel sets $B \subset \mathbf{R}^n$. The definition of the push-forward $T_{\#}\mu$ can equally well be expressed by asserting

$$\int b(y)\, d(T_{\#}\mu)(y) = \int b(T(x))\, d\mu(x) \tag{6}$$

for all Borel functions $b : \mathbf{R}^n \to \mathbf{R}_+$.

THEOREM 3.2 (Optimal mass transport for strictly convex costs [**8, 9, 3**]). *Fix a strictly convex function $c : \mathbf{R}^n \to \mathbf{R}_+$. Let $\mu \ll \mathrm{vol}$ and ν be two compactly supported Borel probability measures on $\mathbf{R}^n$, and choose two compact sets X and $Y \subset \mathbf{R}^n$ containing the support of μ and ν, respectively. Then there exists $\phi \in \mathcal{I}^c(X,Y)$ such that the Borel map*

$$T(x) := x - \nabla c^*(\nabla \phi(x)) \tag{7}$$

pushes μ forward to ν. This map is uniquely characterized among all maps pushing μ forward to ν by formula (7) with $\phi \in \mathcal{I}^c(X,Y)$. Furthermore T is the unique minimizer of the c-cost $\int c(x - G(x))\, d\mu(x)$ among all Borel maps $G : \mathbf{R}^n \to \mathbf{R}^n$ pushing μ forward to ν (apart from variations on sets of μ-measure zero).

Let us mention that for a set-valued map T on $\mathbf{R}^n$ one can equally well define the push-forward of a measure μ by $(T_{\#}\mu)(B) := \mu(T^{-1}(B))$ for all Borel sets $B \subset \mathbf{R}^n$. As a matter of fact, it will be convenient to see the optimal map T defined almost everywhere by (7) as a set-valued map. To this aim we introduce

DEFINITION 3.3 (c-superdifferential $\partial^c\phi$). *Let X, Y be two compact sets of $\mathbf{R}^n$. For $\phi \in \mathcal{I}^c(X,Y)$ and $x \in X$, the* c-superdifferential *of ϕ at x is the non-empty set*

$$\begin{aligned} \partial^c\phi(x) &:= \{y \in Y \mid \phi(x) + \phi^c(y) = c(x-y)\} && (8)\\ &= \{y \in Y \mid \phi(z) \le \phi(x) + c(z-y) - c(x-y) \ \ \forall z \in X\}. && (9) \end{aligned}$$

Remark 3.4. *The map $\partial^c\phi$ can be seen as a set-valued map on $\mathbf{R}^n$ by extending the definition in the following natural way: first one extends ϕ using (5) and then $\partial^c\phi$ using (8). However, we shall need the set-valued map $\partial^c\phi$ on X only and thus we shall write $\partial^c\phi : X \longrightarrow Y$. It is important to note that since ϕ and c are continuous and Y is compact, we have*

$$\forall x \in X, \quad \partial^c\phi(x) \neq \emptyset \tag{10}$$

Remark 3.5 (c-duality). *Observe that for a c-concave function $\phi \in \mathcal{I}^c(X,Y)$, one has the following duality:*

$$\forall (x,y) \in X \times Y, \quad y \in \partial^c\phi(x) \Longleftrightarrow x \in \partial^c\phi^c(y). \tag{11}$$

In particular we have $\partial^c\phi(X) = Y$. We shall summarize this by writing $(\partial^c\phi)^{-1} = \partial^c\phi^c : Y \longrightarrow X$.

Lemma 3.6 (Multivalued point on view on transport). *Let $\mathcal{X}, Y$ be an open relatively compact and a compact subset of $\mathbf{R}^n$, respectively. Fix a strictly convex function $c : \mathbf{R}^n \to \mathbf{R}_+$ and $\phi \in \mathcal{I}^c(\overline{\mathcal{X}}, Y)$, and denote by $\mathcal{D}^1(\phi)$ the subset of $\mathcal{X}$ where the Lipschitz map ϕ is differentiable. By Rademacher theorem, we know that* $\mathrm{vol}(\mathcal{X}) = \mathrm{vol}(\mathcal{D}^1(\phi))$. *Let T be the Borel map defined on $\mathcal{D}^1(\phi)$ by (7). Then*

1) The set-valued map $\partial^c\phi : X \longrightarrow Y$ is such that $\mathcal{D}^1(\phi) \subset dom(\partial^c\phi)$ and for every $x \in \mathcal{D}^1(\phi)$,

$$\{\partial^c\phi(x)\} = \{T(x)\}.$$

2) If μ is a Borel measure absolutely continuous with respect to vol *and compactly supported inside $\mathcal{X}$, then $T_\#\mu = (\partial^c\phi)_\#\mu$. In fact, for every Borel $B \subset \mathbf{R}^n$, in view of (11) one has*

$$T_\#\mu(B) = \mu(\partial^c\phi^c(B)).$$

Proof. To prove 1) take $x \in \mathcal{D}^1(\phi)$ and, in view of (10), $y \in \partial^c\phi(x)$. Then the continuous function $h(z) := \phi(z) - c(z-y)$ achieves a maximum at $z = x$ and thus admits 0 as a *supergradient* (see below). This means that $\nabla\phi(x)$ is a subgradient of the strictly convex function $z \to c(z-y)$ at x and so $\nabla c^*(\nabla\phi(x)) = x - y$. Thus we get $y = x - \nabla c^*(\nabla\phi(x)) = T(x)$. To prove 2) take $A \subset \mathbf{R}^n$ and use that $\mu(\mathcal{D}^1(\phi)) = 1$ and 1) to get

$$\mu((\partial^c\phi)^{-1}(A)) = \mu\{x \in \mathcal{D}^1(\phi)\,;\ \{T(x)\} = \partial^c\phi(x)) \subset A\} = \mu(T^{-1}(A)).$$

□

Remark 3.7 (Convention). *The previous result says $T = \partial^c\phi$ μ-a.e. and so in the sequel we will say that $T = \partial^c\phi$ is the c-optimal transport, keeping in mind that T is rather seen as a Borel map defined μ-a.e. by (7) and $\partial^c\phi$ is a set-valued map.*

4. Jacobian equation for optimal transport

Notation: We denote by Ω an open set of $\mathbf{R}^n$. We will always assume that the convex cost function c is non-negative.

Definition 4.1 (Superdifferential). *Let $\phi : \Omega \to \mathbf{R}$ be a continuous function. The* superdifferential *of ϕ is the set-valued map $\partial\phi : \Omega \to \mathbf{R}^n$ defined for $x \in \Omega$, by,*

$$\partial\phi(x) := \{v \in \mathbf{R}^n\,;\ \phi(z) \leq \langle z - x, v\rangle + \phi(x) + o(|z-x|)\,\}.$$

An element of $\partial\phi(x)$ is a supergradient *of ϕ at x.*

Definition 4.2 (Semi-concavity). *A function $\phi : \Omega \to \mathbf{R}$ is* semi-concave *at $x \in \Omega$ if there exists a ball $B_r(x)$ and a smooth function $V : B_r(x) \to \mathbf{R}$ such that $\phi + V$ is concave throughout $B_r(x)$. The function ϕ is semi-concave on Ω if it is semi-concave at each point of Ω.*

One can always takes for V the function $V(z) := -C|z-x|^2$ of $V(z) := -C|z|^2$ for some non-negative constant C.

Remark 4.3 (Semi-concave functions have non-empty superdifferential). *The property of having non-empty superdifferentials extends immediately from concave functions to semi-concave functions. In other words, if $\phi : \Omega \to R$ is semi-concave on Ω, one has $\mathcal{D}(\partial\phi) = \Omega$.*

Now, we state the classical result of Aleksandrov (see [**1**] for a modern proof).

THEOREM 4.4 (Aleksandrov's Hessian). *Let $\phi : \Omega \to \mathbf{R}$ be a semi-concave function on Ω. Then the multivalued map $\partial\phi : \Omega \to \mathbf{R}^n$ has a differential (1) almost everywhere on Ω. The differential of $\partial\phi$ is called Hessian of ϕ (in the sense of Aleksandrov) and where it exists, it is a self-adjoint operator denoted by*

$$\mathrm{Hess}_x\phi := d(\partial\phi)_x. \tag{12}$$

Remember that the existence of a derivative for $\partial\phi$ at x of course supposes that $x \in dom(\partial\phi)$, ie: ϕ is differentiable at x. And since ϕ is locally Lipschitz, it is differentiable almost everywhere on Ω, which is consistent with Aleksandrov theorem. The next result is due to Gangbo and McCann.

LEMMA 4.5 (Semi-concavity of c-concave functions for $\mathcal{C}^2$ costs [**9**]). . *Let $\mathcal{X}$ and Y be an open relatively compact and a compact subset of $\mathbf{R}^n$, respectively. Fix a $\mathcal{C}^2$ convex function c and $\phi \in \mathcal{I}^c(\overline{\mathcal{X}}, Y)$. The function ϕ is semi-concave on $\mathcal{X}$ (and hence $\partial\phi$ is almost everywhere differentiable on $\mathcal{X}$).*

We now give an improvement of lemma 3.6.

LEMMA 4.6 (c-superdifferential versus superdifferential). *Fix a strictly convex function c on $\mathbf{R}^n$. Let $\mathcal{X}$ and Y be an open relatively compact and a compact subset of $\mathbf{R}^n$, respectively, and $\phi \in \mathcal{I}^c(\overline{\mathcal{X}}, Y)$. For every $x \in \mathcal{X}$ one has*

$$\partial^c\phi(x) \subset x - \nabla c^*(\partial\phi(x)).$$

PROOF. Take $y \in \partial^c\phi(x)$ and set $h(z) := \phi(z) - c(z-y)$. Then h is semi-concave and achieves a maximum at $z = x$. Therefore h admits 0 as supergradient at x. Take a subgradient v of c at $x - y$. Then, since $0 \in \partial h(x)$, one has $v \in \partial\phi(x)$. But by strict convexity of c we know that $\nabla c^*(v) = x - y$. □

Combining the previous results with the lemma 2.3 we get

PROPOSITION 4.7 (Differential of $\partial^c\phi$). *Fix a convex function c such that c and c^* are $\mathcal{C}^2$. Let $\mathcal{X}$ and Y be an open relatively compact and a compact subset of $\mathbf{R}^n$, respectively, and $\phi \in \mathcal{I}^c(\overline{\mathcal{X}}, Y)$. At every point $x \in \mathcal{X}$ where $\partial\phi$ is differentiable (and hence almost everywhere) the set-valued map $\partial^c\phi$ is also differentiable and*

$$\begin{aligned} d(\partial^c\phi)_x &= I - \mathrm{Hess}_{\nabla\phi(x)}c^*\, d(\partial\phi)_x \\ &= I - \mathrm{Hess}_{\nabla\phi(x)}c^*\, \mathrm{Hess}_x\phi \end{aligned} \tag{13}$$

Furthermore, $d(\partial^c\phi)_x$ has only real non-negative eigenvalues.

The non-negativity of the eigenvalues follows from c-concavity. Indeed, it tells us that the self-adjoint operator

$$\mathrm{Hess}_{x-T(x)}c - \mathrm{Hess}_x\phi = \mathrm{Hess}_{\nabla c^*(\nabla\phi(x))}c - \mathrm{Hess}_x\phi$$

is non-negative. But,

$$d(\partial^c\phi)_x = \mathrm{Hess}_{\nabla\phi(x)}c^* \left(\mathrm{Hess}_{\nabla c^*(\nabla\phi(x))}c - \mathrm{Hess}_x\phi\right),$$

and the product of a positive matrix and of a non-negative matrix has only real non-negative eigenvalues.

THEOREM 4.8 (Jacobian equation for optimal transport). *Fix a convex function c such that c and c^* are $\mathcal{C}^2$, and $\mu, \nu \ll$ vol two compactly supported Borel probability measures with density f and g, respectively. Let $\mathcal{X}$ and $\mathcal{Y}$ be two open relatively containing the support of μ and ν, respectively, and let $\phi \in \mathcal{I}^c(\overline{\mathcal{X}}, \overline{\mathcal{Y}})$ such that $T = \partial^c\phi\, ae$ is the c-optimal transport (7) pushing μ forward to ν. Then there exists a Borel set $K \subset \mathcal{X}$ of measure 1 for μ such that for every $x \in K$ the function ϕ has a Hessian (12) (hence $\partial^c\phi$ has a differential) at x and the following equation holds*

$$\begin{aligned} 0 \neq f(x) &= g(T(x)) \det d(\partial^c\phi)_x \\ &= g(T(x)) \det\left(I - \mathrm{Hess}_{\nabla\phi(x)}c^*\, \mathrm{Hess}_x\phi\right). \end{aligned} \tag{14}$$

PROOF. Remember that $\phi^c \in \mathcal{I}^c(\overline{\mathcal{Y}}, \overline{\mathcal{X}})$ and that Gangbo and McCann proved that $T^{-1}(y) = y + \nabla c^*(-\nabla\phi^c(y)) = \partial^c\phi^c(y)$ $\nu - a.e.$ (on $\mathcal{Y}$) is the c-optimal transport pushing ν forward to μ. Introduce, as in [**6**], the set

$$\begin{aligned} K_0 &:= \{x \in \mathcal{X}\,;\ \partial\phi \text{ is differentiable at } x \text{ and } \partial\phi^c \text{ is differentiable at } T(x)\} \\ &= \{x \in \mathcal{X}\,;\ \partial^c\phi \text{ is differentiable at } x \text{ and } \partial^c\phi^c \text{ is differentiable at } T(x)\} \end{aligned}$$

The second equality follows from proposition 4.7. Combining lemma 4.5, Aleksandrov theorem applied to ϕ and ϕ^c, together with the definition of mass transport (lemma 3.6), one has $\mu(K_0) = 1$. Now, set

$$\begin{aligned} K := \{x \in K_0 \quad ; \quad & x \text{ is a Lebesgue point of } f, \text{ with } f(x) \neq 0, \text{ and} \\ & T(x) \text{ is a Lebesgue point of } g\} \end{aligned}$$

Again, we have $\mu(K) = 1$. For $x \in K_0$, the map $d(\partial^c\phi(x))$ is invertible in virtue of (11) and of lemma 2.4. The result then follows by applying proposition 2.5 together with the definition of Lebesgue points. □

REMARK 4.9 (How to use it). *Theorem 4.8 can be read as follows: let f and g be compactly supported Borel probability densities and T the c-optimal mass transport pushing f forward to g; then for almost every x there exists a linear map dT_x (satisfying (1)) with only real non-negative eigenvalues such that*

$$f(x) = g(T(x)) \det dT_x.$$

The map dT_x is not the differential of the Borel map T, nor its distributional derivative: it is, where it exists, the differential of the set valued map $\partial^c\phi$. It can be called the "differential of T in the sense of Aleksandrov" since it includes the Hessian of ϕ in the sense of Aleksandrov. We stress that this notion of "in the sense of Aleksandrov" is not exactly the one that was used by Caffarelli in his study of the Monge-Ampère equation. When one is interested in proving functional inequalities, it can be useful to do integration by parts, and so one would like to compare this differential with the distributional derivative. The general philosophy is that in a convex setting, derivatives in the sense of Aleksandrov provide the absolute continuous part of distributional derivatives. Here we prove the following weaker result which is however sufficient in usual applications. This result depends only on the semi-concavity of ϕ and is independent of Theorem 4.8. Since we will work with convex functions, it is convenient to assume that $\mathcal{X}$ is convex (enlarge it if necessary).

PROPOSITION 4.10 (Integration by parts). *As in proposition 4.7, fix a convex function c such that c and c^* are $\mathcal{C}^2$. Let $\mathcal{X}$ and Y be a convex open relatively compact and a compact subset of $\mathbf{R}^n$, respectively, $\phi \in \mathcal{I}^c(\overline{\mathcal{X}}, Y)$ and T be the Borel map $T(x) := x - \nabla c^*(\nabla\phi(x)) = \partial^c\phi(x)$ ae. Denote by dT the map defined for almost every $x \in \mathcal{X}$ by $dT_x := d(\partial^c\phi)_x$ (13). Then one has*

$$\int \alpha(x)\mathrm{tr}(dT_x)\,dx \leq -\int \langle \nabla\alpha,\, T\rangle \tag{15}$$

for every non-negative smooth test function α compactly supported inside $\mathcal{X}$.

PROOF. We first notice that ϕ is (globally) semi-concave. Indeed, we know that for every $x \in \mathcal{X}$, there exists a ball $B_r(x)$ and $k \geq 0$ such that $z \to \phi(z) - k|z|^2$ is concave in $B_r(x)$. Since $\mathcal{X}$ is relatively compact, this number k can be chosen such that for every $x \in \mathcal{X}$, the function $z \to \phi(z) - k|z|^2$ is concave in a neighborhood of x, thus $z \to \phi(z) - k|z|^2$ is concave throughout $\mathcal{X}$.

We prefer working with convex functions, so let us set $\tilde{c}(z) := c(-z)$ and $\psi := -\phi + k|x|^2$ so that $T(x) = x + \nabla\tilde{c}^*(\nabla\psi - kx)$. Then, ψ is *convex* on $\mathcal{X}$ and, where it exists,

$$\mathrm{tr}(dT_x) = \mathrm{tr}\left(I + \mathrm{Hess}_{\nabla\psi(x)-kx}\tilde{c}^*\left(\mathrm{Hess}_x\psi - kI\right)\right) = \mathrm{tr}\left(I + A_x(\mathrm{Hess}_x\psi - kI)\right)$$

where $\mathrm{Hess}_x\psi$ is the Hessian of ψ in the sense of Aleksandrov and $A_x := \mathrm{Hess}_{\nabla\psi(x)-kx}\tilde{c}^*$. We now closely follow the arguments of Evans and Gariepy [**7**] (based on Lebesgue's classical method for studying derivatives of measures). We use an approximation ψ_n of ψ obtained by convolution with smooth non-negative functions with support tending to $\{0\}$ (we work in some open convex relatively compact set $\mathcal{O}$ containing the support of α and whose closure is inside $\mathcal{X}$; in $\mathcal{O}$ the gradient $\nabla\psi$ remains essentially bounded). The ψ_n's are smooth and convex and $\nabla\psi_n$ remains uniformly essentially bounded in $\mathcal{O}$. Let $T_n(x) := x + \nabla\tilde{c}^*(\nabla\psi_n(x) - kx)$. Since almost every x is a Lebesgue point of $\nabla\psi$ we have $\nabla\psi_n \to \nabla\psi\, ae$ and thus $T_n \to T\, ae$. Because T_n is smooth we have

$$-\int \langle \nabla\alpha, T_n\rangle = \int \alpha\, \mathrm{tr}(dT_n). \tag{16}$$

We note that $d(T_n)_x = I + \mathrm{Hess}_{\nabla\psi_n(x)-kx}\tilde{c}^*\,(\mathrm{Hess}_x\psi_n - kI)$ and since $\mathrm{Hess}\psi_n \geq 0$ and $\tilde{c}^*$ is $\mathcal{C}^2$, we know that $\mathrm{tr}(dT_n)$ remains uniformly bounded from below on the compact set supporting α. Thus by applying Fatou's lemma on the compact set supporting α we have, in view of (16) and of the fact that the sequence T_n tending ae to T remains essentially bounded:

$$-\int \langle \nabla\alpha, T\rangle \geq \int \alpha \liminf \mathrm{tr}(dT_n) \tag{17}$$

The convex function ψ admits a second order derivative, in the sense of distributions on $\mathcal{X}$, denoted by $[D^2\psi]$, which is a (matrix valued) Radon measure. Let us denote by $D^2\psi$ the density (with respect to the Lebesgue measure) of the absolutely continuous part of $[D^2\psi]$. This is not exactly the definition of the Hessian in the sense of Aleksandrov; however, as proved in [**7**], one has for almost every $x \in \mathcal{X}$, $\mathrm{Hess}_x\psi = D^2\psi(x)$ and thus

$$\int \alpha(x)\mathrm{tr}(dT_x)\, dx = \int \alpha(x)\mathrm{tr}\left(I + A_x(D^2\psi(x) - kI)\right)\, dx. \tag{18}$$

It follows from [**7**, §6.1 p242] that for almost every x one has $\mathrm{Hess}_x\psi_n \longrightarrow D^2\psi(x)$ and therefore for almost every x one has $d(T_n)_x \longrightarrow I + A_x(D^2\psi(x) - kI)$. Then, by (17) and (18) we obtain the desired inequality. ☐

An application of the previous results can be found in [**5**]. Let us mention that theorem 4.8 and proposition 4.10 are closely related to technical results of Otto [**11**].

5. Interpolant map and interpolant measure

Again, throughout this section, we will follow the arguments given in [**6**].

PROPOSITION 5.1 (Interpolant injectivity). *Let ϕ be a c-concave function, for some strictly convex c. For $t \in (0,1)$ introduce the set-valued map M_t defined by the following Minkowski sum:*

$$M_t(x) := (1-t)x + t\partial^c\phi(x).$$

Then M_t is injective on $dom(M_t) = dom(\partial^c\phi)$.

PROOF. Fix $x, x' \in dom(\partial^c\phi)$ and assume $z := M_t(x) = M_t(x')$. Setting $y = \partial^c\phi(x)$ and $y' = \partial^c\phi(x')$ one has (draw a picture!)

$$\begin{aligned} y' - x &= y' - z + z - x \\ &= (1-t)(y' - x') + t(y - x) \end{aligned} \tag{19}$$

and similarly

$$\begin{aligned} y - x' &= y - z + z - x' \\ &= (1-t)(y - x) + t(y' - x'). \end{aligned} \tag{20}$$

By strict convexity of c one has, using (19) and (20),

$$c(y' - x) + c(y - x') < c(y - x) + c(y' - x'). \tag{21}$$

unless $y - x = y' - x' = y' - x = y - x'$. Thus (21) holds unless $x = x'$. But c-cyclic monotony of $\partial^c\phi$, for $y \in \partial^c\phi(x)$ and $y' \in \partial^c\phi(x')$ imposes

$$c(y - x) + c(y' - x') \leq c(y' - x) + c(y - x').$$

Thus (21) cannot hold, leading to $x = x'$ and injectivity of M_t on its domain. ☐

THEOREM 5.2 (Interpolant density). *Fix a $\mathcal{C}^2$ convex function c such that c^* is also $\mathcal{C}^2$. Let μ and ν be two absolutely continuous and compactly supported Borel probability measures. Let $\mathcal{X}$ and $\mathcal{Y}$ be two open relatively compact sets containing the support of μ and ν respectively, and $\phi \in \mathcal{I}^{\,c}(\overline{\mathcal{X}}, \overline{\mathcal{Y}})$ such that $T = \partial^c\phi$ ae is the c-optimal map (7) pushing μ forward to ν. For $t \in (0,1)$, introduce the interpolant map*

$$T_t(x) := x - t\nabla c^*(\nabla\phi(x)) = (1-t)x + t\partial^c\phi(x) =: M_t(x) \;\; \mu - ae.$$

Then the image measure $(T_t)_\# \mu = (M_t)_\# \mu$ is absolutely continuous (and hence has a density).

PROOF. Once again, introduce as in the proof of theorem 4.8 the set

$$\begin{aligned} K_0 &:= \{x \in \mathcal{X}\,;\ \partial\phi \text{ is differentiable at } x \text{ and } \partial\phi^c \text{ is differentiable at } T(x)\} \\ &= \{x \in \mathcal{X}\,;\ \partial^c\phi \text{ is differentiable at } x \text{ and } \partial^c\phi^c \text{ is differentiable at } T(x)\}. \end{aligned}$$

One has $\mu(K_0) = 1$ and since $\mu \ll \mathrm{vol}$, it is enough to prove that the image measure $(M_t)_\#\mathrm{vol}_{K_0}$ of the Lebesgue measure vol restricted to K_0 under the map M_t is absolutely continuous. We will assume that K_0 is σ-compact. If it was not, using regularity of μ one could replace it by a σ-compact subset still carrying the full measure of μ.

Remember that for $x \in K_0$, the set $M_t(x)$ is a single point $M_t(x) = (1-t)x + tT(x) = T_t(x)$ ($K_0 \subset dom(M_t) \cap int(\mathcal{D}(M_t))$). Again, for $x \in K_0$, the linear map $d(\partial^c\phi)_x$ is invertible in virtue of (11) and lemma 2.4. Of course at $x \in K_0$, the set-valued map M_t has also a differential and $d(M_t)_x = (1-t)I + td(\partial^c\phi)_x$. But we know by proposition 4.7 that $d(\partial^c\phi)_x$ has only real non-negative eigenvalues. Thus we know that $d(M_t)_x$ is invertible at every $x \in K_0$. Finally note that the map M_t is continuous on K_0 (by remark 2.2). The end of the proof relies on the following lemma:

LEMMA 5.3. *Let $K \subset K_0$ be a compact subset of K_0. For every $x \in K$ there exists $k_x > 0$ such that for every $z \in K$ one has*

$$|M_t(x) - M_t(z)| \geq k_x\,|x - z|. \tag{22}$$

Consequently $(M_t)_\#\mathrm{vol}_K$ is absolutely continuous.

PROOF. Assume the claimed property was false. One could find a sequence $(x_k) \subset K$ with $x_k \neq x$, such that

$$|M_t(x) - M_t(x_k)| \leq |x - x_k|/k = o(|x - x_k|). \tag{23}$$

By compactness one can take a subsequence, still denoted by x_k, such that x_k converges to some $z \in K$. But we know by proposition 5.1 that M_t is injective on K and thus, by continuity in (23), we have $z = x$ and $x_k \to x$. Differentiability of M_t at x implies that

$$M_t(x_k) = M_t(x) + d(M_t)_x(x - x_k) + o(|x - x_k|).$$

But this, together with invertibility of $d(M_t)_x$, contradicts (23).

Using standard arguments from measure theory, we shall now deduce that the image of vol_K under M_t is absolutely continuous. The previous estimate (22) tells us that $K = \bigcup_k K_k$ where

$$K_k = \{x \in K \mid \forall z \in K,\ |M_t(x) - M_t(z)| \geq \frac{1}{k}|x - z|\,\}.$$

Continuity of M_t shows the sets K_k to be closed, hence compact. But $M_t : K_k \longrightarrow K'_k := M_t(K_k)$ has the property that M_t^{-1} is Lipschitz on K'_k. By a classical argument using the Vitali covering lemma, the Lipschitz map M_t^{-1} cannot increase the volume of any subset of K'_k by factor greater than k^n. Thus the image of vol_{K_k} under $M_t : K_k \longrightarrow K'_k$ is absolutely continuous. The image of vol_K under M_t is an increasing limit of the images of vol_{K_k} under M_t and so itself absolutely continuous, which concludes the proof of the lemma. □

End of the proof of the theorem. To deduce the absolute continuity of $(M_t)_\#\mathrm{vol}_{K_0}$ from the previous lemma, write $K_0 = \bigcup K_j$ as an increasing union of compact sets K_j. The image of vol_{K_0} under M_t is the increasing limit of the images of K_j under M_t which are absolutely continuous from the previous lemma. Consequently the image of vol_{K_0} under M_t is absolutely continuous. This implies the desired absolute continuity for $(M_t)_\#\mu$. □

REMARK 5.4 (Optimality). *When the cost c is the quadratic cost, or any homogeneous cost, the map T_t in Theorem 5.2 is the optimal map pushing μ forward to μ_t (one should note that $t\phi$ is c-concave, see* [**9**]*). Thus, the densities will satisfy some Jacobian equation. However, in the general case, there is no reason for the map T_t to be the c-optimal map pushing μ forward to μ_t.*

References

[1] G. Bianchi, A. Colesanti and C. Pucci, On the second differentiability of convex surfaces. *Geom. Dedicata.* 60 (1996), 39–48.

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

[3] L. Caffarelli, Allocation maps with general cost functions, in *Partial Differential Equations and Applications*, Lecture Notes in Pure and Applied Math. 177 (1996), 29–35.

[4] D. Cordero-Erausquin, Some applications of mass transport to Gaussian type inequalities, *Arch. Rational Mech. Anal.* 161 (2002), 257–269.

[5] D. Cordero-Erausquin, W. Gangbo and C. Houdre, Inequalities for generalized entropy and optimal transportation, *Same volume.*

[6] D. Cordero-Erausquin, R. J. McCann and M. Schmuckenschläger, A Riemannian interpolation inequality à la Borell, Brascamp and Lieb, *Invent. Math.* 146 (2001), 219–257.

[7] L.C. Evans and R.F. Gariepy, *Measure theory and fine properties of functions*, Studies in Advanced Mathematics, CRC Press, 1992.

[8] W. Gangbo and R. J. McCann, Optimal maps in Monge's mass transport problem, *C. R. Acad Sci. Paris Sér. I Math.* 321 (1995), 1653–1658.

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

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

[11] F. Otto, Doubly degenerate diffusion equations as steepest descent, preprint.

Laboratoire d'Analyse et de Mathématiques Appliquées (UMR 8050), Université de Marne la Vallée, 77454 Marne la Vallée Cedex 2, France

E-mail address: cordero@math.univ-mlv.fr