manuscripta math.

Zoltán M. Balogh · Valentina Penso

On singular sets of c-concave functions

Received: 24 March 2017 / Accepted: 30 August 2017

Abstract. We prove that under quite general condition on a cost function c in $\mathbb{R}^n$ the Hausdorff dimension of the singular set of a c-concave function has dimension at most $n-1$. Our result applies for non-semiconcave cost functions and has applications in optimal mass transportation.

1. Introduction

In the theory of optimal mass transportation [14,15] convex/concave functions play an important role. For instance, in the case of the quadratic cost $c: \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$ $c(x, y) = \frac{1}{2}|x-y|^2$ the optimal transport map Ψ between two absolute continuous measures μ and ν with $\Psi_\sharp \mu = \nu$ is given by the gradient of a convex function $\Psi = \nabla u$ almost everywhere.

Let us recall that the optimal transport map is the minimizer of the energy

$$E(\Phi) = \int c(x, \Phi(x)) d\mu(x),$$

among all measurable maps $\Phi: \mathbb{R}^n \to \mathbb{R}^n$ transporting μ into ν i.e. $\Phi_\sharp \mu = \nu$.

For the general case of continuous cost function $c(x, y)$ the situation becomes more involved. In this generality the existence of an optimal transport map is not so clear. However, the existence of an *optimal transport plan* can be guaranteed. Its support lies on the graph of the c-superdifferential $\partial^c u$ of a c-concave function u. If $\partial^c u(x)$ is single valued for μ almost every x, then the optimal transport map exists and it is given by the map $x \mapsto \partial^c u(x)$ for μ almost every x. It is therefore an important matter to study the singularity set of $\partial^c u$, i.e. the set of those points x where $\partial^c u(x)$ is not a singleton. Recent developments in the theory of optimal mass transportation [3] especially in sub-Riemannian geometries [4,6,13] or geometric optics [8,9] motivates the consideration of general cost functions c.

The authors acknowledge the support of the Swiss National Science Foundation Grant No. 200020-146477.

Z. M. Balogh (✉)· V. Penso: Mathematisches Institute, Universität Bern, Sidlerstrasse 5, 3012 Bern, Switzerland. e-mail: zoltan.balogh@math.unibe.ch

V. Penso: e-mail: valentina.penso@math.unibe.ch

Mathematics Subject Classification: 26B25 · 28A78

DOI: 10.1007/s00229-017-0971-2

Published online: 11 September 2017

Alberti et al. [1] studied the singularity set of convex/concave and more general semiconvex/semiconcave functions. Let us recall that for a concave function $u\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ the *superdifferential* of u at the point $x \in \mathbb{R}^n$ is the set

$$\partial u(x) := \left\{p \in \mathbb{R}^n \mid u(z) \leq u(x) + \langle p, x - z\rangle, \forall z \in \mathbb{R}^n\right\}.$$

The *singular set* of u is defined by

$$S(u) = \{x \in \mathbb{R}^n \mid \# \partial u(x) > 1\}.$$

Then by a result going back to Reshetnyak [12] (see also [1,2,7]), it follows that $\dim S(u) \leq n-1$.

A comprehensive overview of these results can be found in the book of Cannarsa and Sinestrari [5]. The main idea is to use the more general concept of semiconvexity or semiconcavity.

Let us recall that if Ω is an open subset of $\mathbb{R}^n$ and $\omega\colon \mathbb{R}^+ \longrightarrow \mathbb{R}^+$ is continuous and such that $\omega(r) = o(r)$, as $r \to 0$, we say that a function $f\colon \Omega \longrightarrow \mathbb{R} \cup \{+\infty\}$ is *semiconcave with modulus* ω if, for any $x, y \in \mathbb{R}^n$ and for any $t \in [0, 1]$, the following inequality is verified

$$f((1-t)x + ty) \geq (1-t)f(x) + tf(y) + t(1-t)\omega(|x-y|).$$

In this case we can define the superdifferential of u at $x \in \Omega$ as

$$\partial u(x) := \left\{p \in \mathbb{R}^n \mid \limsup_{z \to x} \frac{u(z) - u(x) - \langle p, x - z\rangle}{|z - x|} \leq 0\right\}.$$

The singular set of $u\colon \mathbb{R}^n \to \mathbb{R}$ is defined as above and in this more general case also follows that its Hausdorff dimension is less or equal than $n-1$.

Using these results one can conclude the same bound on the Hausdorff dimension for singular set for the c-superdifferential of c-concave functions under the assumption of semiconcavity of the cost function c.

The purpose of this paper is to study the size of the singular set of $\partial^c u$ for a rather general class of cost functions c that go beyond the class of semiconcave functions. In order to state our result we have to first recall the concepts we need and fix some notation.

We say that u is *c-concave* if

$$u(x) = \inf_{y \in Y} c(x, y) - \varphi(y), \quad \forall x \in \mathbb{R}^n,$$

for some suitable non empty set $Y \subset \mathbb{R}^n$ and $\varphi\colon Y \longrightarrow \mathbb{R} \cup \{\infty\}$, $\varphi \not\equiv \infty$. We call the *c-transform* of u the function

$$u^c(y) = \inf_{x \in \mathbb{R}^n} c(x, y) - u(x), \tag{1}$$

for every $y \in \mathbb{R}^n$. Then, for every x and $y \in \mathbb{R}^n$, it holds:

$$\begin{aligned} u(x) + u^c(y) &= u(x) + \inf_{\tilde{x} \in \mathbb{R}^n} c(\tilde{x}, y) - u(\tilde{x}) \\ &\leq u(x) + c(x, y) - u(x) = c(x, y). \end{aligned}$$

Definition 1.1. Let $u\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ and let $x \in \mathbb{R}^n$ be a point. We define the *c-superdifferential* of u at x as the (possibly empty) set,

$$\begin{aligned}\partial^c u(x) := &\left\{p \in \mathbb{R}^n \mid c(x,p) - u(x) \le c(z,p) - u(z),\ \text{for every } z \in \mathbb{R}^n\right\}\\ = &\left\{y \in \mathbb{R}^n \,\middle|\, u(x) + u^c(y) = c(x,y)\right\}.\end{aligned}$$

Definition 1.2. Given a function $u\colon \mathbb{R}^n \longrightarrow \mathbb{R}$, we call *singular set* of u the set

$$\Sigma(u) = \left\{x \in \mathbb{R}^n \,\middle|\, \#\partial^c u(x) > 1\right\}.$$

In the remaining part of this paper we shall use the following standing assumptions on the cost function c.

Assumption 1.1. (*Differentiability Condition*) The cost function $c\colon \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}$ is a continuous function. For every $p \in \mathbb{R}^n$, the function $x \longmapsto c(x,p)$ is differentiable almost everywhere. We denote by

$$C(p) = \left\{x \in \mathbb{R}^n \mid \nabla_x c(x,p) \text{ exists}\right\}.$$

We assume that for every $x \in \mathbb{R}^n$, there exists $r = r(x) > 0$ such that for all $0 < r < r(x)$ there exists $\hat{x} \in B(x,r)$ with the property that $\hat{x} \in C(p)$ for all $p \in \mathbb{R}^n$. Furthermore, we assume that Taylor's formula with the above notations holds, in the sense that

$$\lim_{r\to 0} \frac{c(x,p) - c(\hat{x},p) - < \nabla_x c(\hat{x},p), (x-\hat{x}) >}{r} = 0. \tag{2}$$

Assumption 1.2. (*Twist Condition*) For any compact set $K \subset \mathbb{R}^n$ and for any $0 < \eta < 1$, there exists $0 < \xi = \xi(\eta) < 2$ such that if $p_1, p_2 \in K$ and $x \in K$ are so that $|p_1 - p_2| > \eta$ and $x \in C(p_1) \cap C(p_2)$, then

$$|\nabla_x c(x,p_1) - \nabla_x c(x,p_2)| > \xi. \tag{3}$$

Assumption 1.3. (*Gradient Continuity Condition*) For any compact set $K \subset \mathbb{R}^n$, for any $\varepsilon > 0$ and for any $\eta > 0$, there exists $\delta = \delta(\varepsilon,\eta) > 0$ such that if $p_1, p_1', p_2, p_2' \in K$ and $x \in K$ are so that $|p_1 - p_2| > \eta$, $|p_1 - p_1'| < \delta$, $|p_2 - p_2'| < \delta$ and $x \in C(p_1) \cap C(p_1') \cap C(p_2) \cap C(p_2')$, then either

$$|\nabla_x c(x,p_1) - \nabla_x c(x,p_1')| < \varepsilon \tag{4}$$

or

$$|\nabla_x c(x,p_2) - \nabla_x c(x,p_2')| < \varepsilon. \tag{5}$$

Let us comment briefly on the above conditions. Note first, that in the first condition the required differentiability at the point $\hat{x}$ does not necessarily imply the validity of Taylor's formula in the form that it is written in (2). The reason for this is that the usual Taylor's formula is written in terms of the point $\hat{x}$. On the other hand, in our case the radius $r = r(x) > 0$ depends on x and not on $\hat{x}$ and this is why we need to require in addition the validity of (2). In practical applications however it is rather easy to check this condition as for typical cost functions the differentiability fails only on the diagonal $x = p$.

The second condition formulates a certain quantitative injectivity of the gradient of the cost function in its differentiability points. This condition is rather standard in the theory of optimal mass transportation [15].

The third condition is new. It says that the mapping $p \mapsto \nabla_x c(x, p)$ cannot have too many discontinuities close to each other. In some sense, this condition replaces the missing semiconcavity assumption.

We are now ready to state our main theorem.

Theorem 1.1. *Let $u: \mathbb{R}^n \longrightarrow \mathbb{R}$ be a c-concave function. Suppose that c satisfies Assumptions 1.1, 1.2 and 1.3. Then* $\dim \Sigma(u) \leq n - 1$.

The paper is organized as follows: in Sect. 2 we give the proof of Theorem 1.1. Our approach is different to the classical one [1,5] and uses dimension estimates of porous sets due to Mattila [11]. In Sect. 3 we shall discuss our standing assumptions from the statement of Theorem 1.1 through examples. Also here we shall give an application of to optimal mass transportation.

2. Proof of the main result

As mentioned above the proof of Theorem 1.1 uses a result due to Mattila, about Hausdorff dimension estimates of porous sets which we are going to recall. The interested reader is referred to the paper [10] and also to Chapter 11 of [11].

We start by recalling some notation. Let $A \subset \mathbb{R}^n$, $x \in \mathbb{R}^n$ and $r > 0$ be fixed. We set

$$p(A, x, r) := \sup \left\{ \rho > 0 \,|\, B(z, \rho) \subset B(x, r) \setminus A, \text{ for some } z \in \mathbb{R}^n \right\}$$

We call the *porosity* of A at the point x the number

$$p(A, x) := \liminf_{r \to 0} \frac{p(A, x, r)}{r}.$$

Notice that for any point $x \in A$ we have $0 \leq p(A, x, r) \leq \frac{r}{2}$. Consequently, it follows that $0 \leq p(A, x) \leq \frac{1}{2}$.

Using this notation the result of Mattila is stated as:

Theorem 2.1. *Let $p \in]0, \frac{1}{2}[$. Then there exists $d(p) \in [n - 1, n]$ such that*

$$\lim_{p \to \frac{1}{2}} d(p) = n - 1$$

and $\dim A \leq d(p)$*, for every $A \subset \mathbb{R}^n$ with the property that $p(A, x) \geq p$, for every $x \in A$.*

The following lemma, which will be a key ingredient in the proof of Theorem 1.1, generates a condition for a set to be porous.

Lemma 2.1. *Let $x \in \mathbb{R}^n$, $r > 0$, $w \in \mathbb{S}^{n-1}$ and $0 < \rho < \frac{1}{2}$ be fixed. If $x' \in B(z, (1/2 - \rho)r)$ where $z = x + \frac{1}{2}rw$ then*

$$\left\langle \frac{x' - x}{r}, w \right\rangle > \rho. \tag{6}$$

Proof. Let us write $x' \in B(z, (1/2 - \rho)r)$ in the form $x' = z + y$ where $|y| < (\frac{1}{2} - \rho)r$.

Since $z = x + \frac{1}{2}rw$ we obtain $x' = x + \frac{1}{2}rw + y$ for $y \in \mathbb{R}^n$, $|y| < (\frac{1}{2} - \rho)r$. Observe that

$$\begin{aligned}\langle x' - x, w\rangle &= \left\langle \frac{1}{2}rw + y, w \right\rangle = \frac{1}{2}r + \langle y, w\rangle \\ &\geq \frac{1}{2}r - |y| > \frac{1}{2}r - \left(\frac{1}{2} - \rho\right) r = \rho r > 0.\end{aligned}$$

Dividing this inequality by $r > 0$ yields the claim. □

Remark 2.1. Notice that this Lemma provides a sufficient condition for porosity. Let $A \subset \mathbb{R}^n$ and let $x \in A$ be fixed. If there exists $r_0 = r_0(x) > 0$ such that for all $0 < r < r_0(x)$ there exists a vector $w = w(x, r) \in \mathbb{S}^{n-1}$ such that for all $x' \in A \cap B(x, r)$ we have

$$\left\langle \frac{x' - x}{r}, w \right\rangle \leq \rho, \tag{7}$$

then according to the above Lemma $x' \notin B\left(z, \left(\frac{1}{2} - \rho\right) r\right)$, for every $0 < r < r_0$. This implies that $p(A, r, x) \geq \left(\frac{1}{2} - \rho\right) r$.

We are now ready to give the proof of Theorem 1.1.

Proof of Theorem 1.1. Our goal is to show that, given a c-concave function $u: \mathbb{R}^n \longrightarrow \mathbb{R}$, the singular set $\Sigma(u)$ has dimension at most $n - 1$.

Notice that we can restrict to the case of non isolated points. Indeed, the set of isolated point of $\Sigma(u)$ has zero dimension. For simplicity of notation, we denote by Σ the set $\{x \in \Sigma(u) \mid x \text{ is not isolated}\}$.

The proof will be divided in several steps, with the purpose of simplifying the set to which we apply our argument.

Step 1 Let us define, for every ν and $N \in \mathbb{N}$,

$$\Sigma_{\nu,N} = \left\{ x \in \overline{B(0, N)} \,\middle|\, \exists\, p, q \in \overline{B(0, N)} \cap \partial^c u(x),\ d(p, q) > \frac{1}{\nu} \right\}. \tag{8}$$

Therefore, we can assert that

$$\Sigma = \bigcup_{\nu, N \in \mathbb{N}} \Sigma_{\nu,N}. \tag{9}$$

Let us check that (9) is true. If $x \in \bigcup_{\nu,N\in\mathbb{N}} \Sigma_{\nu,N}$, there are ν and $N \in \mathbb{N}$ such that $p, q \in \overline{B(0,N)} \cap \partial^c u(x)$ with $d(p,q) > \frac{1}{\nu}$. This implies that $p \neq q$ so $\#\partial^c u(x) > 1$.

Now the opposite inclusion should be proved. Let $x \in \Sigma$. Since $\#\partial^c u(x) > 1$, there are $p, q \in \partial^c u(x)$ such that $d := d(p,q) > 0$. We choose $\bar{N} = \lceil \max\{|p|, |q|\} + 1 \rceil$ and $\bar{\nu} \in \mathbb{N}$ such that $\bar{\nu} > \frac{1}{d}$. With this choice, we can conclude that $x \in \Sigma_{\bar{\nu},\bar{N}}$.

Decomposition (9) will help us to reach our goal. Indeed, it is clear that

$$\dim \Sigma = \dim\left(\bigcup_{\nu,N} \Sigma_{\nu,N}\right) = \sup\{\dim \Sigma_{\nu,N} \mid \nu, N \in \mathbb{N}\}.$$

The decomposition built in this first step allows us to focus our attention on a set of the form

$$\Sigma_{\nu,N} = \left\{ x \in \overline{B(0,N)} \,\middle|\, \partial^c u(x) \subset \overline{B(0,N)} \text{ and } \operatorname{diam} \partial^c u(x) > \frac{1}{\nu} \right\}.$$

We shall namely prove that $\dim \Sigma_{\nu,N} \le n - 1$ for general ν and $N \in \mathbb{N}$ fixed.

Step 2 From now on, let $\nu \in \mathbb{N}$ and $N \in \mathbb{N}$ be fixed. The proof of the Theorem is a direct consequence of the following

Claim. *Let* $0 < \rho < \frac{1}{2}$. *Then there exists a finite family of sets* $\left\{\Sigma^i_{\nu,N}\right\}_{i=1}^{\sigma}$ *such that*

- $\Sigma_{\nu,N} = \bigcup_{i=1}^{\sigma} \Sigma^i_{\nu,N}$;
- $p\left(\Sigma^i_{\nu,N}, x\right) > \frac{1}{2} - \rho$, *for every* $x \in \Sigma^i_{\nu,N}$.

Let us postpone the proof of this Claim for now, and see why it implies the theorem. By Theorem 2.1, if the Claim is true, it follows that, for every $i \in \{1, \dots, \sigma\}$,

$$\dim \Sigma^i_{\nu,N} \le d\left(\frac{1}{2} - \rho\right),$$

with the property that

$$\lim_{\rho\to 0} d\left(\frac{1}{2} - \rho\right) = n - 1.$$

This implies that, for $\varepsilon > 0$, there exists $\rho = \rho(\varepsilon) > 0$, sufficiently small, such that $\dim \Sigma^i_{\nu,N} \le n - 1 + \varepsilon$, which clearly entails that

$$\dim \Sigma_{\nu,N} \le n - 1 + \varepsilon.$$

Letting $\varepsilon \longrightarrow 0$, the Claim follows. The remaining steps are devoted to proving the Claim. From now on, let $0 < \rho < \frac{1}{2}$ be fixed.

Step 3 The task of this step is to obtain a finite decomposition for the set $\Sigma_{\nu,N}$, proving the first part of Claim.

Our efforts will focused in building a δ-covering for the ball $\overline{B(0,N)}$, where the c-superdifferential lives. The question now is: how do we choose the radius $\delta > 0$, depending on ρ?

If we set $\eta = \frac{1}{2\nu}$, we can choose $\xi > 0$, dependent on η therefore on ν, such that inequality (3) holds. We set $\varepsilon = \frac{1}{2}\xi\rho$ and then select the relative $\delta = \delta(\rho,\nu) > 0$, for which Assumption 1.2 is satisfied. Without loss of generality, we assume that $\delta < \frac{1}{8\nu}$.

We consider a δ-covering $\mathcal{B} := \{B(q_i,\delta)\}_{i=1}^M$ of $\overline{B(0,N)}$. Clearly, we can assume that it is a finite family because $\overline{B(0,N)}$ is a compact set. We then define

$$\Sigma_{\nu,N}^{(j,l)} := \Big\{x \in \Sigma_{\nu,N} \,\Big|\, \partial^c u(x) \cap B(q_j,\delta) \neq \emptyset,\ \partial^c u(x) \cap B(q_l,\delta) \neq \emptyset,\ \mathrm{dist}(B(q_j,\delta), B(q_l,\delta)) > \frac{1}{4\nu}\Big\}.$$

Let us prove that

$$\Sigma_{\nu,N} = \bigcup_{j,l} \Sigma_{\nu,N}^{(j,l)}. \tag{10}$$

If $x \in \Sigma_{\nu,N}$, then there exist p_1 and $p_2 \in \partial^c u(x)$ such that $|p_1 - p_2| > \frac{1}{2\nu}$. Since p_1 and $p_2 \in \partial^c u(x) \subset \overline{B(0,N)}$, there are $B(q_j,\delta)$ and $B(q_l,\delta) \in \mathcal{B}$ such that $p_1 \in B(q_j,\delta)$ and $p_2 \in B(q_l,\delta)$. Then $x \in \Sigma_{\nu,N}^{(j,l)}$, because

$$\mathrm{dist}\,(B(q_k,\delta), B(q_l,\delta)) \geq d(p_1,p_2) - 2\delta > \frac{1}{4\nu}.$$

Moreover, the union in (10) is finite; indeed the family $\mathcal{B}$ has a finite number of elements. This concludes the proof of the first part of the Claim.

Step 4 Let us now consider the second part of the Claim: the aim is to show that, for every $j, l \in \{1,\ldots,M\}$ fixed,

$$p\left(\Sigma_{\nu,N}^{(j,l)}, x\right) \geq \frac{1}{2} - \rho, \tag{11}$$

for every $x \in \Sigma_{\nu,N}^{(j,l)}$.

In this step we show that for a point $x \in \Sigma_{\nu,N}^{(j,l)}$ there exists $r_0 = r_0(x) > 0$ such that for all $0 < r < r_0(x)$ we find a unit vector $w = w(x,r) \in \mathbb{S}^{n-1}$ such that if $x' \in \Sigma_{\nu,N}^{(j,l)} \cap B(x,r)$, then

$$\left\langle \frac{x'-x}{r}, w \right\rangle < \rho. \tag{12}$$

Let $x \in \Sigma_{\nu,N}^{(j,l)}$ be fixed. By definition of $\Sigma_{\nu,N}^{(j,l)}$, we can select $p_1 \in \partial^c u(x) \cap B(q_j,\delta)$ and $p_2 \in \partial^c u(x) \cap B(q_l,\delta)$. Saying that p_1 and $p_2 \in \partial^c u(x)$ means that the following two inequalities hold

$$c(x,p_1) - u(x) \leq c(z,p_1) - u(z), \tag{13}$$

$$c(x,p_2) - u(x) \leq c(z,p_2) - u(z), \tag{14}$$

for every $z \in \mathbb{R}^n$.

By our differentiability Assumption 1.1 we find a $r_0 = r_0(x) > 0$ such that for all $0 < r < r_0(x)$ there exists a point $\hat{x} \in B(x, r)$ that is a differentiability point for the function $x \to c(x, p)$ for any $p \in \mathbb{R}^n$. Without loss of generality we assume that for the point $\hat{x}$, the second option in our Continuity Assumption 1.3 (i.e. (5)) holds.

We set

$$w := \frac{\nabla_x c(\hat{x}, p_2) - \nabla_x c(\hat{x}, p_1)}{|\nabla_x c(\hat{x}, p_2) - \nabla_x c(\hat{x}, p_1)|}.$$

In the following we shall prove the estimate (12) for this choice of w. In order to do that, consider a point $x' \in \Sigma_{\nu,N}^{(j,l)} \cap B(x, r)$. Hence there exist $p_1' \in \partial^c u(x') \cap B(q_j, \delta)$ and $p_2' \in \partial^c u(x') \cap B(q_l, \delta)$. Therefore, again by the definition of c-superdifferential, one has

$$c(x', p_1') - u(x') \le c(z, p_1') - u(z), \tag{15}$$
$$c(x', p_2') - u(x') \le c(z, p_2') - u(z), \tag{16}$$

for every $z \in \mathbb{R}^n$.

We focus our attention to inequalities (13) and (15). In the first we set $z = x'$ and in the second $z = x$. Consequently, they read as follows

$$c(x, p_1) - c(x', p_1) \le u(x) - u(x'). \tag{17}$$

In a similar way, using inequalities (14) and (16) we obtain

$$c(x, p_2') - c(x', p_2') \ge u(x) - u(x'). \tag{18}$$

Inequalities (17) and (18) combined give

$$c(x, p_1) - c(x', p_1) \le c(x, p_2') - c(x', p_2'). \tag{19}$$

Using Assumption (1.1) we can apply Taylor's formula at the point $\hat{x}$ as indicated in that condition. We have

$$c(x, p_1) = c\left(\hat{x}, p_1\right) + \left\langle \nabla_x c(\hat{x}, p_1), x - \hat{x} \right\rangle + o(r), \tag{20}$$
$$c(x', p_1) = c\left(\hat{x}, p_1\right) + \left\langle \nabla_x c(\hat{x}, p_1), x' - \hat{x} \right\rangle + o(r). \tag{21}$$

We subtract Eq. (20) to (21) in order to get

$$c(x, p_1) - c(x', p_1) = \left\langle \nabla_x c(\hat{x}, p_1), x - x' \right\rangle + o(r). \tag{22}$$

With a similar argument, one has also

$$c(x, p_2') - c(x', p_2') = \left\langle \nabla_x c(\hat{x}, p_2'), x - x' \right\rangle + o(r). \tag{23}$$

Now, we combine (22) and (23) with inequality (19) and obtain

$$\left\langle \nabla_x c(\hat{x}, p_1), x - x' \right\rangle + o(r) \le \left\langle \nabla_x c(\hat{x}, p_2'), x - x' \right\rangle + o(r),$$

or, equivalently,

$$\langle \nabla_x c(\hat{x}, p_2') - \nabla_x c(\hat{x}, p_1), x' - x \rangle \leq o(r),$$

which can be written as

$$\langle \nabla_x c(\hat{x}, p_2) - \nabla_x c(\hat{x}, p_1), x' - x \rangle \leq \langle \nabla_x c(\hat{x}, p_2) - \nabla_x c(\hat{x}, p_2'), x' - x \rangle + o(r).$$

Dividing this relation by $\left|\nabla_x c(\hat{x}, p_2) - \nabla_x c(\hat{x}, p_1)\right| \cdot r$ yields:

$$\left\langle \frac{\nabla_x c(\hat{x}, p_2) - \nabla_x c(\hat{x}, p_1)}{|\nabla_x c(\hat{x}, p_2) - \nabla_x c(\hat{x}, p_1)|}, \frac{x' - x}{r} \right\rangle \leq \left\langle \frac{\nabla_x c(\hat{x}, p_2) - \nabla_x c(\hat{x}, p_2')}{\left|\nabla_x c(\hat{x}, p_2) - \nabla_x c(\hat{x}, p_1)\right|}, \frac{x' - x}{r} \right\rangle + o(1).$$

Note that $\left|\nabla_x c(\hat{x}, p_2) - \nabla_x c(\hat{x}, p_2')\right| \leq \epsilon$ and $\left|\nabla_x c(\hat{x}, p_2) - \nabla_x c(\hat{x}, p_1)\right| > \xi$ while $\frac{\epsilon}{\xi} < \frac{\rho}{2}$ by our choices of parameters made at the beginning of Step 3.

This implies that the first term on the right side of the above estimate is less than $\frac{\rho}{2}$ since the second term converges to 0 as $r \to 0$ we obtain that the second term is also less than $\frac{\rho}{2}$ for $r > 0$ small enough. Consequently we obtain that

$$\left\langle \frac{\nabla_x c(\hat{x}, p_2) - \nabla_x c(\hat{x}, p_1)}{|\nabla_x c(\hat{x}, p_2) - \nabla_x c(\hat{x}, p_1)|}, \frac{x' - x}{r} \right\rangle = \left\langle w, \frac{x' - x}{r} \right\rangle \leq \rho,$$

for $r > 0$ small enough, as required. □

3. Examples and applications

The first aim of this section is to give some examples illustrating the Assumptions needed in Theorem 1.1. Let us observe that the standard quadratic cost function: $c\colon \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}$, $c(x, p) = \frac{1}{2}|x - p|^2$ satisfies Assumptions 1.1, 1.2 and 1.3.

The next example, the case of the linear cost function $c(x, p) = |x - p|$, shows that second assumption (Twist Condition) is necessary. In this case the first and the third assumptions (and not the second one) hold and there exists a c-concave function whose singular set is of full dimension.

Example 3.1. Consider the following cost function

$$\begin{aligned} c\colon \mathbb{R}^n \times \mathbb{R}^n &\longrightarrow \mathbb{R} \\ (x, p) &\longmapsto |x - p|. \end{aligned}$$

This function is continuous and, for every $p \in \mathbb{R}^n$ fixed, $x \longmapsto |x - p|$ is differentiable except for the point $x = p$. This shows Assumption 1.1. Moreover, for $x \neq p$ we have that $\nabla_x c(x, p) = \frac{x-p}{|x-p|}$. Since the only discontinuity point of this map is at $x = p$ it easy to see that Assumption 1.3 is also satisfied. On the other hand we can observe that, $\nabla_x c(x, \cdot)$ is not injective, i.e. there exist x, p_1 and $p_2 \in \mathbb{R}^n$ such that $p_1 \neq p_2$ but

$$\nabla_x c(x, p_1) = \nabla_x c(x, p_2).$$

Let us fix $x = 0$ and consider $y \neq x$. We have that

$$\nabla_x c(0, p) = -\frac{p}{|p|}.$$

Now, if $p_1 = \lambda p_2$, for some $\lambda \in \mathbb{R}^+$, then

$$\nabla_x c(0, p_1) = \nabla_x c(0, \lambda p_2) = -\frac{\lambda p_2}{\lambda |p_2|} = \nabla_x c(0, p_2).$$

It is clear that this fact conflicts with Assumption 1.2.

Let us now indicate a c-concave function $u : \mathbb{R}^n \longrightarrow \mathbb{R}$ such that $\dim \Sigma = n$. Let $x_0 \in \mathbb{R}^n$ be fixed and consider the function

$$\psi(p) = |p - x_0|.$$

We write the respective c-concave function

$$u(x) = \inf_{p \in \mathbb{R}^n} (|x - p| - |p - x_0|). \tag{24}$$

By the triangle inequality, it holds that

$$|x - p| - |p - x_0| \geq -|x - x_0|.$$

This implies that the infimum in (24) is achieved when $p = x$. Therefore

$$u(x) = -|x - x_0|.$$

We prove now that, for every $x \in \mathbb{R}^n$, it holds that

$$\#\partial^c u(x) > 1.$$

Let us choose an arbitrary $x \in \mathbb{R}^n$. By definition of c-superdifferential, we have that $p \in \partial^c u(x)$ if and only if

$$|x - p| + |x - x_0| \leq |z - p| + |z - x_0|, \tag{25}$$

for every $z \in \mathbb{R}^n$.

Define now $p_t := x_0 + t(x - x_0)$ for $t \in \mathbb{R}$. We want to show that p_t satisfies (25) for every $t > 1$. Let us perform some calculations:

$$\begin{aligned} |x - p_t| + |x - x_0| &= |x - x_0 - t(x - x_0)| + |x - x_0| \\ &= t|x - x_0| = |t(x - x_0) + x_0 - x_0| \\ &= |p_t - x_0| \leq |p_t - z| + |z - x_0|, \end{aligned} \tag{26}$$

where the last inequality trivially follows from the triangular inequality and is true for each choice of $z \in \mathbb{R}^n$. This means that

$$\left\{ p_t \in \mathbb{R}^n \mid p_t = x_0 + t(x - x_0), \text{ with } t > 1 \right\} \subset \partial^c u(x),$$

and, since x was choosen arbitrarily, implies that

$$\Sigma(u) = \mathbb{R}^n.$$

The next example is a cost function that is simply the sum of two previously considered cost functions. As we shall see, all three Assumptions satisfied and so the statement of Theorem 1.1 applies.

Example 3.2. Consider the cost function

$$\begin{aligned} c\colon \mathbb{R}^n \times \mathbb{R}^n &\longrightarrow \mathbb{R} \\ (x, p) &\longmapsto |x-p| + |x-p|^2. \end{aligned}$$

Let us mention that $c(x, p) = |x - p|$ is the basic example of a function that is non semiconcave. The interested reader may check this fact or see [5]. As a consequence, we obtain that the above considered cost function $c(x, p) = |x - p| + |x - p|^2$ is not semiconcave either. This provides us with an example where the standard theory fails while the statement of Theorem 1.1 still applies.

In order to check the assumptions let us note first, that c is a continuous function and that, for every fixed $p \in \mathbb{R}^n$, $x \longmapsto c(x, p)$ is differentiable in $\mathbb{R}^n \setminus \{p\}$. Moreover, the validity of Taylor's formula (2) is easily verified.

Let us prove that c satisfies the Twist Condition. First of all we compute the gradient of the cost function with respect to the x variable:

$$\nabla_x c(x, p) = \frac{x-p}{|x-p|} + 2(x-p). \tag{27}$$

In order to proceed we need the following:

Claim. *Let v and w be two nonzero vectors in $\mathbb{R}^n$. Then*

$$\left| \left(v + \frac{v}{|v|} \right) - \left(w + \frac{w}{|w|} \right) \right| \geq |v - w|.$$

Proof. For simplicity of notation, we set $a\colon = |v|$ and $b\colon = |w|$. Therefore, $a + 1 = \left| v + \frac{v}{|v|} \right|$ and $b + 1 = \left| w + \frac{w}{|w|} \right|$. Now, it is well known that

$$|v - w| = \sqrt{a^2 + b^2 - 2ab\cos\gamma},$$

where γ is the angle between v and w on the plane spanned by the two vectors. Analogously, we can compute

$$\begin{aligned} \left| \left(v + \frac{v}{|v|} \right) - \left(w + \frac{w}{|w|} \right) \right| &= \sqrt{(a+1)^2 + (b+1)^2 - 2(a+1)(b+1)\cos\gamma} \\ &= \sqrt{a^2 + b^2 - 2ab\cos\gamma + 2(1+a+b)(1-\cos\gamma)} \\ &\geq \sqrt{a^2 + b^2 - 2ab\cos\gamma} = |v - w|, \end{aligned}$$

and this completes the proof of the claim. □

Fix $0 < \eta < 1$ and $x \in \mathbb{R}^n$. We aim to show that there exists $0 < \xi < 2$, dependent on η but not on x, such that, for every $p_1, p_2 \in \mathbb{R}^n$ so that $p_1 \neq x \neq p_2$ and $|p_1 - p_2| > \eta$, it follows

$$|\nabla_x c(x, p_1) - \nabla_x c(x, p_2)| > \xi.$$

It is here that we need to apply the Claim:

$$\begin{aligned}|\nabla_x c(x, p_1) - \nabla_x c(x, p_2)| &= \left|\left(\frac{x-p_1}{|x-p_1|} + 2(x-p_1)\right) - \left(\frac{x-p_2}{|x-p_2|} + 2(x-p_2)\right)\right| \\ &\geq |2(x-p_1) - 2(x-p_2)| = 2|p_1 - p_2| > 2\eta.\end{aligned}$$

If we choose $\xi < \eta$, we can conclude that Assumption 1.2 is satisfied.

Let us check Assumption 1.3. Fix $\varepsilon > 0$ and $\eta > 0$. Fix also $x \in \mathbb{R}^n$. We would like to find a suitable $\delta = \delta(\varepsilon, \eta) > 0$, not dependent on x, such that if p_1, p_1' and $p_2, p_2' \in \mathbb{R}^n$ are so that $|p_1 - p_1'| < \delta$, $|p_2 - p_2'| < \delta$, $|p_1 - p_2| > \eta$ and $x \notin \{p_1, p_1', p_2, p_2'\}$, then either

$$\left|\nabla_x c(x, p_1) - \nabla_x c(x, p_1')\right| < \varepsilon \tag{28}$$

or

$$\left|\nabla_x c(x, p_2) - \nabla_x c(x, p_2')\right| < \varepsilon. \tag{29}$$

We need to do some calculations, for $i = 1, 2$.

$$\begin{aligned}\left|\nabla_x c(x, p_i) - \nabla_x c(x, p_i')\right| &= b\left|\left(\frac{x-p_i}{|x-p_i|} + 2(x-p_i)\right) - \left(\frac{x-p_i'}{|x-p_i'|} + 2(x-p_i')\right)\right| \\ &= \left|\frac{x-p_i}{|x-p_i|} - \frac{x-p_i'}{|x-p_i'|} + 2(p_i - p_i')\right|.\end{aligned}$$

Let us fix $\delta = \frac{1}{16}\eta\varepsilon$ and check that it is a good choice. We can assume that $|x - p_2| > \frac{\eta}{4}$ and $|x - p_2'| > \frac{\eta}{4}$. If these inequalities are not satisfied, then we can switch and consider the case $|x - p_1| > \frac{\eta}{4}$ and $|x - p_1'| > \frac{\eta}{4}$. Let us prove why this is true. Assume that $|x - p_2| > \frac{\eta}{4}$ and $|x - p_2'| \leq \frac{\eta}{4}$. The cases with $|x - p_2| \leq \frac{\eta}{4}$ and $|x - p_2'| > \frac{\eta}{4}$, or with $|x - p_2| \leq \frac{\eta}{4}$ and $|x - p_2'| \leq \frac{\eta}{4}$, can be treated in the same way. First, we notice that, since $\varepsilon < 1$,

$$|p_1' - p_2| \geq |p_1 - p_2| - |p_1 - p_1'| \geq \eta - \frac{1}{16}\eta\varepsilon > \frac{15}{16}\eta. \tag{30}$$

Moreover, we can also write that

$$|p_1' - p_2'| \geq |p_1' - p_2| - |p_2 - p_2'| \geq \frac{15}{16}\eta - \frac{1}{16}\eta\varepsilon = \frac{7}{8}\eta. \tag{31}$$

Extimates (31) allow us to conclude that

$$|x - p_1'| \geq |p_1' - p_2'| - |x - p_2'| \geq \frac{7}{8}\eta - \frac{\eta}{4} = \frac{5}{8}\eta > \frac{\eta}{4}.$$

On the other hand, it holds also that

$$|x-p_1| \geq |p_1-p_2'|-|x-p_2'| \geq |p_1-p_2|-|p_2'-p_2|-|x-p_2'| > \eta-\frac{1}{16}\eta-\frac{\eta}{4} > \frac{\eta}{4}.$$

This implies that it is not loss of generality if we assume that

$$|x-p_2| > \frac{\eta}{4} \quad \text{and} \quad |x-p_2'| > \frac{\eta}{4}. \tag{32}$$

We are now ready to prove that, with our choice of δ, (29) holds. First of all, one has

$$
\begin{aligned}
|\nabla_x c(x,p_2)-\nabla_x c(x,p_2')| &= \left|\frac{x-p_2}{|x-p_2|}-\frac{x-p_2'}{|x-p_2'|}+2(p_2-p_2')\right| &(33)\\
&\leq \left|\frac{x-p_2}{|x-p_2|}-\frac{x-p_2'}{|x-p_2'|}\right|+2|p_2-p_2'| &(34)\\
&< \left|\frac{x-p_2}{|x-p_2|}-\frac{x-p_2'}{|x-p_2'|}\right|+\frac{1}{8}\eta\varepsilon. &(35)
\end{aligned}
$$

Writing $r=|x-p_2|$ and $r'=|x-p_2'|$, there exist two unit vectors v and $v'\in\mathbb{R}^n$ such that $p_2=x+rv$ and $p_2'=x+r'v'$. It holds that

$$\left|\frac{x-p_2}{|x-p_2|}-\frac{x-p_2'}{|x-p_2'|}\right| = |v-v'|.$$

Without loss of generality, we can assume that $r\leq r'$. We have

$$|p_2-p_2'| = |rv-r'v'| = |(r-r')v+r'(v-v')| \leq r'|v-v'|-(r-r').$$

This inequality implies the following extimate

$$
\begin{aligned}
|v-v'| &\leq \frac{r-r'}{r'}+\frac{|p_2-p_2'|}{r'} \leq \frac{4}{\eta}((r-r')+|p_2-p_2'|)\\
&\leq \frac{4}{\eta}\big(|x-p_2|-|x-p_2'|)+|p_2-p_2'|\big) \leq \frac{8}{\eta}|p_2-p_2'| < \frac{\varepsilon}{2}.
\end{aligned}
$$

Let us come back to (33). We have that

$$|\nabla_x c(x,p_2)-\nabla_x c(x,p_2')| < \left|\frac{x-p_2}{|x-p_2|}-\frac{x-p_2'}{|x-p_2'|}\right|+\frac{1}{8}\eta\varepsilon < \frac{\varepsilon}{2}+\frac{1}{8}\eta\varepsilon < \varepsilon,$$

which is exactly what we wanted.

In our final example we illustrate applications to our result to the case of sub-Riemannian type singular metrics. For illustrative purposes we restrict ourselves to the case of the first Heisenberg group [4]. Let us start with by introducing the following:

Notation 3.1. We denote by $\xi_1 : \mathbb{R}^3 \longrightarrow \mathbb{R}^2$, $\xi_1(x, y, t) = (x, y)$, the projection to the first components and by $\xi_2 : \mathbb{R}^3 \longrightarrow \mathbb{R}$, $\xi_2(x, y, t) = t$, the projection to the last variable.

Moreover, if we have two points $p = (x_1, y_1, t_1) \in \mathbb{R}^3$ and $q = (x_2, y_2, t_2) \in \mathbb{R}^3$, we introduce the Heisenberg group operation by

$$p^{-1} \cdot q = (x_2 - x_1, y_2 - y_1, t_2 - t_1 - 2(x_2 y_1 - x_1 y_2)) . \tag{36}$$

Finally, we introduce this function

$$\begin{aligned} N : \mathbb{R}^3 &\longrightarrow \mathbb{R} \\ (x, y, t) &\longmapsto \left(\left(x^2 + y^2\right)^2 + t^2\right)^{\frac{1}{4}}, \end{aligned}$$

that is the so-called *Korànyi norm* on the first Heisenberg group $\mathbb{H}^1$.

Example 3.3. We define a cost function as follows

$$\begin{aligned} c : \mathbb{R}^3 \times \mathbb{R}^3 &\longrightarrow \mathbb{R} \\ (p, q) &\longmapsto N^2(p^{-1} \cdot q) + \xi_2(q - p)^2. \end{aligned}$$

Our aim is to show that Theorem 1.1 can be applied for this cost function, which is not semiconcave. Clearly, the function $p \longmapsto c(p, q)$ is differentiable for every $p \in \mathbb{R}^3$ with $p \neq q$. Some comptaions that are left to the reader show that (2) also holds. This means that Assumption 1.1 is satisfied.

Let us prove that Assumption 1.2 is satisfied too. First, we show that the map

$$\begin{aligned} \mathbb{R}^3 \setminus \{q\} &\longrightarrow \mathbb{R}^3 \\ p &\longmapsto \nabla_q c(p, q) \end{aligned}$$

is injective for every $q \in \mathbb{R}^3$. Without loss of generality we can assume that $q = 0$. Hence, one has that, for $p = (x, y, t) \neq 0$,

$$\nabla_q c(p, 0) = \frac{1}{2} \frac{1}{N^2(p)} \left(4(x^2 + y^2)x - 4ty, 4(x^2 + y^2)y + 4tx, 2t\right) + (0, 0, 2t). \tag{37}$$

Now, let $p_1 = (x_1, y_1, t_1)$ and $p_2 = (x_2, y_2, t_2) \in \mathbb{R}^3 \setminus \{0\}$ be such that

$$\nabla_q c(p_1, 0) = \nabla_q c(p_2, 0). \tag{38}$$

We want to prove that $p_1 = p_2$. From formula (37), it is easy to see that, for $i = 1, 2$,

$$\left\| \xi_1 \left(\nabla_q c(p_i, 0)\right) \right\|_{\mathbb{R}^2} = 4\left(x_i^2 + y_i^2\right).$$

Hence, thanks to equality (38), it holds that $x_1^2 + y_1^2 = x_2^2 + y_2^2 : = k$. Consider now the third component of $\nabla_q c(p_i, 0)$. If we show that the function

$$f(t) = \frac{t}{(k^2 + t^2)^{\frac{1}{2}}} + 2t$$

is injective, then (38) implies that $p_1 = p_2$. This fact is clearly true, because f is strictly monotone, and it guarantees injectivity of the map $p \longmapsto \nabla_q(p, 0)$ on $\mathbb{R}^3 \setminus \{0\}$.

Let us consider Assumption 1.2. We argue by contradiction and we assume that there exist a compact set $K \subset \mathbb{R}^3$ and $0 < \eta < 1$ such that for every $n \in \mathbb{N}$ there are $p_1^{(n)}, p_2^{(n)} \in K$ and $q_n \in K$ so that $\left|p_1^{(n)} - p_2^{(n)}\right| > \eta$, $q_n \in C\left(p_1^{(n)}\right) \cap C\left(p_2^{(n)}\right)$ and

$$\left|\nabla_q c\left(p_1^{(n)}, q_n\right) - \nabla_q c\left(p_2^{(n)}, q_n\right)\right| \leq \frac{1}{n}. \tag{39}$$

Now, since $\left\{p_1^{(n)}\right\}_{n\in\mathbb{N}}, \left\{p_2^{(n)}\right\}_{n\in\mathbb{N}} \subset K$ and $(q_n)_{n\in\mathbb{N}} \subset K$ and K is a compact set, eventually restricting to subsequences, we can assume that there exist p_1^∞, p_2^∞ and $q_0 \in K$ such that $p_1^{(n)} \longrightarrow p_1^\infty$, $p_2^{(n)} \longrightarrow p_2^\infty$ and $q_n \longrightarrow q_0$, as $n \to \infty$. Without loss of generality, we can assume that $q_0 = 0$. We need to consider two cases:

(i) $p_1^\infty \neq 0$ and also $p_2^\infty \neq 0$;
(ii) $p_1^\infty = 0$ or $p_2^\infty = 0$.

In the first case, it holds that, for $i = 1, 2$,

$$\nabla_q c(p_i^{(n)}, q_n) \longrightarrow \nabla_q c(p_i^\infty, 0),$$

as $n \to \infty$. Hence, it follows that

$$\left|\nabla_q c(p_1^\infty, 0) - \nabla_q c(p_2^\infty, 0)\right| = 0,$$

but we have a contradiction with the injectivity, because $\left|p_1^\infty - p_2^\infty\right| > \eta$.

Consider now the second case. Assume that $p_1^\infty = 0$, the other case is totally analogous. By restricting to subsequence if needed, we can assume that the limit $v = \lim_{n\to\infty} \nabla_q c\left(p_1^{(n)}, 0\right)$ exists. From explicit formula (37), one can easily see that

$$v = \lim_{n\to\infty} \nabla_q c\left(p_1^{(n)}, 0\right) \in \{0\} \times \{0\} \times [-1, 1]. \tag{40}$$

Again, we have two cases $p_2^\infty = (0, 0, t_2^\infty)$ or $p_2^\infty = (x_2^\infty, y_2^\infty, t_2^\infty) \notin \{0\} \times \{0\} \times \mathbb{R}$.

In the first case, we notice that

$$\left|\xi_2\left(\nabla_q c(p_2^\infty, 0)\right)\right| = \left|\frac{t_2}{|t_2|} + 4t_2\right| > 1.$$

Therefore, $\nabla_q c(p_2^\infty, 0) \notin \{0\} \times \{0\} \times [-1, 1]$. This fact, and (40), provide a contradiction to (39) for large values of n.

On the other hand, when $p_2^\infty \neq \left(0, 0, t_2^\infty\right)$, we know that

$$\xi_1\left(\nabla_q c\left(p_2^\infty, 0\right)\right) = 4\left(\left(x_2^\infty\right)^2 + \left(y_2^\infty\right)^2\right) \neq 0.$$

Therefore, also in this case $\nabla_q c(p_2^\infty, 0) \notin \{0\} \times \{0\} \times [-1, 1]$. Again, we have a contradiction with (39) and (40).

It remains to prove that this cost function satisfies Assumption 1.3. Let $K \subset \mathbb{R}^3$ be a compact set and let $\varepsilon > 0$ and $\eta > 0$ be fixed. We want to find $\delta = \delta(\varepsilon, \eta) > 0$ such that if $p_1, p_2, p_1', p_2', q \in K$ are so that $|p_1 - p_2| > \eta$, $|p_i - p_i'| < \delta$, for $i = 1, 2$, and $q \in C(p_1) \cap C(p_1') \cap C(p_2) \cap C(p_2')$, then one among (4) and (5) holds.

Set $\delta_1 \le \frac{1}{16}\eta\varepsilon$. With this choice of δ_1, analogously to Example 3.2, we can assume that $|p_2 - q| \ge \frac{\eta}{4}$ and $|p_2' - q| \ge \frac{\eta}{4}$.

Let us denote

$$\Delta_\eta = \left\{(p, q) \in \mathbb{R}^3 \times \mathbb{R}^3 \;\middle|\; |p - q| < \frac{\eta}{4}\right\}$$

and consider the continuous function

$$\begin{aligned} F\colon K \times K \setminus \Delta_\eta &\longrightarrow \mathbb{R}^3 \\ (p, q) &\longmapsto \nabla_q c(p, q). \end{aligned}$$

Since F is a continuous function over a compact set, it is uniformly continuous. Therefore, there exists $\delta_2 = \delta_2(\varepsilon, \eta) > 0$ such that, if $|p - p'| < \delta_2$, then

$$\left|\nabla_q c(p, q) - \nabla_q c(p', q)\right| < \varepsilon.$$

Now, if we set $\delta = \min\{\delta_1, \delta_2\}$, Assumption 1.3 is satisfied.

As the main application of Theorem 1.1 we shall now formulate a result about the existence of optimal transport map with respect to a cost function satisfying Assumptions 1.1, 1.2, 1.3:

Theorem 3.1. *Let us assume that $c\colon \mathbb{R}^n \times \mathbb{R}^n \mapsto \mathbb{R}_+$ satisfies the Assumptions 1.1, 1.2, 1.3. Let μ and ν be two Borel regular probability measures such that μ does not give mass to $n - 1$ dimensional sets. Then there is an optimal transport map $\Psi\colon \mathbb{R}^n \mapsto \mathbb{R}^n$ transporting μ to ν i.e. $\Psi_\sharp \mu = \nu$ such that*

$$\int c(x, \Psi(x) d\mu(x) \le \int c(x, \Phi(x) d\mu(x)$$

for any measurable map $\Phi\colon \mathbb{R}^n \mapsto \mathbb{R}^n$ such that $\Phi_\sharp \mu = \nu$. Moreover, there exists a c concave function u such that $\Psi(x) = \partial^c(x)$ for μ a.e. x.

Proof. Based on the general theory of optimal mass transportation [14,15] there exists an optimal transport plan π with marginals μ and ν supported on the graph of the c-superdifferential $\partial^c u$ of some c-concave function u. If the the multivalued map $x \mapsto \partial^c u(x)$ is single valued for μ a.e. x then this will give rise to an optimal transport map defined by $\Psi(x) = \partial^c u(x)$ for μ almost every x.

According to the above consideration we only need to check that $x \mapsto \partial^c u(x)$ is single valued for μ almost every x. Clearly, $x \mapsto \partial^c u(x)$ is single valued outside the singular set $\Sigma(u)$. By Theorem 1.1 dim $\Sigma(u) \le n - 1$. By our assumption on the measure μ we have $\mu(\Sigma(u)) = 0$ and thus the claim of the theorem follows. □

Acknowledgements The authors thank the referee and Rita Pini for carefully reading the text and for pointing out mistakes in the earlier versions of this paper. We especially thank the referee for his comments related to the application of Taylor's formula in Assumption (1.1).

References

[1] Alberti, G., Ambrosio, L., Cannarsa, P.: On the singularities of convex functions. Manuscr. Math. **76**, 421–436 (1992)

[2] Alberti, G., Ambrosio, L.: A geometrical approach to monotone functions in $\mathbb{R}^n$. Math. Z **230**, 259–316 (1999)

[3] Ambrosio, L., Gigli, N., Savaré, G.: Gradient Flows in Metric Spaces and in the Space of Probability Measures, Lectures in Mathematics. Birkhäuser Verlag, Basel (2005)

[4] Ambrosio, L., Rigot, S.: Optimal mass transportation in Heisenberg group. J. Funct. Anal. **208**, 261–301 (2004)

[5] Cannarsa, P., Sinestrari, C.: Semiconcave Functions, Hamilton–Jacobi Equations, and Optimal Control. Birkäuser, Basel (2004)

[6] Figalli, A., Rifford, L.: Mass transportation on sub-Riemannian manifolds. Geom. Funct. Anal. **20**, 124–159 (2010)

[7] Fu, J.H.G.: Monge ampere functions. I. Indiana Math. J. **38**(3), 745–771 (1989)

[8] Gutierrez, C.E., Huang, Q.: The refractor problem in reshaping light beams. Arch. Ration. Mech. Anal. **193**(2), 423–443 (2009)

[9] Gutierrez, C.E., Huang, Q.: The near field refractor. Annales de l'Institut Henri Poincaré **31**(4), 655–684 (2014)

[10] Mattila, P.: Distribution of sets and measures along planes. J. Lond. Math. Soc. **38**, 125–132 (1988)

[11] Mattila, P.: Geometry of Sets and Measures in Euclidean Spaces. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge (1995)

[12] Reshetnyak, Y.G.: Generalized derivative and differentiability almost everywhere. Math. USSR Sbornik **4**, 293–302 (1968)

[13] Rifford, L.: Sub-riemannian Geometry and Optimal Transport. Springer Briefs in Mathematics. Springer, Cham (2014)

[14] Villani, C.: Topics in Optimal Transportation. AMS, Providence (2003)

[15] Villani, C.: Optimal Transport, Old and New, Grundlehren Math. Wiss, vol. 338. Springer, Cham (2008)