PARTIAL REGULARITY RESULTS IN OPTIMAL TRANSPORTATION

G. DE PHILIPPIS, A. FIGALLI

Abstract. This note describes some recent results on the regularity of optimal transport maps. As we shall see, in general optimal maps are not globally smooth, but they are so outside a "singular set" of measure zero.

1. The optimal transportation problem

The optimal transportation problem, whose origin dates back to Monge [19], aims to find a way to transport a distribution of mass from one place to another by minimizing the transportation cost. Mathematically, the problem can be formulated as follows: given two probability measures μ and ν (representing respectively the initial and final configuration of the mass that we want to transport) defined on the measurable spaces X and Y, one says that a map $T : X \to Y$ *transports* μ onto ν if $T_\sharp \mu = \nu$, i.e.,

$$\nu(A) = \mu\left(T^{-1}(A)\right) \qquad \forall\, A \subset Y \text{ measurable.}$$

Then, given a cost function $c : X \times Y \to \mathbb{R}$ (so that $c(x, y)$ represents the cost to transport a unit of mass from x to y), one wants to minimize the transportation cost among all possible transport maps.

Since transporting a unitary mass from x to $T(x)$ costs $c(x, T(x))$, the cost to transport the whole mass μ is simply given by $\int_X c(x, T(x))\, d\mu(x)$. Hence the optimal transportation problem consists in solving the minimization problem

$$\min_{T_\# \mu = \nu} \left\{ \int_X c(x, T(x))\, d\mu(x) \right\}. \tag{1}$$

When $T : X \to Y$ minimizes the transportation cost we call it an *optimal transport map.*

Even in Euclidean spaces with the cost c given by the Euclidean distance $|x - y|$ or its square $|x - y|^2$, the problem of the existence of an optimal transport map is far from being trivial. Moreover, it is easy to build examples where the Monge problem is ill-posed simply because there is no transport map: this happens for instance when μ is a Dirac mass while ν is not. This means that one needs some restrictions on the measures μ and ν.

We notice that when $X, Y \subset \mathbb{R}^n$, $\mu(dx) = f(x)dx$, and $\nu(dy) = g(y)dy$, if $T : X \to Y$ is a sufficiently smooth transport map one can rewrite the transport condition $T_\# \mu = \nu$ as a Jacobian equation. Indeed, if $\chi : \mathbb{R}^n \to \mathbb{R}$ denotes a test function, the condition $T_\# \mu = \nu$

gives

$$\int_{\mathbb{R}^n} \chi(T(x))f(x)\,dx = \int_{\mathbb{R}^n} \chi(y)g(y)\,dy.$$

Now, assuming in addition that T is a diffeomorphism, we can set $y = T(x)$ and use the change of variable formula to obtain that the second integral is equal to

$$\int_{\mathbb{R}^n} \chi(T(x))g(T(x))\big|\det(\nabla T(x))\big|\,dx.$$

By the arbitrariness of χ, this gives the Jacobian equation

$$f(x) = g(T(x))\,\big|\det(\nabla T(x))\big|. \tag{2}$$

1.1. **The quadratic cost on $\mathbb{R}^n$.** In [2, 3], Brenier considered the case $X = Y = \mathbb{R}^n$ and $c(x,y) = |x-y|^2/2$, and proved the following theorem (which was also obtained independently by Cuesta-Albertos and Matrán [8] and by Rachev and Rüschendorf [20]).

Theorem 1.1. *Let μ and ν be two compactly supported probability measures on $\mathbb{R}^n$. If μ is absolutely continuous with respect to the Lebesgue measure, then:*

- (i) *There exists a unique solution $\hat{T}$ to the optimal transport problem with cost $c(x,y) = |x-y|^2/2$.*
- (ii) *The optimal map $\hat{T}$ is characterized by the structure $\hat{T}(x) = \nabla u(x)$ for some convex function $u : \mathbb{R}^n \to \mathbb{R}$, which is called the "potential" associated to $\hat{T}$.*

Let us point out, for further use, that the minimization problem for the cost $|x-y|^2/2$ is equivalent to the the minimization problem for the cost $-x\cdot y$. Indeed, for any transport map T we have

$$\int_{\mathbb{R}^n} \frac{|T(x)|^2}{2}\,d\mu(x) = \int_{\mathbb{R}^n} \frac{|y|^2}{2}\,d\nu(y)$$

(this is a direct consequence of the condition $T_\#\mu = \nu$), hence

$$\begin{aligned}\int_{\mathbb{R}^n} \frac{|x-T(x)|^2}{2}\,d\mu(x) &= \int_{\mathbb{R}^n} \frac{|x|^2}{2}\,d\mu(x) + \int_{\mathbb{R}^n} \frac{|T(x)|^2}{2}\,d\mu(x) + \int_{\mathbb{R}^n} \big(-x\cdot T(x)\big)\,d\mu(x)\\ &= \int_{\mathbb{R}^n} \frac{|x|^2}{2}\,d\mu(x) + \int_{\mathbb{R}^n} \frac{|y|^2}{2}\,d\nu(y) + \int_{\mathbb{R}^n} \big(-x\cdot T(x)\big)\,d\mu(x),\end{aligned}$$

and since the first two integrals in the right hand side are independent of T we see that the two problems

$$\min_{T_\#\mu=\nu} \int_{\mathbb{R}^n} \frac{|x-T(x)|^2}{2}\,d\mu(x) \qquad \text{and} \qquad \min_{T_\#\mu=\nu} \int_{\mathbb{R}^n} \big(-x\cdot T(x)\big)\,d\mu(x)$$

are equivalent (that is, they have the same minimizers).

Having found a solution to (1), a natural question is the one concerning its *regularity*:

Assuming that X and Y are two bounded smooth open sets in $\mathbb{R}^n$, let $\mu(dx) = f(x)dx$ and $\nu(y) = g(y)dy$ be two probability measures with smooth densities f and g such that $X = \{f > 0\}$ and $Y = \{g > 0\}$. Then, is it true that the optimal map $\hat{T}$ (or equivalently the "potential" u) is smooth?

As observed by Caffarelli [6], one cannot expect any general regularity result for u without making some geometric assumptions on the support of the target measure. Indeed, let $n = 2$ and suppose that $X = B_1$ is the unit ball centered at the origin and $Y = \big(B_1^+ + e_1\big) \cup \big(B_1^- - e_1\big)$ is the union of two half-balls (here (e_1, e_2) denote the canonical basis of $\mathbb{R}^2$), where

$$B_1^+ := \big(B_1 \cap \{x_1 > 0\}\big), \qquad B_1^- := \big(B_1 \cap \{x_1 < 0\}\big).$$

Then, if $f = \frac{1}{|X|}\mathbf{1}_X$ and $g = \frac{1}{|Y|}\mathbf{1}_Y$, it is easily seen that the optimal map $\hat{T}$ is given by

$$\hat{T}(x) := \begin{cases} x + e_1 & \text{if } x_1 > 0, \\ x - e_1 & \text{if } x_1 < 0, \end{cases}$$

which corresponds to the gradient of the convex function $u(x) = |x|^2/2 + |x_1|$.

Thus (as one could also show by an easy topological argument) in order to hope for a regularity result for u we need at least to assume the connectedness of Y. However, not even this is sufficient. Indeed, starting from the above construction and considering a sequence of domains X'_ε where one adds a small strip of width $\varepsilon > 0$ to glue together $\big(B_1^+ + e_1\big) \cup \big(B_1^- - e_1\big)$, one can also show that for $\varepsilon > 0$ small enough the optimal map will still be discontinuous (see [6] or [25, Theorem 12.3] for more details).

In order to understand what is happening in the previous example, let us try to write down what is the equation satisfied by u. Since $\hat{T} = \nabla u$, the Jacobian equation (2) gives that u formally solves the Monge-Ampère equation

$$\det(D^2 u(x)) = \frac{f(x)}{g(\nabla u(x))} \tag{3}$$

coupled with the "boundary condition"

$$\nabla u(X) = Y \tag{4}$$

which corresponds to the fact that $\hat{T}$ transports $f(x)dx$ onto $g(y)dy$ (recall that $X = \{f > 0\}$ and $Y = \{g > 0\}$). So, one may in principle hope to apply the regularity theory for Monge-Ampère in order to show that u is actually smooth.

However this is just a formal computation, and what one can rigorously show is the following: the transport condition $(\nabla u)_\# f = g$ means that

$$\int_{(\nabla u)^{-1}(A)} f = \int_A g \qquad \forall\, A \subset Y.$$

From this fact it is possible to prove (see for instance [24, Lemma 4.6]) that

$$\int_E f = \int_{\partial u(E)} g \qquad \forall\, E \subset X,$$

where ∂u denotes the subdifferential of u:

$$\partial u(x) := \big\{p \in \mathbb{R}^n \,:\, u(z) \geq u(x) + p \cdot (z - x) \quad \forall\, z \in \mathbb{R}^n\big\}, \qquad \partial u(E) := \bigcup_{x \in E} \partial u(x).$$

Hence, since $Y = \{g > 0\}$ we get

$$\int_E f = \int_{\partial u(E) \cap Y} g \qquad \forall\, E \subset X,$$

and, if $\lambda \leq f, g \leq 1/\lambda$ on X and Y respectively, we deduce that [1]

$$\lambda^2 |E| \leq |\partial u(E) \cap Y| \leq |E|/\lambda^2 \qquad \forall\, E \subset X. \tag{5}$$

Hence, we can see this as a "weak" form of the Monge-Ampère equation.

On the other hand, the "right" notion of weak solution of (3) (i.e., a notion of solution which allows one to obtain a satisfactory regularity theory) is the one of *Alexandrov solution* [1]: we say that a convex function $u : X \to \mathbb{R}$ is an Alexandrov solution of

$$\lambda^2 \leq \det D^2 u \leq 1/\lambda^2 \tag{6}$$

if

$$\lambda^2 |E| \leq |\partial u(E)| \leq |E|/\lambda^2 \qquad \forall\, E \subset X. \tag{7}$$

Note that (6) is the condition which would follow from (3) when $\lambda \leq f, g \leq 1/\lambda$. Moreover, if $u \in C^2(X)$ then (6) implies (7), indeed in this case $\partial u = \nabla u$, hence by the Area Formula

$$|\partial u(E)| = |\nabla u(E)| = \int_E \det(D^2 u),$$

and (7) follows from (6).

The difference between (5) and (7) is at the base of the previous counterexample. Indeed, (7) provides enough control on the behavior of ∂u to show that if u is strictly convex [2] then $\partial u(x)$ is a singleton for all $x \in X$. By convexity of u, this implies that u is continuously differentiable in X, and actually one can also show that ∇u is Holder continuous (see [4, 5] and [10, Section 2.4]).

On the other hand, (5) only gives information on the behavior of the intersection $\partial u(E) \cap Y$ for $E \subset X$. In the counterexample above with $X = B_1$ and $Y = \big(B_1^+ + e_1\big) \cup \big(B_1^- - e_1\big)$, the potential u was given by $u(x) = |x|^2/2 + |x_1|$. Hence, for any x of the form $x = (0, x_2)$ the set ∂u is multivalued, namely $\partial u(x) = [-1, 1] \times \{x_2\}$. Thus

$$\partial u(\{0\} \times [-1, 1]) = [-1, 1]^2.$$

---

[1] Here and in the sequel, $|E|$ denotes the Lebesgue measure of a set E.

[2] By an example of Pogorelov this turns out to be a necessary condition, see for instance [24, Section 4.1.3].

This would not be possible if u satisfied (7) since ∂u has to map sets of measure zero onto sets of measure zero. However, since the intersection of $[-1,1]^2$ with $Y = \{g > 0\}$ has measure zero, one is not able to detect the singularity of u using (5).

Hence, in order to avoid this kind of counterexamples one should make sure that the target Y always covers the image of X through the subdifferential map ∂u. A way to ensure this is that Y is convex. Indeed (see for instance [6] or [10, Theorem 3.3]) if Y is convex then $\partial u(X) \subset \overline{Y}$ and (5) becomes (7). This information allows one to prove regularity [6, 7]. More precisely the following holds:

Theorem 1.2. *Let $X, Y \subset \mathbb{R}^n$ be two bounded open sets, let $f : X \to \mathbb{R}^+$ and $g : Y \to \mathbb{R}^+$ be two probability densities bounded away from zero and infinity on X and Y respectively, and denote by $\hat{T} = \nabla u : X \to Y$ the unique optimal transport map sending f onto g for the cost $|x-y|^2/2$. Assume that Y is convex. Then:*

(a) $\hat{T} \in C^{0,\alpha}_{\mathrm{loc}}(X)$.
(b) *If in addition $f \in C^{k,\beta}_{\mathrm{loc}}(X)$ and $g \in C^{k,\beta}_{\mathrm{loc}}(Y)$ for some $\beta \in (0,1)$, then $\hat{T} \in C^{k+1,\beta}_{\mathrm{loc}}(X)$.*
(c) *Furthermore, if $f \in C^{k,\beta}(\overline{X})$, $g \in C^{k,\beta}(\overline{Y})$, and both X and Y are smooth and uniformly convex, then $\hat{T} : \overline{X} \to \overline{Y}$ is a global diffeomorphism of class $C^{k+1,\beta}$.*

Even if this result is very satisfactory, one still would like to understand how "bad" can be the set where u is not regular when one removes the convexity assumption on the target. As shown in [14] (see also [12] for a more precise description of the singular set in two dimensions), in this case one can prove that the optimal transport map is actually smooth outside a closed set of measure zero. More precisely, the following holds:

Theorem 1.3. *Let $X, Y \subset \mathbb{R}^n$ be two bounded open sets, let $f : X \to \mathbb{R}^+$ and $g : Y \to \mathbb{R}^+$ be two probability densities bounded away from zero and infinity on X and Y respectively, and denote by $\hat{T} = \nabla u : X \to Y$ the unique optimal transport map sending f onto g for the cost $|x-y|^2/2$. Then there exist two relatively closed sets $\Sigma_X \subset X$ and $\Sigma_Y \subset Y$, with $|\Sigma_X| = |\Sigma_Y| = 0$, such that $\hat{T} : X \setminus \Sigma_X \to Y \setminus \Sigma_Y$ is a homeomorphism of class $C^{0,\alpha}_{\mathrm{loc}}$ for some $\alpha > 0$. In addition, if $c \in C^{k+2,\beta}_{\mathrm{loc}}(X \times Y)$, $f \in C^{k,\beta}_{\mathrm{loc}}(X)$, and $g \in C^{k,\beta}_{\mathrm{loc}}(Y)$ for some $k \geq 0$ and $\beta \in (0,1)$, then $\hat{T} : X \setminus \Sigma_X \to Y \setminus \Sigma_Y$ is a diffeomorphism of class $C^{k+1,\beta}_{\mathrm{loc}}$.*

Sketch of the proof. As explained above, when Y is not convex there could be points $x \in X$ such that $\partial u(x) \not\subset Y$. Let us define [3]

$$\mathrm{Reg}_X := \{x \in X : \ \partial u(x) \subset Y\} \qquad \Sigma_X := X \setminus \mathrm{Reg}_X.$$

By the continuity property of the subdifferential it is immediate to see that Reg_X is open. Moreover it follows from the condition $(\nabla u)_\sharp(f\,dx) = g\,dy$ that $\nabla u(x) \in Y$ for a.e. $x \in X$, thus $|\Sigma_X| = 0$. Hence

$$\lambda^2 |E| \leq |\partial u(E)| \leq |E|/\lambda^2 \qquad \forall\, E \subset \mathrm{Reg}_X$$

[3]Actually, in [12, 14] the regular set is defined in a slightly different way and it is in principle smaller. However, the advantage of that definition is that it allows for a more refined analysis of the singular set (see [12]).

provided $\lambda \leq f, g \leq 1/\lambda$. A (non-trivial) adaptation of Caffarelli's techniques permits to prove that u is smooth inside Reg_X (the main issue here is to show that u is strictly convex). □

1.2. **The case of a general cost.** After Theorem 1.1 many researchers started to work on the problem of showing existence (and regularity) of optimal maps in the case of more general costs, both in an Euclidean setting and in the case of Riemannian manifolds. Since, at least locally, any Riemannian manifold looks like $\mathbb{R}^n$, here we shall only focus on the Euclidean case (see [13] and [10] for more results).

Let us introduce first some conditions on the cost function $c : X \times Y \to \mathbb{R}$, where $X, Y \subset \mathbb{R}^n$:

- **(C0)** The cost function $c : X \times Y \to \mathbb{R}$ is of class C^2 with $\|c\|_{C^2(X\times Y)} < \infty$.
- **(C1)** For any $x \in X$, the map $Y \ni y \mapsto -D_x c(x, y) \in \mathbb{R}^n$ is injective.
- **(C2)** For any $y \in Y$, the map $X \ni x \mapsto -D_y c(x, y) \in \mathbb{R}^n$ is injective.
- **(C3)** $\det(D_{xy}c)(x, y) \neq 0$ for all $(x, y) \in X \times Y$.

We also introduce the concept of c-convex functions which generalizes the one of convex functions that appeared in the case $c(x, y) = -x \cdot y$ (see Theorem 1.1 and recall that, by the discussion immediately after that theorem, the costs $-x \cdot y$ and $|x - y|^2/2$ are equivalent): a function $u : X \to \mathbb{R} \cup \{+\infty\}$ is *c-convex* if it can be written as

$$u(x) = \sup_{y\in Y}\{-c(x, y) + \lambda_y\}$$

for some family of constants $\lambda_y \in \mathbb{R}$.

The following is a basic result in optimal transport theory.

Theorem 1.4. *Let $c : X \times Y \to \mathbb{R}$ satisfy* **(C0)-(C1)**. *Given two probability densities f and g supported on X and Y respectively, there exists a c-convex "potential" $u : X \to \mathbb{R}$ such that the map $\hat{T} : X \to Y$ implicitly defined by*

$$D_x c(x, \hat{T}(x)) + \nabla u(x) = 0 \tag{8}$$

is the unique optimal transport map sending f onto g.

Since c satisfies **(C1)** we can define the *c-exponential map*:

$$\text{for any } x \in X,\ y \in Y,\ p \in \mathbb{R}^n, \qquad \text{c-exp}_x(p) = y \quad \Leftrightarrow \quad p = -D_x c(x, y).$$

This allows us to rewrite (8) as $\hat{T}(x) = \text{c-exp}_x(\nabla u(x))$.

Let us try again to understand which PDE is satisfied by the "potential" u. Assuming that u is smooth, we see that the c-convexity of u implies that

$$D^2u(x) + D_{xx}c\big(x, \text{c-exp}_x(\nabla u(x))\big) \geq 0. \tag{9}$$

Moreover, using (**C2**) one can show that $\hat{T}$ is injective and that $\hat{T}^{-1}$ is the optimal map between ν and μ for the symmetrized cost $c^*(x, y) = c(y, x)$.

Hence, differentiating (8) with respect to x and using (2) and (9), we obtain

$$\det\Big(D^2u(x) + D_{xx}c\big(x, \text{c-exp}_x(\nabla u(x))\big)\Big) = \left|\det\left(D_{xy}c\big(x, \text{c-exp}_x(\nabla u(x))\big)\right)\right| \frac{f(x)}{g(\text{c-exp}_x(\nabla u(x)))}. \tag{10}$$

Hence, at least formally, u solves a Monge-Ampère type equation of the form

$$\det\big(D^2u - \mathcal{A}(x, \nabla u)\big) = h(x, \nabla u) \tag{11}$$

with

$$\mathcal{A}(x, p) := -D_{xx}c\big(x, \text{c-exp}_x(p)\big) \tag{12}$$

and

$$h(x, p) := \left|\det\left(D_{xy}c\big(x, \text{c-exp}_x(p)\big)\right)\right| \frac{f(x)}{g(\text{c-exp}_x(p))}.$$

Observe that, in the case $c(x, y) = -x \cdot y$, $\mathcal{A} \equiv 0$ and (11) reduces to the classical Monge-Ampère equation. As we showed in the previous section, in order to get regularity of u one needs to assume the convexity of the target domain. The issue is in some sense the following: the Monge-Ampère equation enjoys some a-priori regularity estimates (these are the so called Pogorelov estimates, see for instance [16, Section 17.6]) which allows one to obtain regularity of solutions provided one has suitable boundary conditions. In the case $c(x, y) = -x \cdot y$ the boundary condition was $\nabla u(X) = Y$ and convexity of Y was enough to ensure regularity.

Now, for the general case we have to face two difficulties: in addition to identify some suitable notion of convexity on the domains to handle the boundary conditions, one also needs some analogous of the Pogorelov estimates for the general class of equations (11).

The breakthrough for the regularity of solutions to this class of equations came with the paper of Ma, Trudinger and Wang [18], where the authors found a mysterious condition on the cost functions that turned out to be sufficient to prove the regularity of u. More precisely, the condition to be imposed on the cost (that we call here "MTW condition") is the following:

$$D^2_{p_\eta p_\eta}\mathcal{A}(x, p)[\xi, \xi] \leq 0 \qquad \forall\, x, p,\ \forall\, \xi \perp \eta. \tag{13}$$

Since $\mathcal{A}$ depends on first and second order derivatives of the cost (see (12)), the MTW condition is a fourth-order condition on c.

Under this condition, Ma, Trudinger, and Wang could prove the following result [18, 22, 23] (see also [21]), that generalizes Theorem 1.2(c):

Theorem 1.5. *Let $c : X \times Y \to \mathbb{R}$ satisfy* **(C0)**-**(C3)**. *Assume that the MTW condition holds, and that f and g are smooth and bounded away from zero and infinity on their respective supports X and Y. Also, suppose that:*

- *X and Y are smooth;*
- *$D_xc(x, Y)$ is uniformly convex for all $x \in X$;*
- *$D_yc(X, y)$ is uniformly convex for all $y \in Y$.*

Then $u \in C^\infty(\overline{X})$ *and the map* $T_u : \overline{X} \to \overline{Y}$ *defined as* $T_u(x) := \text{c-exp}_x(\nabla u(x))$ *is a smooth diffeomorphism.*

In [17] Loeper started a systematic study of the MTW condition and its relation to the geometry of optimal transport maps. Among other things, he was able to prove that the MTW condition (13) is essentially equivalent to the following (see [17], [25, Chapter 12] for a more precise discussion):
For any c-convex function u, its c-subdifferntial

$$\partial_c u(x) = \big\{ y \in \overline{Y} : u(z) + c(z,y) \geq u(x) + c(x,y) \qquad \forall\, z \in X \big\} \tag{14}$$

is connected *for every* $x \in X$.

Note that, when $c(x,y) = -x \cdot y$, c-convex function are just convex and $\partial_c u(x)$ reduces to $\partial u(x)$ (which is convex, thus connected).

Connectdness of the c-subdifferential turns out to be a necessary condition for the regularity of optimal maps, see [17], [25, Theorem 12.7], [15]. Hence, in view of its equivalence with the MTW condition, Loeper proved the following: if there exist $x, p, \xi \perp \eta$ such that the MTW condition fail, then one can construct smooth positive probability densities f and g (whose supports satisfy the appropriate global convexity assumptions) such that the optimal map between $\mu = f dx$ and $\nu = g dy$ is discontinuous. Moreover, Loeper also found a nice connection with geometry: if $c = d^2/2$ with d a Riemannian distance, then

$$D^2_{p_k p_\ell} A_{ij}(x,0)\xi^i \xi^j \eta^k \eta^\ell = -\frac{2}{3}\mathrm{Sect}_x([\xi,\eta]) \qquad \forall\, \xi, \eta \in T_x M,\ \xi \perp \eta,$$

where $\mathrm{Sect}_x([\xi,\eta])$ denotes the sectional curvature of the 2-plane generated by ξ and η. Since (as we just mentioned above) the MTW condition is necessary for regularity, one gets the following: [4]

Corollary 1.5.1. *Let* $c = d^2/2$ *on a smooth Riemannian manifold* M*, and assume that* $\mathrm{Sect}_x < 0$ *at some point along some 2-plane in* $T_x M$*. Then one can construct* $f, g \in C^\infty(M)$ *with* $f, g > 0$ *such that* $\hat{T} \notin C^0$.

Let us also mention that the MTW condition is quite restrictive and it is satisfied only in very particular cases. These include the costs:

- $|x-y|^2/2$ (or equivalently $-x \cdot y$);
- $-\log|x-y|$;
- $\sqrt{a^2 - |x-y|^2}$;
- $\sqrt{a^2 + |x-y|^2}$;
- $|x-y|^p$ with $-2 < p < 1$;

and the case $c = d^2/2$ on the following manifolds:

- $\mathbb{R}^n$ and $\mathbb{T}^n$;
- $\mathbb{S}^n$, its quotients (like $\mathbb{RP}^n$), and its submersions (like $\mathbb{CP}^n$ or $\mathbb{HP}^n$);

[4] Although we did not state them here, many existence and uniqueness result for optimal transport maps on Riemannian manifolds are known (see for instance [11]), and they include for instance the case $c(x,y) = d(x,y)^2/2$.

- products of any of the examples listed above (for instance, $\mathbb{S}^{n_1} \times \ldots \times \mathbb{S}^{n_k} \times \mathbb{R}^\ell$ or $\mathbb{S}^{n_1} \times \mathbb{CP}^{n_2} \times \mathbb{T}^{n_3}$);
- smooth perturbations of $\mathbb{S}^n$.

Because the MTW condition is usually false, a natural question is: Can one prove a partial regularity result for general cost functions?

Notice that in the case $-x \cdot y$ one exploits the fact that $\partial u(x) \subset \Lambda$ a.e. (see the sketch of the proof of Theorem 1.3), which means (very roughly) that at most points we are as in the case of a convex target, and hence a local regularity theory is in principle available. However, in the general case, besides the global obstruction given by the geometry of the source and target domains there is also the local obstruction given by the failure of the MTW condition. For instance, by Loeper's result, if M has negative sectional curvature then MTW fails *at every point*! This means that there is no hope to say that, as in the quadratic case, we are in a "good" situation at almost every point. Still, in [9] we have been able to prove the following result:

Theorem 1.6. *Let $X, Y \subset \mathbb{R}^n$ be two bounded open sets, and let $f : X \to \mathbb{R}^+$ and $g : Y \to \mathbb{R}^+$ be two continuous probability densities bounded away from zero and infinity on X and Y respectively. Assume that the cost $c : X \times Y \to \mathbb{R}$ satisfies* **(C0)**-**(C3)**, *and denote by $\hat{T} : X \to Y$ the unique optimal transport map sending f onto g. Then there exist two relatively closed sets $\Sigma_X \subset X$ and $\Sigma_Y \subset Y$, with $|\Sigma_X| = |\Sigma_Y| = 0$, such that $\hat{T} : X \setminus \Sigma_X \to Y \setminus \Sigma_Y$ is a homeomorphism of class $C^{0,\beta}_{\rm loc}$ for any $\beta < 1$. In addition, if $c \in C^{k+2,\alpha}_{\rm loc}(X \times Y)$, $f \in C^{k,\alpha}_{\rm loc}(X)$, and $g \in C^{k,\alpha}_{\rm loc}(Y)$ for some $k \geq 0$ and $\alpha \in (0,1)$, then $\hat{T} : X \setminus \Sigma_X \to Y \setminus \Sigma_Y$ is a diffeomorphism of class $C^{k+1,\alpha}_{\rm loc}$.*

This result can be suitably localized to obtain a regularity result for the squared distance function on Riemannian manifolds:

Theorem 1.7. *Let M be a smooth Riemannian manifold, and let $f, g : M \to \mathbb{R}^+$ be two continuous probability densities, locally bounded away from zero and infinity on M. Let $\hat{T} : M \to M$ denote the optimal transport map for the cost $c = d^2/2$ sending f onto g, d being the Riemannian distance on M. Then there exist two closed sets $\Sigma_X, \Sigma_Y \subset M$, with $|\Sigma_X| = |\Sigma_Y| = 0$, such that $\hat{T} : M \setminus \Sigma_X \to M \setminus \Sigma_Y$ is a homeomorphism of class $C^{0,\beta}_{\rm loc}$ for any $\beta < 1$. In addition, if both f and g are of class $C^{k,\alpha}$, then $\hat{T} : M \setminus \Sigma_X \to M \setminus \Sigma_Y$ is a diffeomorphism of class $C^{k+1,\alpha}_{\rm loc}$.*

Idea of the proof of Theorem 1.6. Let x_0 be a point where the potential u associated to $\hat{T}$ is twice differentiable (since u is c-convex, one can use Alexandrov's Theorem to show that u is twice differentiable at almost every point), set $y_0 := \hat{T}(x_0)$, and assume without loss of generality that $x_0 = y_0 = 0$. Then u looks like a parbola near zero, and up to subtracting a linear function we have

$$u(x) = Mx \cdot x + o(|x|^2).$$

We now observe that the cost $c(x, y)$ is equivalent to

$$\hat{c}(x, y) := c(x, y) - c(x, 0) - c(0, y) + c(0, 0).$$

Indeed, as in the quadratic case, for any transport map T we have

$$\begin{aligned}\int_{\mathbb{R}^n} \hat{c}(x, T(x))\, d\mu(x) &= \int_{\mathbb{R}^n} c(x, T(x))\, d\mu(x) - \int_{\mathbb{R}^n} c(x, 0)\, d\mu(x) \\ &\quad - \int_{\mathbb{R}^n} c(0, T(x))\, d\mu(x) + \int_{\mathbb{R}^n} c(0, 0)\, d\mu(x) \\ &= \int_{\mathbb{R}^n} c(x, T(x))\, d\mu(x) - \int_{\mathbb{R}^n} c(x, 0)\, d\mu(x) \\ &\quad - \int_{\mathbb{R}^n} c(0, y)\, d\nu(y) + c(0, 0),\end{aligned}$$

and the last three terms are independent of T.

So, without loss of generality we can assume that $\hat{c} = c$, and by Taylor's expansion we get

$$c(x, y) = Ax \cdot y + O(|x|^2|y| + |y|^2|x|).$$

Hence, up to applying the linear transformations $x \mapsto M^{1/2}x$ and $y \mapsto -M^{-1/2}A^*y$, we can assume that

$$u(x) = \frac{|x|^2}{2} + o(|x|^2)$$

and

$$c(x, y) = -x \cdot y + O(|x|^2|y| + |y|^2|x|)$$

near $(0, 0)$.

In addition, since f and g are continuous,

$$f(x) = f(0) + \omega(|x|), \qquad g(y) = g(0) + \omega(|y|),$$

for some modulus of continuity $\omega : \mathbb{R}^+ \to \mathbb{R}^+$.

By compactness we prove the following key result:

Lemma 1.7.1. *Assume*

$$\left| u(x) - \frac{|x|^2}{2} \right| \le \eta \quad \text{in } B_1,$$

$$\|c(x, y) + x \cdot y\|_{C^2(B_1 \times B_1)} \le \delta,$$

$$\|f - 1\|_{L^\infty(B_1)} + \|g - 1\|_{L^\infty(B_1)} \le \delta.$$

Then, provided $\eta > 0$ is universally small, there exists a modulus of continuity $\hat{\omega}$ such that

$$|u - \phi| \le \hat{\omega}(\delta) \quad \text{in } B_{1/2},$$

where $\nabla\phi$ is an optimal transport map for the quadratic cost between two constant densities. In addition

$$\|\phi\|_{C^3(B_{1/2})} \le C.$$

We apply the lemma as follows: first we rescale u, c, f, g once:

$$\psi(x) \mapsto u_1(x) := \frac{u(hx)}{h^2}, \qquad c(x,y) \mapsto c_1(x,y) := \frac{1}{h^2}c(hx,hy),$$
$$f(x) \mapsto f_1(x) := f(hx), \qquad g(x) \mapsto g_1(x) := g(hx)$$

for some $h \ll 1$.

Since

$$u(x) = \frac{|x|^2}{2} + o(|x|^2),$$

for h small we have

$$\left| u_1(x) - \frac{|x|^2}{2} \right| \leq \eta \quad \text{in } B_1.$$

Thus we can apply Lemma 1.7.1 with $\delta = \min\{\omega(h), Ch\}$ to obtain

$$|u_1 - \phi| \leq \overline{\omega}(h) \quad \text{in } B_{1/2}, \qquad \|\phi\|_{C^3(B_{1/2})} \leq C,$$

for some modulus of continuity $\overline{\omega}$. Let now $P(x) := \frac{1}{2}D^2\phi(0)x \cdot x$. Then

$$|\phi(x) - \phi(0) - \nabla\phi(0) \cdot x - P(x)| \leq Cr^3 \quad \text{in } B_r$$

for any $r \in (0, 1/2)$, therefore

$$|u_1 - \phi(0) - \nabla\phi(0) \cdot x - P(x)| \leq \overline{\omega}(h) + Cr^3 \quad \text{in } B_r.$$

We are now in position to iterate the rescaling argument: set

$$u_2(x) := \frac{u_1(rx) - \phi(0) - \nabla\phi(0) \cdot x}{r^2} \qquad c_2(x,y) := \frac{c_1(rx,ry) - \phi(0) - \nabla\phi(0) \cdot x}{r^2},$$
$$f_2(x) := f_1(rx), \qquad g_2(x) := g_1(rx)$$

Then, since $P(rx)/r^2 = P(x)$ we obtain

$$|u_2(x) - P(x)| \leq \frac{\overline{\omega}(h)}{r^2} + Cr \leq \eta \qquad \text{in } B_1$$

provided we choose first $r = r(\eta) \ll 1$ and then $h = h(r, \eta) \ll 1$. Let us observe that $P(x)$ is not exactly $|x|^2/2$ (as we would need to iterate the argument again), but we can show that it is of the form $Ax \cdot x$ for some symmetric matrix A satisfying $\lambda \mathrm{Id} \leq A \leq \Lambda \mathrm{Id}$ for some universal constants $0 < \lambda \leq \Lambda < \infty$. This is actually enough for us to keep iterating this argument and show that, for any $\alpha < 1$, there exists $C > 0$ such that

$$|u(x) - u(0) - \nabla u(0) \cdot x| \leq C|x|^{1+\alpha}.$$

Since this argument can be reapplied at any point near 0, we get $u \in C^{1,\alpha}$ in a neighborhood of 0.

This is the main step of the proof since it allows us to get rid of the local obstruction given by the failure of the MTW condition. Indeed, since u is C^1 near 0, recalling (14) it is easy to see that (for x in a neighborhood of 0)

$$\partial_c u(x) = \{\text{c-exp}(\nabla u(x))\},$$

in particular $\partial_c u(x)$ is connected. Relying on this, we can show that u enjoys a comparison principle, and this allows us to use a second approximation argument with solutions of the classical Monge-Ampère equation to conclude that u is $C^{2,\sigma'}$ in a smaller neighborhood for some $\sigma' > 0$. Then higher regularity follows from Schauder's theory.

These results imply that $\hat{T}$ is of class $C^{0,\beta}$ in neighborhood of $\bar{x}$ (resp. $\hat{T}$ is of class $C^{k+1,\alpha}$ if $c \in C^{k+2,\alpha}_{\rm loc}$ and $f, g \in C^{k,\alpha}_{\rm loc}$). Being our assumptions completely symmetric in x and y, we can apply the same argument to the optimal map T^* sending g onto f (here optimal means with respect to the cost $c^*(x, y) = c(y, x)$). Since $T^* = \hat{T}^{-1}$, it follows that $\hat{T}$ is a global homeomorphism of class $C^{0,\beta}_{\rm loc}$ (resp. $\hat{T}$ is a global diffeomorphism of class $C^{k+1,\alpha}_{\rm loc}$) outside a closed set of measure zero, concluding the proof. □

References

[1] Alexandrov, A. D. Existence and uniqueness of a convex surface with a given integral curvature. *C. R. (Doklady) Acad. Sci. URSS (N. S.)* 35 (1942), 131–134.

[2] Brenier, Y. Décomposition polaire et réarrangement monotone des champs de vecteurs. (French) *C. R. Acad. Sci. Paris Sér. I Math.* 305 (1987), no. 19, 805–808.

[3] Brenier, Y. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.* 44 (1991), no. 4, 375–417.

[4] Caffarelli, L.A. – A localization property of viscosity solutions to the Monge-Ampère equation and their strict convexity. *Ann. of Math. (2)* 131 (1990), no. 1, 129–134.

[5] Caffarelli, L.A. – Some regularity properties of solutions of Monge-Ampère equation. *Comm. Pure Appl. Math.* 44 (1991), no. 8-9, 965–969.

[6] Caffarelli, L.A. The regularity of mappings with a convex potential. *J. Amer. Math. Soc.* 5 (1992), no. 1, 99–104.

[7] Caffarelli, L.A. Boundary regularity of maps with convex potentials. II. *Ann. of Math. (2)* 144 (1996), no. 3, 453–496.

[8] Cuesta-Albertos, J.A.; Matrán, C. Notes on the Wasserstein metric in Hilbert spaces. *Ann. Prob.* 17 (1989), no. 3, 1264–1276.

[9] De Philippis, G.; Figalli, A. Partial regularity for optimal transport maps. Preprint, 2012.

[10] De Philippis, G.; Figalli, A. The Monge-Ampère equation and its link to optimal transportation. Preprint, 2013.

[11] Fathi, A; Figalli, A. – Optimal transportation on non-compact manifolds. *Israel J. Math.*, 175 (2010), 1–59.

[12] Figalli, A. Regularity properties of optimal maps between nonconvex domains in the plane. *Comm. Partial Differential Equations*, 35 (2010), no. 3, 465–479.

[13] Figalli, A. Regularity of optimal transport maps [after Ma-Trudinger-Wang and Loeper]. Séminaire Bourbaki. Vol. 2008/2009. Exposés 997Ð1011. *Astérisque*, 332 (2010), Exp. No. 1009.

[14] Figalli, A.; Kim, Y.-H. Partial regularity of Brenier solutions of the Monge-Ampère equation. *Discrete Contin. Dyn. Syst.*, 28 (2010), no. 2, 559–565.

[15] Figalli, A.; Rifford, L.; Villani, C. Necessary and sufficient conditions for continuity of optimal transport maps on Riemannian manifolds. *Tohoku Math. J. (2)*, 63 (2011), no. 4, 855–876.

[16] Gilbarg, D.; Trudinger, N. S. – *Elliptic Partial Differential Equations of Second Order.* Springer-Verlag, New York, 1983.

[17] Loeper, G. On the regularity of solutions of optimal transportation problems. *Acta Math.*, 202 (2009), no. 2, 241–283.

[18] Ma, X.-N.; Trudinger, N.S.; Wang, X.-J. Regularity of potential functions of the optimal transportation problem. *Arch. Ration. Mech. Anal.* 177 (2005), no. 2, 151–183.

[19] Monge, G. Mémoire sur la Théorie des Déblais et des Remblais. *Hist. de l'Acad. des Sciences de Paris* (1781), 666–704.

[20] Rachev, S.T.; Rüshendorf, L. *Mass Transportation Problems. Vol I: Theory, Vol II: Applications.* Probability and its applications. Spinger-Verlag, New York, 1998.

[21] Trudinger, N. S. A note on global regularity in optimal transportation. Preprint.

[22] Trudinger, N. S.; Wang, X.-J. On the second boundary value problem for Monge-Ampère type equations and optimal transportation. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)* 8 (2009), no. 1, 143–174.

[23] Trudinger, N. S.; Wang, X.-J. On strict convexity and continuous differentiability of potential functions in optimal transportation. *Arch. Ration. Mech. Anal.* 192 (2009), no. 3, 403–418.

[24] Villani, C. – *Topics in optimal transportation.* Graduate Studies in Mathematics, 58. American Mathematical Society, Providence, RI, 2003.

[25] Villani, C. – *Optimal transport, old and new.* Grundlehren des mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Vol. 338, Springer-Verlag, Berlin-New York, 2009.