STABILITY OF A 4TH-ORDER CURVATURE CONDITION ARISING IN OPTIMAL TRANSPORT THEORY

C. VILLANI

Abstract. A certain curvature condition, introduced by Ma, Trudinger and Wang in relation with the regularity of optimal transport, is shown to be stable under Gromov–Hausdorff limits, even though the condition implicitly involves fourth order derivatives of the Riemannian metric. Two lines of reasoning are presented with slightly different assumptions, one purely geometric, and another one combining geometry and probability. Then a converse problem is studied: prove some partial regularity for the optimal transport on a perturbation of a Riemannian manifold satisfying a strong form of the Ma–Trudinger–Wang condition.

Introduction

Stability properties of geometric invariants are one indicator of their flexibility and generality. Sometimes an invariant is stable under limit processes requiring much less structure than what was (apparently) used to define the invariant itself. A well-known example is the property of nonnegative (or nonpositive) sectional curvature, whose definition involves second-order derivatives of a metric tensor, but which is nevertheless stable under the purely metric notion of Gromov–Hausdorff convergence [1]. Another example which was studied by Lott and me [16], and independently by Sturm [18], is the stability of Ricci curvature lower bounds under measured Gromov–Hausdorff convergence.

In the present paper I shall consider an example which in some sense is even more striking since it will involve fourth-order derivatives of a metric tensor, and still there will be some stability under Gromov–Hausdorff convergence. Two schemes of proof will be presented, with slightly different results: the first one is purely metric, while the other one involves a probabilistic interpretation in terms of optimal transport. The second approach seems to be more robust, even though there is no measure theory in the original formulation of the problem.

In the last part of this paper, I shall consider the converse stability problem (stability under perturbation rather than under limit): What can be said, in terms of regularity of optimal transport, of a perturbation of a manifold satisfying the above-mentioned fourth-order condition?
This work is a prolongation of my collaborations with John Lott [16] on the one hand, and Grégoire Loeper [15] on the other hand. Various results and ideas from these works will be used directly or indirectly; and I shall also rely on the ideas of Loeper in [14]. Further thanks are due to Philippe Delanoë, Young-Heon Kim and Robert McCann for useful discussions; to Alessio Figalli for a careful reading of the manuscript; and to the Suzuki music school for an inspiring concert.

This paper is most respectfully dedicated to Paul Malliavin, whose work had a deep influence in reshaping the boundaries of probability, partial differential equations and differential geometry — those same boundaries that optimal transport is currently reshaping through the work of a large community of researchers.

1. Main results

The Ma–Trudinger–Wang curvature condition (MTW condition in short) was introduced a few years ago [17], as a key to the derivation of a priori smoothness estimates for solutions of optimal transport problems in a non-Euclidean setting. I shall consider it only in the particular (but arguably most important) case when the cost function is the square of the geodesic distance on a Riemannian manifold.

So let \( M \) be a smooth complete connected \( n \)-dimensional Riemannian manifold \((n \geq 2)\) with geodesic distance \( d \), and let \( c(x, y) = d(x, y)^2/2 \). I shall denote by \( T_x M \) the tangent space of \( M \) at \( x \), and \( \text{cut}(x) \) the cut locus of \( x \). Whenever \( x \) and \( y \) are two points in \( M \), with \( y \notin \text{cut}(x) \), pick up coordinate systems \((x_i)_{1 \leq i \leq n} \) and \((y_j)_{1 \leq j \leq n} \) around \( x \) and \( y \) respectively, and define a 4-tensor on \( T_x M \times T_y M \) by

\[
(1.1) \quad S_{(x,y)}(\xi, \eta) = \frac{3}{2} \sum_{ijklrs} \left( c_{ij,r} c_{r,s} c_{s,kt} - c_{ij,kt} \right) \xi^i \xi^j \eta^k \eta^l,
\]

where \( c_i = \partial c/\partial x_i \), \( c_{-j} = \partial c/\partial y_j \), \( c_{i,j} = \partial^2 c/\partial x_i \partial y_j \), etc., and \((c^{i,j})\) stands for the matrix elements of the inverse of \((c_{i,j})\).

The covariant nature of (1.1) has been established by Loeper [14], Kim and McCann [12]. In the latter reference it is shown that (1.1) is in fact the sectional curvature of \( M \times M \), equipped with the metric tensor \(-d^2_{x,y} c = -d_x d_y c\), along the plane generated by \((\xi, \eta)\). Note that \(-d^2_{x,y} c\) is not a Riemannian metric, indeed it has signature \((n, n)\). The generality of the underlying construction is discussed in [11]; it is actually (as I learnt through Robert Bryan and John Lott) an instance of para-Kähler structure [2, Section 2.2]. To simplify notation, I shall write \( \langle \cdot, \cdot \rangle \)
for the metric tensor \(-d_{x,y}^2\) \(c\), even though it is not Riemannian. Explicitly,

\[
\langle \xi, \eta \rangle = -(d_{x,y}^2) \cdot (\xi, \eta) = - \sum_{ij} c_{i,j} \xi^i \eta^j,
\]

and this also coincides with \(m_x((d_v \exp_x)^{-1} \eta, \xi)\), where \(m_x\) is the Riemannian metric at \(x\), \(\exp_x\) the Riemannian exponential map starting from \(x\), and \(v = (\exp_x)^{-1}(y) \in T_xM\). (By convention, \((\exp_x)^{-1}\) is the inverse of the exponential map restricted to its domain of injectivity, see below.)

The Ma–Trudinger–Wang condition MTW\((K_0)\) can be formulated as follows:

\[
\left[ \langle \xi, \eta \rangle = 0 \right] \implies \mathcal{S}_{(x,y)}(\xi, \eta) \geq K_0 |\xi|^2 |\tilde{\eta}|^2,
\]

where \(K_0 > 0\) (strong MTW condition) or \(K_0 = 0\) (weak MTW condition), and \(\tilde{\eta} = (d_v \exp_x)^{-1}(\eta)\). This condition automatically implies that all sectional curvatures (in the usual Riemannian sense) of \(M\) are bounded below by \(K_0\); in fact, \(\mathcal{S}_{(x,x)}(\xi, \eta)\) coincides with the sectional curvature at \(x\) along the plane \(\langle \xi, \eta \rangle\), if \(|\xi| = |\eta| = 1\) and \(\langle \xi, \eta \rangle = 0\) (this observation is due to Loeper [14] and was recast in [21, Particular Case 12.29]).

Even though \(\mathcal{S}_{(x,x)}\) only involves the sectional curvature, \(\mathcal{S}_{(x,y)}\) seems to be “genuinely fourth order” for \(x \neq y\); note in particular that fourth-order derivatives of the distance implicitly involve fourth-order derivatives of the metric tensor, via the third-derivative of the exponential map. The study by Delanoë and Ge [4] suggests that one needs a control on second-order derivatives of the Gaussian curvature to control \(\mathcal{S}\) close to the sphere.

There is by now plenty of evidence that these conditions play a key role in the regularity theory of optimal transport [10, 12, 13, 14, 15, 17, 19, 20]. Some of these results are reviewed in [21, Chapter 12]. Throughout this paper, I shall abbreviate MTW\((0)\) into just MTW.

My first goal here is to investigate the stability of this notion under the weak and popular notion of Gromov–Hausdorff convergence. By definition, a family of compact metric spaces \((\mathcal{X}_k, d_k)_{k \in \mathbb{N}}\) converges to some metric space \((\mathcal{X}, d)\) in Gromov–Hausdorff sense if there are (Borel measurable) maps \(f_k : \mathcal{X}_k \to \mathcal{X}\), called approximate isometries, such that

\[
\begin{align*}
\forall x, y \in \mathcal{X}_k, \quad \left| d(f_k(x), f_k(y)) - d_k(x, y) \right| & \leq \varepsilon_k, \\
\forall y \in \mathcal{X}, \quad \exists x \in \mathcal{X}_k, \quad d(f_k(x), y) & \leq \varepsilon_k,
\end{align*}
\]

(1.4)
where \((\varepsilon_k)_{k \in \mathbb{N}}\) is a sequence converging to 0. This notion does not preserve the local structure and allows for wild behavior such as reduction of dimension (collapse) at the limit.

A property which goes along well with the MTW condition is the convexity of tangent injectivity loci (CTIL condition for short), which I shall now explain. Let \(M\) be a compact Riemannian manifold, let \(x \in M\) and let \(U_x M\) stand for the unit sphere in \(T_x M\). For any \(\xi \in U_x M\), define \(t_C(\xi)\) as the first cut time of the geodesic \(\gamma(t) = \exp_x (t \xi)\), or equivalently the largest \(t_0\) such that \(\gamma\) is minimizing on \([0, t_0]\).

The tangent cut locus and tangent injectivity locus at \(x\) are subsets of \(T_x M\) defined respectively by

\[
\text{TCL}(x) = \{ t\xi; \; t = t_C(\xi); \; \xi \in U_x M \}; \\
\text{TIL}(x) = \{ t\xi; \; 0 \leq t < t_C(\xi); \; \xi \in U_x M \}.
\]

Note that \(\text{TCL}(x) = \partial \text{TIL}(x)\). I shall write \(\overline{\text{TIL}}(x) = \text{TIL}(x) \cup \text{TCL}(x) = \overline{\text{TIL}}(x)\).

I shall say that \(M\) satisfies the CTIL condition if all its tangent injectivity loci are convex.

The MTW and CTIL conditions occur together in the theory of regularity of optimal transport. Moreover, these conditions are not independent: there is reason to suspect that MTW automatically implies CTIL. It was actually proven in [15] that the strong MTW condition does imply CTIL, at least under a rather restrictive assumption of nonfocality of the cut locus, which means that \(d_{t\xi} \exp_x\) is invertible for \(t = t_C(\xi)\) (for any \(x\) and any \(\xi \in U_x M\)).

The first main result of the present paper is the following:

**Theorem 1.1** (Gromov–Hausdorff stability of MTW in presence of CTIL). Let \((M_k)_{k \in \mathbb{N}}\) be a sequence of compact Riemannian manifolds, each of them satisfying MTW and CTIL. Assume that \(M_k\) converges in Gromov–Hausdorff topology to some Riemannian manifold \(M\). Then also \(M\) satisfies MTW.

The combination of Theorem 1.1 with the results in [15] imply the

**Corollary 1.2.** Let \((M_k)_{k \in \mathbb{N}}\) be a sequence of compact Riemannian manifolds, each of them satisfying MTW(\(\delta_k\)) for some \(\delta_k > 0\), such that the cut locus of \(M_k\) is nonfocal. Assume that \(M_k\) converges in Gromov–Hausdorff topology to some Riemannian manifold \(M\). Then \(M\) satisfies MTW(0).

I shall explore two different strategies to prove the stability of MTW. The first one, purely geometric, relies on a synthetic reformulation of the MTW and CTIL
conditions, which has some common points with the Cartan–Alexandrov–Toponogov criterion for nonnegative sectional curvature [1, Chapter 4] — but one central element will be the notion of *bisector*, rather than median. This reformulation will allow to get the conclusion of Theorem 1.1 only under an additional assumption of *uniform upper bound on the sectional curvature*. On the other hand, the conclusion will be more complete than what is stated in Theorem 1.1, since $M$ itself will be shown to satisfy CTIL.

The second strategy relies on a reinterpretation of MTW in terms of properties of optimal transport of probability measures. So it does involve measure theory, not just geometry. This scheme of proof seems to be more efficient and will be enough to get the full generality of Theorem 1.1.

Theorem 1.1 leaves room for improvement. Once could conjecture that the CTIL condition is not explicitly needed, but automatically results from MTW. This striking conclusion would generalize the results in [15] and bring a lot of coherence in the subject, but does not seem easy to prove.

The stability is treated as follows. Section 2 is devoted to various reformulations of the CTIL and MTW conditions. In Section 3 I shall pass to the limit using elementary geometry; then in Section 4 I shall consider the strategy based on optimal transport.

A natural issue left open in this paper is the stability of the strong Ma–Trudinger–Wang condition, say MTW($K$) for $K > 0$. At present, there does not seem to be any convincing idea around to attack this problem.

Once the stability under limit has been explored, it is natural to investigate the “converse” problem of stability under perturbation: If a sequence $(M_k)_{k \in \mathbb{N}}$ converges in Gromov–Hausdorff topology to $M$ satisfying an MTW condition, does the approximating manifold inherit some properties related to that condition? In the last part of this paper I shall prove some results in this spirit, assuming however the strong MTW condition, and replacing the Gromov–Hausdorff convergence by stronger notions of convergence.

In particular, it will be shown in Section 6 that if a Riemannian manifold $M = (M, m)$ satisfies certain assumptions (including the strong MTW condition), and $M_k = (M, m_k)$, where $m_k$ is a $C^2$-approximation of $m$, then $M_k$ inherits from $M$ a coarse-grained version of the regularity theory of optimal transport which is typical of the strong MTW condition — at least under pointwise a priori bounds on the probability densities. This is interesting because $m_k$ is only assumed to approximate $m$ in $C^2$ topology, so it might not satisfy any sort of MTW condition. (Note that the $C^2$ convergence of the metric is not much stronger than the Gromov–Hausdorff...
convergence: under bounds from above and below on the sectional curvature, the noncollapsing Gromov–Hausdorff convergence implies $C^{1,\alpha}$ convergence for all $\alpha < 1$ [6, Theorem 4.1].

If the approximation is in $C^4$ topology there is a stability of MTW at least away from the cut locus, and more complete regularity results follow, see Section 5.

2. Synthetic reformulations

The problem addressed in this section is the reformulation of the MTW and CTIL conditions in terms of distances, lengths and angles. Let us start by recalling some of these concepts.

The length $L(\gamma)$ of a Lipschitz path $\gamma : [0, T] \to M$ is defined by the classical formula $L(\gamma) = \sup \sum_i d(\gamma(t_i), \gamma(t_{i+1}))$, where the sum extends over all nondecreasing subdivisions $(t_j)_{0 \leq j \leq J}$ of $[0, T]$. Minimizing geodesics are Lipschitz paths minimizing the length between their endpoints. Except otherwise stated, all geodesics in this paper will be parameterized by $[0, 1]$ and have constant speed, i.e. $d(\gamma(s), \gamma(t)) = L|t - s|$; or equivalently, $\gamma(s) = \exp_x(s \dot{\gamma}(0))$, where $\exp_x$ is the Riemannian exponential map starting from $x$.

If $\xi, \eta$ are two vectors in a Euclidean space, their (nonoriented) angle $\theta$ is defined by $\theta = \hat{\langle \xi, \eta \rangle} := \cos^{-1}\left(\frac{\langle \xi, \eta \rangle}{|\xi||\eta|}\right) \in [0, \pi]$.

If $\gamma_0$ and $\gamma_1$ are two geodesics originating from $x$ (i.e. $\gamma_0(0) = \gamma_1(0) = x$), the angle $\hat{\langle \gamma_0, \gamma_1 \rangle}$ between $\gamma_0$ and $\gamma_1$ (at $x$) is by definition the angle between the initial velocities $\dot{\gamma}_0(0)$ and $\dot{\gamma}_1(0)$ in $T_x M$. Angles can be expressed in purely metric terms with the help of the law of cosines [1, Proposition 3.6.27]:

$$\hat{\langle \gamma_0, \gamma_1 \rangle} = \lim_{s \to 0_+, t \to 0_+} \cos^{-1}\left(\frac{d(\overline{x}, \gamma_0(s))^2 + d(\overline{x}, \gamma_1(t))^2 - d(\gamma_0(s), \gamma_1(t))^2}{2 d(\overline{x}, \gamma_0(s)) d(\overline{x}, \gamma_1(t))}\right).$$

I shall now introduce two auxiliary functions $\lambda$ and $\mathcal{L}$ defined in terms of plane geometry. If $b, c > 0$ and $\alpha \in (0, \pi)$, let $\Delta = (ABC)$ be a triangle in $\mathbb{R}^2$ with sidelengths $AB = b$, $AC = c$ and angle $\alpha$ at $A$. ($\Delta$ is unique up to an isometry.) Whenever $\beta$ and $\gamma$ in $(0, \pi)$ satisfy $\beta + \gamma = \alpha$, define $\Gamma$ as the line segment originating from $A$, forming angles $\beta$ and $\gamma$ with the vectors $\overrightarrow{AB}$ and $\overrightarrow{AC}$ respectively. Let $D$ be the intersection of $\Gamma$ with $[B, C]$, and $\lambda = BD/BC$ (ratio of lengths), so that $\overrightarrow{AD} = (1 - \lambda) \overrightarrow{AB} + \lambda \overrightarrow{AC}$. Let further $\mathcal{L}(b, c, \beta, \gamma)$ be the length $\overrightarrow{AD}$. Both $\lambda$ and $\mathcal{L}$ can be computed (more or less explicitly) by invoking classical formulas of plane geometry.
geometry; for instance

\[ L(b, c, \beta, \beta) = \frac{2bc \cos \beta}{b+c}. \]

**Figure 1.** Definition of the functions \( \lambda \) and \( \mathcal{L} \) in terms of \( b, c, \beta, \gamma \).

**Lemma 2.1.** Let \( M \) be a Riemannian manifold and \( x \in M \), let \( v_0, v_1, v \) be three elements of \( T_xM \) and let \( t \in [0, 1] \). Write \( \gamma_0(s) = \exp_x(sv_0), \gamma_1(s) = \exp_x(sv_1), \gamma(s) = \exp_x(sv) \). Then the following two statements are equivalent:

(i) \( v = (1-t)v_0 + tv_1; \)

\[ (\widehat{\gamma_0}, \widehat{\gamma_1}) = (\widehat{\gamma_0}, \gamma) + (\gamma, \widehat{\gamma_1}) \]

(ii) \( \mathcal{L}(\gamma) = \mathcal{L}(L(\gamma_0), L(\gamma_1), \widehat{\gamma_0}, \widehat{\gamma_1}) \)

\[ t = \lambda(L(\gamma_0), L(\gamma_1), \widehat{\gamma_0}, \widehat{\gamma_1}). \]

The proof of Lemma 2.1 is more or less obvious once one has noticed that the angle condition \( (\widehat{\gamma_0}, \gamma) = (\widehat{\gamma_0}, \gamma) + (\gamma, \widehat{\gamma_1}) \) forces \( v \) to belong to the plane generated by \( v_0 \) and \( v_1 \). The following definition will save some words:

**Definition 2.2** (bisection). Let \( M \) be a Riemannian manifold and \( x \in M \). If \( \gamma_0 \) and \( \gamma_1 \) are two geodesics originating from \( x \) with an angle \( \theta \in (0, \pi) \), and \( \gamma \) is another
geodesic originating from \( x \), forming an angle \( \theta/2 \) with both \( \gamma_0 \) and \( \gamma_1 \) at \( x \), it is said that \( \gamma \) bisects \( \gamma_0 \) and \( \gamma_1 \) at \( x \).

After the above preparations, the proof of the following proposition will be easy:

**Proposition 2.3.** Let \( M \) be a Riemannian manifold and \( x \in M \). Then the following two properties are equivalent:

1. \( \text{TIL}(x) \) is convex in \( T_x M \);
2. For any two minimizing geodesics \( \gamma_0, \gamma_1 \) in \( M \), originating from \( x \) with some angle \( \theta \in (0, \pi) \), there is a minimizing geodesic \( \gamma \) bisecting \( \gamma_0 \) and \( \gamma_1 \) at \( x \), such that

\[
L(\gamma) \geq \mathcal{L}(L(\gamma_0), L(\gamma_1), \theta/2, \theta/2).
\]

**Remark 2.4.** The inequality in (2.2) can be replaced by an equality (just restrict \( \gamma \) if necessary).

**Proof of Proposition 2.3.** First note that the convexity of \( \text{TIL}(x) \) is equivalent to the convexity of \( \overline{\text{TIL}}(x) \).

Assume (i) holds true. Let \( \gamma_0, \gamma_1 \) be two minimizing geodesics originating from \( x \), with angle \( \theta \). Then \( \gamma_0(t) = \exp_x(tV_0\xi_0), \gamma_1(t) = \exp_x(tV_1\xi_1) \), where \( \xi_0, \xi_1 \in U_x M \), \( \xi_0 \cdot \xi_1 = \cos \theta \), \( 0 \leq V_0 \leq t_c(\xi_0), 0 \leq V_1 \leq t_c(\xi_1) \). Let \( \xi = (\xi_0 + \xi_1)/|\xi_0 + \xi_1| \); \( \xi \) forms an angle \( \theta/2 \) with \( \xi_0 \) and \( \xi_1 \), and belongs to the same plane as \( \xi_0, \xi_1 \). Let \( \ell = \mathcal{L}(V_0, V_1, \theta/2, \theta/2) \): the vector \( \ell \xi \) belongs to the segment \([V_0\xi_0, V_1\xi_1]\), which is entirely included in \( \overline{\text{TIL}}(x) \) because the latter set is convex and contains both \( V_0\xi_0 \) and \( V_1\xi_1 \). So \( \gamma(\ell \xi) = \exp_x(\ell \xi) \) is a minimizing geodesic with the desired properties.

Conversely, assume (ii). Pick up any two vectors \( v_0, v_1 \) in \( \overline{\text{TIL}}(x) \), and let \( \gamma_0(t) = \exp_x(tv_0), \gamma_1(t) = \exp_x(tv_1) \), let \( \theta \) be the angle between \( \xi_0 = v_0/|v_0| \) and \( \xi_1 = v_1/|v_1| \). By assumption there is \( v \) forming an angle \( \theta/2 \) with both \( v_0 \) and \( v_1 \), such that \( |v| \geq \mathcal{L}(v_0, |v_1|, \theta/2, \theta/2) \) and the geodesic \( \exp_x(tv)_{0 \leq t \leq 1} \) is minimizing. Since \( \langle v_0, v_1 \rangle = (v_0, v) + (v, v_1) \), \( v \) belongs to the plane generated by \( v_0 \) and \( v_1 \) (in fact \( v = (|v|/|v_0 + v_1|)(v_0 + v_1) \)), so \( v \) is an intermediate point between \( v_0 \) and \( v_1 \), and it belongs to \( \overline{\text{TIL}}(x) \). But now we can repeat the construction with \( v_0 \) replaced by \( v_{1/2} = v \), and find some \( v_{3/4} \in [v_{1/2}, v_1] \), belonging to \( \overline{\text{TIL}}(x) \), forming an angle \( 3\theta/4 \) with \( v_0 \) and \( \theta/4 \) with \( v_1 \). Continuing in this way, we can construct inductively a sequence \( (v_{j/2^{-k}}) \) \( (k \in \mathbb{N}, 0 \leq j \leq 2^k) \) of points in \([v_0, v_1]\), such that \( v_{j/2^{-k}} \) lies in \([v_0, v_1]\), belongs to \( \overline{\text{TIL}}(x) \), and forms an angle \( \theta j^{-2^{-k}} \) with \( v_0 \). This sequence is dense in \([v_0, v_1]\); so the whole segment \([v_0, v_1]\) lies in \( \overline{\text{TIL}}(x) \). This proves (i). \( \square \)

Next let us consider the (weak) Ma–Trudinger–Wang condition. Loeper [14] established striking consequences of this condition in terms of \( c\)-convex functions. By
definition, a function $\psi : M \to \mathbb{R}$ is said to be $c$-convex if there is $\zeta : M \to \mathbb{R} \cup \{-\infty\}$ such that
\[ \forall x \in M, \quad \psi(x) = \sup_{y \in M} \{ \zeta(y) - c(x, y) \}. \]
The $c$-transform of $\psi$ is then defined by
\[ \psi^c(y) = \inf_{x \in M} \{ \psi(x) + c(x, y) \}, \]
and the $c$-subdifferential of $\psi$ at $x$, or contact set of $\psi$ at $x$, is
\[ \partial_c \psi(x) = \{ y \in M; \ \psi^c(y) - \psi(x) = c(x, y) \}. \]
These definitions generalize the usual notions of convexity, Legendre transform and subdifferential (take $c(x, y) = -x \cdot y$) and play a crucial role in the theory of optimal transport [21, Chapter 5]. The “elementary” $c$-convex functions are defined by
\[ \psi_{x,y_0,y_1}(x) = \max \left( c(\overline{x}, y_0) - c(x, y_0), c(\overline{x}, y_1) - c(x, y_1) \right), \]
where $\overline{x}$, $y_0$ and $y_1$ are arbitrary in $M$. In the sequel, I shall only consider the case when $c(x, y) = d(x, y)^2/2$. Under MTW and CTIL, Loeper proved that $(\exp_x)^{-1}(\partial_c \psi(x) \setminus \text{cut}(x))$ is convex for any $c$-convex function $\psi$ and any $x \in M$; and this property is more or less characteristic of MTW. Moreover, it is equivalent to impose this for all $\psi$, or only for all $\psi_{x,y_0,y_1}$. For the latter class of functions, $(\exp_x)^{-1}(\partial_c \psi(x) \setminus \text{cut}(x))$ is included in a line segment, so its convexity is equivalent to its connectedness; thus MTW also comes close to be equivalent to the property of connectedness of contact sets, referred to as Assumption (C) in [21]. (Ma, Trudinger and Wang [17] were the first to suggest a link between their differential condition and this assumption.)

Below is a more precise statement which can be proven by pushing the arguments of Loeper. I shall say that a set $C$ is $L$-Lipschitz-connected if for any two $y_0, y_1$ in $C$ there is an $L$-Lipschitz path $[0, 1] \to C$ joining $y_0$ to $y_1$.

**Proposition 2.5** (Reformulations of MTW). Let $M$ be a Riemannian manifold satisfying the CTIL condition. Then the following statements are equivalent:

(i) $M$ satisfies the MTW condition;

(ii) For any $d^2/2$-convex function $\psi : M \to \mathbb{R}$ and any $\overline{x} \in M$, $(\exp_{\overline{x}})^{-1}(\partial_c \psi(\overline{x}) \setminus \text{cut}(\overline{x}))$ is convex;

(iii) For any $d^2/2$-convex function $\psi : M \to \mathbb{R}$, any $\overline{x} \in M$, and any $v_0, v_1 \in (\exp_{\overline{x}})^{-1}(\partial_c \psi(\overline{x}))$, let $y_0 = \exp_{\overline{x}} v_0$, $y_1 = \exp_{\overline{x}} v_1$, then
\[ \forall t \in [0, 1], \ \exp_{\overline{x}}((1-t)v_0 + tv_1) \in \partial_c \psi(\overline{x}); \]
(iv) For any $d^2/2$-convex function $\psi : M \rightarrow \mathbb{R}$ and any $x \in M$, $\partial_c \psi(x)$ is $L$-Lipschitz-connected, with $L = 2 \text{diam}(M)$;

(v) For any $d^2/2$-convex function $\psi : M \rightarrow \mathbb{R}$ and any $x \in M$, $\partial_c \psi(x)$ is connected.

Moreover, one may replace the arbitrary $d^2/2$-convex function $\psi$ in (ii)–(v) by $\psi_{x,y_0,y_1}$, where $y_0$ and $y_1$ are arbitrary; and the equivalence still holds.

Moreover, any of the statements (ii)–(v) implies (i) even without the CTIL condition.

Since these properties hold true in a more general context than for just $c = d^2/2$, I preferred to prove this statement (in a slightly lighter version) under more general assumptions in my book [21, Theorem 12.42]. In this reference, property (ii) is called the “regularity property”, property (v) is called “Assumption (C)”, and the CTIL property corresponds to the assumptions of “total $c$-convexity of $\text{Dom}'(\nabla_x c)$” and “total $\tilde{c}$-convexity of $\text{Dom}'(\nabla_x \tilde{c})$”. Assumptions (STwist) and (Cut$^{n-1}$) used in [21] are automatically satisfied for the squared geodesic distance on a Riemannian manifold, see [21, Appendix of Chapter 12]. Theorem 12.42 in [21] only asserts the equivalence of (i), (ii) and (v) under CTIL; but the other equivalences are easily deduced from the proof of that theorem. For the convenience of the reader I shall sketch the proof of Proposition 2.5 and refer to [21] for the technical details.

**Sketch of proof of Proposition 2.5.** First, a simple computation (as in the proof of [21, Proposition 12.14]) shows that if $\psi$ is any $d^2/2$-convex function and $y_0, y_1 \in \partial_c \psi(x)$, then $\partial_c \psi_{x,y_0,y_1}(x) \subset \partial_c \psi(x)$; so it becomes clear that $\psi$ can be replaced by $\psi_{x,y_0,y_1}$ in each of the statements (ii) to (v).

Next, (iii) is a generalization of (ii) where $v_0, v_1$ may belong to $\text{TIL}(x)$ rather than just $\text{TIL}(\overline{x})$. To deduce (iii) from (ii), it suffices to approximate $y_0$ and $y_1$ by $y_0^{(k)}$ and $y_1^{(k)}$ respectively, in such a way that $y_i^{(k)} \notin \text{cut}(x)$, and then pass to the limit as $k \rightarrow \infty$.

Then (iii) and (iv) trivially imply (v). (Let us forget for the moment the implication (iii) $\Rightarrow$ (iv).)

Now assume (v). Let $x \in M$ and let $y \in M \setminus \text{cut}(x)$. When $y_0, y_1$ converge to $y$, the functions $\psi_{x,y_0,y_1}$ converge uniformly to $\psi = \psi_{x,y,y}$, which is just a translate of $-d^2(y, \cdot)/2$. It follows easily that $\limsup_{y_0,y_1} \partial_c \psi_{x,y_0,y_1}(\overline{x})$ is included in $\partial_c \psi(\overline{x}) = \{y\}$ (the fact that $\partial_c \psi(\overline{x})$ is single-valued is a consequence of the differentiability of $\psi$ at $\overline{x}$). So $\partial_c \psi_{x,y_0,y_1}(\overline{x})$ is included in $M \setminus \text{cut}(x)$ if $y_0$ and $y_1$ are close enough to $y$. Then $(\exp_x)^{-1}(\partial_c \psi_{x,y_0,y_1}(\overline{x})) \subset \nabla^- \psi_{x,y_0,y_1}(\overline{x})$, where $\nabla^- \psi(x)$ is
the set of subgradients of \( \psi \) at \( \overline{x} \) [21, Theorem 10.25]. By general properties of semi-convex functions [21, Remark 10.51], \( \nabla^{-1}_{\overline{x}} \psi_{\overline{x},y_0,y_1}(\overline{x}) \) is the convex hull of limits of \( \nabla \psi_{\overline{x},y_0,y_1}(x_k) \) as \( x_k \to \overline{x} \), so it is just the convex segment \([v_0,v_1] \), where \( \exp_{\overline{x}} v_i = y_i \).

The set \( (\exp_{\overline{x}})^{-1}(\partial c \psi_{\overline{x},y_0,y_1}(\overline{x})) \) is connected by assumption and included in a segment, so it is convex. Then one can apply the strategy of [14] (based on Taylor expansion, and recast in [21, Theorem 12.36]) to prove the MTW property (i).

Once MTW is established, one can use the CTIL property and the strategy of Kim and McCann [10] (recast in the second implication of [21, Theorem 12.36]) to prove (ii).

Moreover, the condition MTW implies that \( M \) has nonnegative sectional curvature, so the exponential map \( \exp_{\overline{x}} \) is 1-Lipschitz, and then (iii) obviously implies (iv) since \( |y_0 - v_1| \leq 2 \text{diam}(M) \).

Properties (iv) or (v) in Proposition 2.5 are appealing because they do not involve the Riemannian structure of \( M \), and thus can be hoped to pass to the limit nicely. But even (iii) can be reformulated in “purely metric” terms:

**Proposition 2.6** (Further reformulation of MTW). Let \( M \) be a Riemannian manifold satisfying CTIL. Then \( M \) satisfies MTW if and only if the following property holds true: Whenever two geodesics \( \gamma_0 \) and \( \gamma_1 \) originate from \( \overline{x} \in M \) with some angle \( \theta \in (0, \pi) \), there is a minimizing geodesic \( \gamma \) bisecting \( \gamma_0 \) and \( \gamma_1 \) at \( \overline{x} \), having length \( L(\gamma) = L(L(\gamma_0), L(\gamma_1), \theta/2, \theta/2) \), such that if \( y = \gamma(1), y_0 = \gamma_0(1), y_1 = \gamma_1(1) \) then

\[
(2.4) \quad \forall x \in M \quad d(\overline{x}, y)^2 - d(x, y)^2 \leq \max \left( d(\overline{x}, y_0)^2 - d(x, y_0)^2, d(\overline{x}, y_1)^2 - d(x, y_1)^2 \right).
\]

**Remark 2.7.** Inequality (2.4) relates distances between certain configurations of five points \( \overline{x}, x, y_0, y_1, y \); compare with the Cartan–Alexandrov–Toponogov characterization of nonnegative (sectional) curvature, which involves four points [1, Chapter 4]. Since MTW implies nonnegative sectional curvature, it is not absurd to think that this implication may be seen at the level of (2.4). (At least when \( d(\overline{x}, y_0) = d(\overline{x}, y_1) \) I would guess that the Cartan–Alexandrov–Toponogov characterization can be retrieved by choosing \( x = y \).)

**Proof of Proposition 2.6.** First note that (2.4) is equivalent to \( y \in \partial c \psi_{\overline{x},y_0,y_1}(\overline{x}) \). Since \( \gamma \) in Proposition has length exactly \( L(L(\gamma_0), L(\gamma_1), \theta/2, \theta/2) \), \( y \) in the statement of Proposition 2.6 coincides with \( \exp_x ((1-t) \dot{\gamma}_0(0) + t \dot{\gamma}_1(0)) \) for the particular choice \( t = \lambda(L(\gamma_0), L(\gamma_1), \theta/2, \theta/2) \). So the property in Proposition 2.5(iii) does imply the property in Proposition 2.6.

The converse is not immediate because in Proposition 2.6 there is only one \( t \in [0,1] \), while we need them all in Proposition 2.5(iii). Let us write \( y_{1/2} = y, \gamma_{1/2} = \gamma \),
so (2.4) means \( y_{1/2} \in \partial_c \psi_{x,y_0,y_1}(\bar{x}) \). This implies that \( \partial_c \psi_{x,y_{1/2},y_1}(\bar{x}) \subset \partial_c \psi_{x,y_0,y_1}(\bar{x}) \). Applying the property again, but this time with \( y_0 \) replaced by \( y_{1/2} \), we construct a geodesic \( \gamma_{3/4} \) forming angles \( 3\theta/4 \) and \( \theta/4 \) with \( \gamma_0 \) and \( \gamma_1 \) respectively, with length at least \( \mathcal{L}(\gamma_{1/2}), \mathcal{L}(\gamma_1), \theta/4, \theta/4) = \mathcal{L}(\gamma_0), \mathcal{L}(\gamma_1)), 3\theta/4, \theta/4) \), such that \( y_{3/4} = \gamma_{3/4}(1) \) belongs to \( \partial_c \psi_{x,y_{1/2},y_1}(\bar{x}) \), and therefore to \( \partial_c \psi_{x,y_0,y_1}(\bar{x}) \). Repeating the process again and again, we can construct a sequence \( (v_{j2^{-k}}) (k \in \mathbb{N}, 0 \leq j \leq 2^k) \) such that \( \exp_{\bar{x}}(t v_{j2^{-k}}) \) belongs to \( \partial_c \psi_{x,y_0,y_1}(\bar{x}) \) and \( v_{j2^{-k}} \) forms angles \( \theta j 2^{-k} \) and \( \theta(1 - j 2^{-k}) \) with \( v_0 \) and \( v_1 \) respectively. Then the statement in Proposition 2.5(iii) is obtained by passing to the limit as \( k \to \infty \). \( \square \)

3. PASSING TO THE LIMIT VIA ELEMENTARY GEOMETRY

In this section I shall establish the following variant of Theorem 1.1:

**Theorem 3.1.** Let \((M_k)_{k \in \mathbb{N}}\) be a sequence of compact Riemannian manifolds, each of them satisfying MTW and CTIL, and having sectional curvatures bounded above by \( C \), independently of \( k \). Assume that \( M_k \) converges in Gromov–Hausdorff sense to some Riemannian manifold \( M \). Then also \( M \) satisfies MTW and CTIL.

For that I shall use the following (very easy) lemma.

**Lemma 3.2.** Let \((M_k)_{k \in \mathbb{N}}\) be a sequence of compact Riemannian manifolds, converging in Gromov–Hausdorff topology to some limit \( M \), by means of approximate isometries \( f_k \). Let \( x \in M \) and \( x_k \in M_k \) such that \( f_k(x_k) \to x \). Let \( \gamma \) be a minimizing geodesic on \( M \) with \( \gamma(0) = x \). Then there are geodesics \( \gamma_k \) on \( M_k \) such that \( \gamma_k(0) = x_k \) and

\[
\sup_{0 \leq t \leq 1} d(f_k(\gamma_k(t)), \gamma(t)) \longrightarrow 0.
\]

**Proof of Lemma 3.2.** First consider the case when \( \gamma \) is the unique (constant-speed, minimizing) geodesic between its endpoints. Let \( y = \gamma(1) \), let \( y_k \in M_k \) be such that \( d(f_k(y_k), y) \leq k^{-1} \), and let \( \gamma_k \) be a minimizing geodesic between \( x_k \) and \( y_k \). Up to extraction of a subsequence, \( f_k(\gamma_k(t)) \) converges uniformly to a minimizing geodesic joining \( x \) to \( y \), which is necessarily \( \gamma \).

If \( \gamma \) is not unique, for each \( \ell \in \mathbb{N} \) let \( \tilde{\gamma}_\ell \) be the restriction of \( \gamma \) to \([0, 1 - 1/\ell] \), reparameterized on \([0, 1] \), then \( \tilde{\gamma}_\ell \) is the unique minimizing geodesic between its endpoints and \( \tilde{\gamma}_\ell \) converges to \( \gamma \) uniformly. A diagonal extraction concludes the argument. \( \square \)

**Proof of Theorem 3.1.** Let \( M_k, M \) satisfy the assumptions of Theorem 1.1. Let \( \bar{x} \in M \) and let \( \gamma_0, \gamma_1 \) be two geodesics originating from \( \bar{x} \) with some angle \( \theta \). By definition
of Gromov–Hausdorff convergence we can find \( \overline{x}^{(k)} \) in \( M_k \) such that \( f_k(\overline{x}^{(k)}) \to x \). By Lemma 3.2 we can find a sequence \( (k_j)_{j \in \mathbb{N}} \) going to infinity (still denoted with an index \( k \) for simplicity), a sequence \( \eta_k \to 0 \), and geodesics \( \gamma_0^{(k)}, \gamma_1^{(k)} \) in \( M_k \), originating from \( \overline{x}^{(k)} \), such that \( d(f_k(\gamma_0^{(k)}(t)), \gamma(t)) \leq \eta_k \) for all \( t \in [0, 1], i \in \{0, 1\} \).

Let \( d_k \) and \( d \) stand for the geodesic distances on \( M_k \) and \( M \) respectively. For any \( s, t \) in \([0, 1]\), we have

\[
|d_k(\gamma_0^{(k)}(t), \gamma_1^{(k)}(s)) - d(\gamma_0(t), \gamma_1(s))| \leq |d_k(\gamma_0^{(k)}(t), \gamma_1^{(k)}(s)) - d(f_k(\gamma_0^{(k)}(t)), f_k(\gamma_1^{(k)}(s)))| + d(f(\gamma_0^{(k)}(t)), \gamma_0(t)) + d(\gamma_1(s), f_k(\gamma_1^{(k)}(s)))
\leq \varepsilon_k + 2\eta_k \xrightarrow{k \to \infty} 0.
\]

So

\[
\alpha_k(t, s) := \cos^{-1} \left( \frac{d(\overline{x}, \gamma_0^{(k)}(s))^2 + d(\overline{x}, \gamma_1^{(k)}(t))^2 - d(\gamma_0^{(k)}(s), \gamma_1^{(k)}(t))^2}{2 d(\overline{x}, \gamma_0^{(k)}(s)) d(\overline{x}, \gamma_1^{(k)}(t))} \right)
\]
converges to

\[
\alpha(t, s) := \cos^{-1} \left( \frac{d(\overline{x}, \gamma_0(s))^2 + d(\overline{x}, \gamma_1(t))^2 - d(\gamma_0(s), \gamma_1(t))^2}{2 d(\overline{x}, \gamma_0(s)) d(\overline{x}, \gamma_1(t))} \right)
\]
as \( k \to \infty \).

By assumption each \( M_k \) satisfies MTW, in particular it has nonnegative sectional curvature; so \( \alpha_k \) is a nonincreasing function of \( t \) and \( s \) [1, Section 4]. Also the sectional curvature of \( M_k \) is uniformly bounded above, so the derivatives of \( \alpha_k \) with respect to \( t \) and \( s \) are uniformly bounded above. This is sufficient to deduce that

\[
\lim_{s, t \to 0} \alpha_k(t, s) = \lim_{s, t \to 0} \alpha(t, s).
\]

So the angle \( \theta_k \) formed by \( \gamma_0^{(k)} \) and \( \gamma_1^{(k)} \) at \( \overline{x}^{(k)} \) converges to \( \theta \). (This is the same reasoning as in [1, Theorem 4.3.11]; in the absence of upper bound on the sectional curvature we would only get \( \theta \leq \lim \inf \theta_k \).)

Since \( M_k \) satisfies CTIL, by Proposition 2.3 we can find a geodesic \( \gamma^{(k)} \) bisecting \( \gamma_0^{(k)} \) and \( \gamma_1^{(k)} \) at \( \overline{x}^{(k)} \), such that

\[
L(\gamma^{(k)}) = L(L(\gamma_0^{(k)}), L(\gamma_1^{(k)}), \theta_k/2, \theta_k/2).
\]

Then up to extraction of a further subsequence, \( f_k(\gamma^{(k)}(t)) \) converges to a geodesic \( (\gamma(t))_{0 \leq t \leq 1} \) originating from \( \overline{x} \). By the same reasoning as above,

\[
\widetilde{\gamma}_0, \gamma = \lim_{k \to \infty} (\widetilde{\gamma}_0^{(k)}, \gamma^{(k)}) = \lim_{k \to \infty} \frac{\theta_k}{2} = \frac{\theta}{2}.
\]
Similarly, \((\overline{\gamma}, \gamma_1) = \theta/2\), so \(\gamma\) bisects \(\gamma_0\) and \(\gamma_1\) at \(\bar{x}\). Moreover,

\[
L(\gamma^{(k)}) = d(\gamma^{(k)}(0), \gamma^{(k)}(1)) \xrightarrow[k \to \infty]{} d(\gamma(0), \gamma(1)) = L(\gamma)
\]
as \(k \to \infty\); and similarly \(L(\gamma^{(k)}_i) \to L(\gamma_i)\) for \(i \in \{0, 1\}\). Thus

\[
L(\gamma) = L(L(\gamma_0), L(\gamma_1), \theta/2, \theta/2).
\]

Then by Proposition 2.3 again, \(M\) satisfies CTIL.

Now if \(x\) is any point in \(M\), we can find \(x^{(k)}\) in \(M_k\) such that \(f_k(x^{(k)}) \to x\). Let \(y_i = \gamma_i(1), y = \gamma(1), y^{(k)}_i = \gamma^{(k)}_i(1), y^{(k)} = \gamma^{(k)}(1)\); since \(M_k\) satisfies CTIL and MTW, by Proposition 2.6 we have

\[
d(y^{(k)}_i, y^{(k)})^2 - d(x^{(k)}, y^{(k)})^2
\]

\[
\leq \max\left(d(y^{(k)}_i, y^{(k)}_0)^2 - d(x^{(k)}, y^{(k)}_0)^2, d(y^{(k)}_1, y^{(k)}_1)^2 - d(x^{(k)}, y^{(k)}_1)^2\right).
\]
The limit \(k \to \infty\) yields

\[
d(\bar{x}, y)^2 - d(x, y)^2 \leq \max\left(d(\bar{x}, y_0)^2 - d(x, y_0)^2, d(\bar{x}, y_1)^2 - d(x, y_1)^2\right),
\]
so by Proposition 2.6 again, \(M\) satisfies MTW. \(\square\)

4. PASSING TO THE LIMIT VIA OPTIMAL TRANSPORT

This section is devoted to the proof of Theorem 1.1. Apart from results in the theory of optimal transport I shall use the following simple lemma:

**Lemma 4.1** (Gromov–Hausdorff stability of Lipschitz connectedness). Let \((X_k, d_k)_{k \in \mathbb{N}}\) be a sequence of compact metric spaces converging in Gromov–Hausdorff sense to a compact metric space \((X, d)\), by means of approximate isometries \(f_k\). Let \(L > 0\), and for each \(k\), let \(C_k\) be an \(L\)-Lipschitz-connected closed subset of \(X_k\). Then, up to extraction of a subsequence, \(\overline{f_k(C_k)}\) converges in the Hausdorff metric to an \(L\)-Lipschitz-connected closed set \(C\).

**Remark 4.2.** Of course the sets \(\overline{f_k(C_k)}\) in the above statement are not necessarily connected.

**Remark 4.3.** The lemma does not hold with \(L\)-Lipschitz-connectedness replaced by standard (pathwise) connectedness: think of the case when \(C_k\) is the graph of \(\sin(1/x)\) restricted to \([k^{-1}, 1]\).
Proof of Lemma 4.1. First of all, the sets \( f_k(C_k) \) are compact subsets of the compact metric space \( X \), so they form a precompact family in Hausdorff distance. After extraction of a subsequence, we have convergence of \( f_k(C_k) \) to some compact subset \( C \) of \( X \).

Let \( x, y \in C \), then there are sequences \((x_k)_{k \in \mathbb{N}}\) and \((y_k)_{k \in \mathbb{N}}\) such that \( x_k, y_k \in X \) and \( f_k(x_k) \to x \), \( f_k(y_k) \to y \). By assumption there is an \( L \)-Lipschitz path \( \gamma_k : [0, 1] \to C_k \) such that \( \gamma_k(0) = x_k \), \( \gamma_k(1) = y_k \). By diagonal extraction, there is a further subsequence (still denoted with index \( k \) for simplicity) such that \( f_k(\gamma_k(t)) \) converges for each \( t \in [0, 1] \cap \mathbb{Q} \), to some limit denoted by \( \gamma(t) \). By passing to the limit in the inequality

\[
d(f_k(\gamma_k(t)), f_k(\gamma_k(s))) \leq d_k(\gamma_k(t), \gamma_k(s)) + \varepsilon_k \leq L|t-s| + \varepsilon_k,
\]

we deduce that \( \gamma \) is \( L \)-Lipschitz, and in particular can be extended to the whole of \([0, 1]\). So \( C \) is \( L \)-Lipschitz-connected.

Proof of Theorem 1.1. Let \((M_k)_{k \in \mathbb{N}}\) and \( M \) satisfy the assumptions of the theorem. Let \( d \) be the geodesic distance on \( M \), \( c = d^2/2 \), let \( \psi : M \to \mathbb{R} \) be a \( c \)-convex function and let \( \phi = \psi^c \) be its \( c \)-transform. Let \( \pi \in M \) and \( y_0, y_1 \in \partial_c \psi(\pi) \) \((y_0 \neq y_1)\); the goal is to find a path joining \( y_0 \) to \( y_1 \), entirely contained in \( \partial_c \psi(\pi) \). Then the conclusion will follow from the equivalence \((i) \iff (v)\) in Proposition 2.5.

Let \( \tilde{\mu} \) be the normalized volume measure on \( M \) and let \( T(x) = \exp_x \nabla \psi(x) \). The map \( T \) is well-defined \( \tilde{\mu} \)-almost surely because \( \psi \) is Lipschitz; so \( \nu = T\# \tilde{\mu} \) (push-forward of \( \tilde{\mu} \) by \( T \)) is also well-defined.

Let \( \alpha, \beta_0, \beta_1 > 0 \) such that \( \alpha + \beta_0 + \beta_1 = 1 \) and \( \beta_0, \beta_1 > 1/3 \). Let

\[
\mu = \alpha \tilde{\mu} + (\beta_0 + \beta_1) \delta_y, \quad \nu = \alpha \tilde{\nu} + \beta_0 \delta_{y_0} + \beta_1 \delta_{y_1}.
\]

By construction, \( \pi = (\text{Id}_M, T)\# \tilde{\mu} + \beta_0 \delta_{y_0} + \beta_1 \delta_{y_1} \) is a coupling of \((\mu, \nu)\) whose support is entirely contained in the graph of \( \partial_c \psi \). By [21, Theorem 5.10], \( \pi \) is an optimal coupling in the transport problem from \( \mu \) to \( \nu \) (with cost \( c = d^2/2 \)), and \((\tilde{\psi}, \tilde{\phi})\) is optimal in the dual Kantorovich problem

\[
\sup_{\psi, \phi} \left\{ \int \phi' d\nu - \int \psi' d\mu; \quad \phi'(y) - \psi'(x) \leq c(x, y) \right\}.
\]

Moreover, since \( d\mu/d\text{vol} > 0 \) everywhere, \( \psi \) is the unique solution of this maximization problem, up to an additive constant [21, Remark 10.30]. (This observation is due to Loeper [14].)

Let now \( d_k \) be the geodesic distance on \( M_k \), and \( c_k = d_k^2/2 \). Let \( f_k^\# : M \to M_k \) be an approximate inverse of \( f_k \); this is a \((4\varepsilon_k)\)-isometry satisfying \( d_k(f_k^\# f(x), x) \leq 3\varepsilon_k \),
shall assume, without loss of generality, that \( a = \psi \) so that (i)

Statement (iv) in Proposition 2.5 implies the existence of a \( c_k \)-convex function \( \psi \) and a “mass-counting” argument, both \( (\pi_k, y_{0,k}) \) and \( (\pi_k, y_{1,k}) \) belong to the support of \( \pi_k \). Since the latter is included in \( \partial_{c_k} \psi_k \), and since \( M_2 \) satisfies MTW and CTIL, Statement (iv) in Proposition 2.5 implies the existence of an \( L \)-Lipschitz path \( \gamma_k : [0,1] \to \partial_{c_k} \psi_k(\pi_k) \), such that \( \gamma_k(0) = y_{0,k} \) and \( \gamma_k(1) = y_{1,k} \). (\( L = 2 \sup_j \text{diam}(M_j) \) will do.)

Next, the properties of the approximate inverse \( f' \) imply that

\[
(f_k)^\# \mu_k \xrightarrow{k \to \infty} \mu, \quad (f_k)^\# \nu_k \xrightarrow{k \to \infty} \nu,
\]

in the sense of weak convergence of measures. By stability of the optimal transport under Gromov–Hausdorff convergence (proven in [16] and recast as [21, Theorem 28.9]), we can extract a subsequence (still denoted with the index \( k \) for simplicity) such that \((f_k, f_k)^\# \pi_k \) converges weakly to an optimal transport plan \( \pi \) between \( \mu \) and \( \nu \).

Claim. Up to extraction of a subsequence, there are constants \( a_k \in \mathbb{R} \) such that \((\psi_k - a_k) \circ f' \) and \((\phi_k - a_k) \circ f' \) converge uniformly to a pair \((\psi', \phi')\) which is optimal in the Kantorovich problem between \( \mu \) and \( \nu \). Moreover, for any \( x \in M \),

\[
\limsup_{k \to \infty} f_k(\partial_{c_k} \psi_k(f_k^\#(x))) \subset \partial_{c'} \psi'(x).
\]

Remark 4.4. The short version of this claim is that the dual Kantorovich problem is stable under Gromov–Hausdorff convergence, and that \( c \)-subdifferentials are upper semicontinuous in the same process.

Proof of Claim. First let \( z \) be any element of \( X \), let \( z_k = f_k^\#(z) \) and let \( a_k = \psi_k(z_k) \), so that \((\psi_k - a_k) \circ f_k^\#(z) = 0 \). Being \( d_k^2 \)-2-convex, the function \( \psi_k \) is \( C \)-Lipschitz, where \( C \) is an upper bound on \( \sup_k \|d_k^2/\text{Lip}(M_k) \). So \( \psi_k - a_k \) is bounded, independently of \( k \); and then the formula \((\phi_k - a_k)(y) = \inf_z[(\psi_k - a_k)(z) + d_k(z, y)^2] \) implies that \( \phi_k - a_k \) is also bounded independently of \( k \). By a variant of Ascoli’s theorem (recalled in [16], stated without proof in [21, Proposition 27.20]), up to extraction of a subsequence the maps \((\psi_k - a_k) \circ f_k^\#(z) \) and \((\phi_k - a_k) \circ f_k^\#(z) \) converge uniformly to some \( \psi' : M \to \mathbb{R} \), \( \phi' : M \to \mathbb{R} \). In the sequel, I shall assume, without loss of generality, that \( a_k = 0 \).
For any \( x, y \) in \( M \) and any \( k \in \mathbb{N} \),
\[
\phi_k(f'_k(y)) - \psi_k(f'_k(x)) \leq \frac{d_k(f'_k(x), f'_k(y))^2}{2};
\]
passing to the limit \( k \to \infty \) yields
\[
(4.4) \quad \phi'(y) - \psi'(x) \leq \frac{d(x, y)^2}{2}.
\]
Now let again \( x \) be arbitrary in \( M \) and let \( y_k \in \partial_{c_k} \psi_k(f'_k(x)) \), then
\[
(4.5) \quad \phi_k(y_k) - \psi_k(f'_k(x)) = \frac{d_k(f'_k(x), y_k)^2}{2}.
\]
Up to extraction of a further subsequence, the sequence \( f_k(y_k) \) converges to some \( y = \overline{y} \in M \); then
\[
\phi_k(y_k) - \phi'(y) - \psi'(x) = d(y_k, f_k(y_k)).
\]
which converges to 0 as \( k \to \infty \). So one may pass to the limit in (4.5) and conclude that for any \( x \) there is \( y = \overline{y} \) satisfying
\[
\phi'(\overline{y}) - \psi'(x) = \frac{d(x, \overline{y})^2}{2}.
\]
This combined with (4.4) shows that \( \psi'(x) = \sup_y [\phi'(y) - c(x, y)] \), in particular \( \psi' \) is \( c \)-convex. At the same time, we have established (4.3) since \( \overline{y} \) could be any cluster point of \( f_k(\partial_{c_k} \psi_k(f'_k(x))) \).

Let \( W_2(\mu, \nu) \) be the Wasserstein distance of order 2 between \( \mu \) and \( \nu \), i.e. the square root of \( \inf \int d^2 d\pi \), where the infimum runs over all couplings \( \pi \) of \( \mu \) and \( \nu \). The optimality of \( (\psi_k, \phi_k) \) means exactly that
\[
(4.6) \quad \int \phi_k d\nu_k - \int \psi_k d\mu_k = \frac{1}{2} W_2(\mu_k, \nu_k)^2.
\]
From the definition of \( \mu_k \) and \( \nu_k \), the left-hand side can be rewritten as
\[
\int (\phi_k \circ f'_k) d\nu - \int (\psi_k \circ f'_k) d\mu,
\]
which obviously converges to \( \int \phi d\nu - \int \psi d\mu \). On the other hand, by stability of Wasserstein distance under Gromov–Hausdorff distance (proven in [16], recast in
\[ W_2(\mu_k, \nu_k) \text{ converges to } W_2(\mu, \nu). \]

In the end,
\[
\int \phi' \, d\nu - \int \psi' \, d\mu = \frac{1}{2} W_2(\mu, \nu)^2.
\]

This proves the optimality of \((\psi', \phi')\) and concludes the proof of the claim. \(\square\)

The end of the proof of Theorem 1.1 is easy. It follows from the Claim and the uniqueness of the solution of (4.1) that \(\psi' = \psi\) (up to some nonessential additive constant).

Now recall that \(\partial c_k \psi_k(\overline{\gamma}_k)\) contains an \(L\)-Lipschitz path \(\gamma_k\) joining \(y_{0,k}\) and \(y_{1,k}\). Since \(\overline{x}_k = f'_k(\overline{\gamma})\) we may apply (4.3) to deduce that
\[
(4.7) \quad \limsup_{k \to \infty} f_k(\gamma_k([0, 1])) \subset \partial c \psi(\overline{x}).
\]

Applying Lemma 4.1 with \(C_k = \gamma_k([0, 1])\), we find that the left-hand side of (4.7) is connected. Since it contains \(\lim f_k(y_{k,0}) = y_0\) and \(\lim f_k(y_{k,1}) = y_1\), the proof is complete. \(\square\)

5. Stability of the optimal transport map

In \([12, 14, 15, 17]\) strong versions of the MTW condition have been exploited to derive regularity estimates on the optimal transport map. When the cost function is the squared geodesic distance on a Riemannian manifold, this condition is however somewhat frigthening because, in constrast with the curvature, the distance is a nonlocal notion and is therefore quite difficult to compute. (See the discussion at the beginning of \([7]\) for some simple but illuminating remarks in this respect.) The situation with the MTW condition is even much worse since we need fourth order derivatives of the distance!

At the time when I am writing these lines, it has not been proved that these four derivatives are really necessary; as a matter of fact, when \(x = y\) the MTW tensor \(\mathcal{G}(x, y)\) reduces to sectional curvature, which of course is of second order in the metric. However, this might be due to some “algebraic” cancellation occurring only for \(x = y\); so I would bet that MTW is a genuine nonlocal fourth-order condition.

If one is working, say, on a numerical approximation of some optimal transport problem in curved geometry, there is hardly any hope to accurately approximate this fourth-order condition; so the stability of the above-mentioned regularity results under perturbation is a natural and relevant issue.

A first approach to this stability problem is the use of the implicit function theorem in the style of Delanoë \([3]\); this has been implemented by Delanoë and Ge \([4]\) near the
sphere, which satisfies the strong MTW condition. This strategy yields strong regularity estimates for the perturbed problem; on the other hand, it requires closeness in somewhat strong sense, typically $C^4$, and certain restrictions on the data. (How much one is allowed to perturb the geometry, depends on the probability measures that one is considering.)

An alternative approach is based on the pointwise stability of optimal transport under weak convergence of the probability measure and uniform convergence of the cost function. In general situations, this pointwise stability holds true only up to a set of small measure, by a Lusin argument [21, Corollary 5.23]. However, here we have much more structure and we know that the limiting transport is smooth! Using this information, we shall obtain a genuinely pointwise stability theorem.

Before giving such a statement, let me recall some results from [15]. First, the definition of uniform regularity:

**Definition 5.1 (Uniform regularity).** A Riemannian manifold $M$ is said to be uniformly regular if there are $\varepsilon_0, \kappa, \lambda > 0$ such that

(a) $\mathrm{TIL}(x)$ is $\kappa$-uniformly convex for any $x \in M$;

(b) For any $\overline{x} \in M$, let $(p_t)_{0 \leq t \leq 1}$ be a $C^2$ curve drawn in $\mathrm{TIL}(\overline{x})$, and let $y_t = \exp_{\overline{x}} p_t$; let further $x \in M$. If

$$\forall t \in (0, 1), \quad |\dot{p}_t| \leq \varepsilon_0 \, d(\overline{x}, x) \, |\dot{p}_t|^2, \quad |\dot{\overline{p}}_t| \leq 2 \, \mathrm{diam}(M),$$

then for any $t \in (0, 1)$,

$$d(x, y_t)^2 - d(\overline{x}, y_t)^2 \geq \min \left( d(x, y_0)^2 - d(\overline{x}, y_0)^2, \, d(x, y_1)^2 - d(\overline{x}, y_1)^2 \right)$$

$$+ 2 \lambda t(1 - t) \, d(\overline{x}, x)^2 \, |p_1 - p_0|^2.$$

The following notation will also be useful. I shall denote by $\mathrm{MTW}(K_0, C_0)$ the following precised version of (1.3):

$$\mathcal{G}(x,y)(\xi, \eta) \geq K_0 |\xi|^2 |\tilde{\eta}|^2 - C_0 \langle \xi, \tilde{\eta} \rangle |\xi| |\tilde{\eta}|,$$

where $\tilde{\eta} = (d_{\exp_x})^{-1}(\eta)$, $v = (\exp_x)^{-1}(y)$, and $(x, y)$ varies in $(M \times M) \setminus \text{cut}(M)$. Finally I shall introduce the following number measuring how much a manifold “avoids focal uniquely minimizing geodesics” (see [15, Appendix B.3]):

$$\delta(M) := \inf_{(x,v) \in \mathrm{TCL}(M)} \mathrm{diam}((\exp_x)^{-1}(\exp_x v)).$$

It is proven in [15] that the sphere is uniformly regular in the sense of Definition 5.1; and so is also any Riemannian manifold with nonfocal cut locus, satisfying a strong MTW condition.
The next result is a reformulation of Corollary 7.3 in [15]; I shall denote by $\exp$, $\nabla$, $\text{vol}$, $d$ respectively the exponential map, gradient, volume measure and geodesic distance associated to the Riemannian metric.

**Theorem 5.2** (Hölder regularity of optimal transport). If a Riemannian manifold $M$ is uniformly regular and satisfies $\delta(M) > 0$, then for any $A,a > 0$ and any probability measures $\mu, \nu$ on $M$ such that $\mu \leq A \text{vol}$, $\nu \geq a \text{vol}$, the optimal transport map between $\mu$ and $\nu$ takes the form $T = \exp(\nabla \psi)$, where

$$\|\psi\|_{C^{1,\alpha}(M)} \leq C(M,A,\alpha), \quad \alpha = \frac{1}{4n-1};$$

in particular,

$$d(T(x),T(y)) \leq C d(x,y)^\alpha.$$ 

As a consequence, one can derive the following pointwise stability theorem. I shall use self-explanatory notation, for instance if a metric $m_k$ depends on some index $k$, I shall denote by $\exp_k$, $\nabla_k$, $d_k$, $\sigma_k$ the associated exponential map, gradient, geodesic distance and sectional curvatures; moreover, the $C^r$ convergence of the metrics means $C^r$ convergence of their coefficients in some fixed system of local charts.

**Theorem 5.3** (Pointwise stability). Let $(M,m)$ be a uniformly regular Riemannian manifold with $\delta(M) > 0$. Let $(m_k)_{k \in \mathbb{N}}$ be a sequence of Riemannian metrics on $M$, such that $m_k \longrightarrow m$ in $C^1_{\text{loc}}(TM)$, and the sectional curvatures $\sigma_k$ are uniformly bounded below. For each $k$ let $\mu_k$, $\nu_k$ be probability measures on $M$ such that $\mu_k$ does not charge sets of Hausdorff dimension $n-1$, and let $T_k = \exp_k(\nabla_k \psi_k)$ be the optimal transport map from $\mu_k$ to $\nu_k$, for the cost function $c_k = d_k^2/2$, where $\psi_k$ is $c_k$-convex. Assume that $\mu_k \longrightarrow \mu = f \text{vol}$ and $\nu_k \longrightarrow \nu = g \text{vol}$ weakly, where $f$ is bounded above and $g$ bounded below by a positive number. Then as $k \longrightarrow \infty$, $\psi_k \longrightarrow \psi$ in $C^1(M)/\mathbb{R}$ and $T_k \longrightarrow T$ in $C(M)$, where $T = \exp(\nabla \psi)$ is the optimal transport map from $\mu$ to $\nu$, for the cost function $c = d^2/2$, and $\psi$ is $c$-convex.

The core of Theorem 5.3 is the following lemma, probably well-known in certain circles; its interest in the present context was pointed out to me by Loeper:

**Lemma 5.4.** Let $\varphi$ be a $C^1$ convex function $\mathbb{R}^n \rightarrow \mathbb{R}$, and let $(\varphi_k)_{k \in \mathbb{N}}$ be a sequence of convex functions converging pointwise to $\varphi$. Then $\varphi_k \longrightarrow \varphi$ in $C^1_{\text{loc}}(\mathbb{R}^n)$, in the sense that

$$\text{dist}(\partial \varphi_k(x), \nabla \varphi(x)) \longrightarrow 0 \quad \text{locally uniformly in } x.$$
**Sketch of proof of Lemma 5.4.** The stability of the subdifferential under pointwise convergence shows that if $x_k \to x$ then
\[ \limsup_{k \to \infty} \partial \varphi_k(x_k) \subset \partial \varphi(x) = \{ \nabla \varphi(x) \}; \]
in particular $\partial \varphi_k(x) \to \nabla \varphi(x)$ pointwise in $x$. Then the passage from pointwise to uniform convergence is because $\partial \varphi_k$ is a monotone map, which allows to use a multidimensional variant of Dini’s theorem. □

**Proof of Theorem 5.3.** First, $c_k = d_k^2/2$ is uniformly Lipschitz, so $\psi_k$ is uniformly Lipschitz too. Adding a nonessential constant to $\psi_k$, applying Ascoli’s theorem and extracting a subsequence if necessary, we may assume that $\psi_k$ converges uniformly to some $\tilde{\psi}$, which is $d^2/2$-convex. By the same reasoning as in the proof of the Claim in Section 4, $\tilde{\psi}$ solves the dual Monge–Kantorovich problem with measures $\mu, \nu$, and cost $c = d^2/2$; so $\tilde{\psi} = \psi$, of course up to an additive constant which without loss of generality will be set to zero.

From the bounds on the sectional curvature, $d_k^2/2$ is uniformly semiconcave [21, Third Appendix of Chapter 10], so the $\psi_k$ are uniformly semiconvex. By Theorem 5.2, $\psi$ lies in $C^1(M)$. By a localization argument and Lemma 5.4 (which obviously extends to uniformly semiconvex functions), the convergence of $\psi_k$ to $\psi$ actually holds in $C^1(M)$. So $d \psi_k \to d \psi$ (uniform convergence of the differential forms). This and the uniform convergence of $m_k$ to $m$ imply that $\nabla_k \psi_k \to \nabla \psi$ uniformly. Also the $C^1$ convergence of $m_k$ to $m$ implies that $\exp_k$ converges uniformly to $\exp$, so $T_k = \exp_k(\nabla_k \psi_k)$ converges uniformly to $T$. □

Here is a consequence of Theorem 5.3 allowing some perturbation of the regularity theory appearing in [14, 15]:

**Theorem 5.5 (Stability of regularity).** Let $(M, \overline{m})$ be a Riemannian manifold such that (a) all tangent injectivity loci in $M$ are uniformly convex; (b) $M$ satisfies $\text{MTW}(K_0, C_0)$ for some $K_0 > 0$, $C_0 < \infty$; (c) $\delta(M) > 0$. Then for any $A, a > 0$ there are $\eta = \eta(M, \overline{m}, A, a) > 0$ and $C = C(M, \overline{m}, A, a)$ such that if $m$ is another Riemannian metric on $M$, satisfying $\|m - \overline{m}\|_{C^4} \leq \eta$, and $\mu, \nu$ are two probability measures on $M$ satisfying $\mu \leq A \text{vol}$, $\nu \geq a \text{vol}$, then the optimal transport map $T = \exp(\nabla \psi)$ between $\mu$ and $\nu$ satisfies
\[ (5.6) \quad \|\psi\|_{C^{1,\alpha}(M)} \leq C. \]

It would seem that the proof of Theorem 5.5 requires a proof of the stability of $\text{MTW}(K_0, C_0)$ under $C^4$ perturbation of the Riemannian metric — a problem which seems really difficult because of the rough dependence of the distance on the metric
near the cut locus. However, as was pointed out to me by McCann, we do not need to worry about what happens close to the cut locus, because we work under pointwise a priori estimates. This observation and the pointwise stability (Theorem 5.3) are the two main ingredients of the proof; also useful will be the next lemma (which will be used again in Section 6):

**Lemma 5.6** (Stability of the cut locus under $C^2$ perturbation). Let $(M, m)$ be a compact Riemannian manifold and let $(m_k)_{k \in \mathbb{N}}$ be a sequence of Riemannian metrics converging to $m$ in $C^2$ topology. Let $x_k, y_k$ be sequences converging to $x$ and $y$ respectively. Then

(i) if $d_k(y_k, \text{cut}(x_k)) \geq \sigma > 0$ then $d(y, \text{cut}(x)) \geq \sigma'$ for some $\sigma' = \sigma'(M, m, \sigma) > 0$;

(ii) if $y_k \in \text{cut}(x_k)$ then $y \in \text{cut}(x)$. In particular,

$$\limsup_{k \to \infty} \text{cut}(x_k) \subset \text{cut}(x).$$

**Remark 5.7.** Statement (i) is probably improvable, but will be sufficient for our needs.

**Proof of Lemma 5.6.** (i) If $x = y$ the result is obvious. Otherwise let $L = d(x, y)$. Let $\gamma_k$ be the unique geodesic joining $x_k$ to $y_k$, parameterized by arc length, and let $L_k = d_k(x_k, y_k)$. From the assumption, the geodesic $\gamma_k$ can be extended into a minimizing geodesic on the time-interval $[0, L_k + \sigma]$. Up to extraction of a subsequence, $\gamma_k$ converges uniformly as $k \to \infty$ to some geodesic $\gamma$ defined on $[0, L + \sigma]$, with $\gamma(0) = x, \gamma(L) = y$. From this and the Lipschitz regularity of $\text{TI}_L(M)$ it follows that $y$ lies a distance $\sigma'$ away from $\text{cut}(x)$ (as in Appendix B of [15]).

(ii) The $C^2$ convergence of $m_k$ to $m$ implies the $C^1$ convergence of the exponential map $\exp_k$ to $\exp$. Take $(x_k, v_k) \in \text{TCL}(M, m_k)$. If $d_{x_k} \exp_{x_k}$ has zero Jacobian determinant this passes to the limit; otherwise we may assume that there is $\tilde{v}_k \neq v_k$ in $\text{TCL}(x_k)$ such that $\exp_{x_k} v_k = \exp_{x_k} \tilde{v}_k$ and $v_k \to v, \tilde{v}_k \to \tilde{v}$. If $v \neq \tilde{v}$ then $\exp_x$ is not injective at $v$, and if $v = \tilde{v}$ we may use the implicit function theorem to conclude that either $\exp_x$ is not injective at $v$, or $d_v \exp_x$ is not invertible. All in all, $v \in \text{TCL}(M, m)$, so $y = \exp_x v$ belongs to $\text{cut}(x)$. \qed

**Proof of Theorem 5.5.** If the property is false, we can find a sequence $(m_k)_{k \in \mathbb{N}}$ of Riemannian metrics, approximating $\overline{\mathfrak{m}}$ in $C^4$ topology, and probability measures $\mu_k \leq A \text{ vol}_k, \nu_k \geq a \text{ vol}_k$, such that the conclusion is violated. Without loss of generality, the probability measures $\mu_k$ and $\nu_k$ respectively converge to some measures $\mu \leq A \text{ vol}$ and $\nu \geq a \text{ vol}$, weakly as $k \to \infty$. 
By [15, Theorem 6.1], there is $\sigma = \sigma(M, m, A, a) > 0$ such that the optimal transport map $T$ between $\mu$ and $\nu$ stays at distance $\sigma$ from the cut locus: $d(T(x), \text{cut}(x)) \geq \sigma$. By Theorem 5.3 and the stability of the cut locus under $C^2$ perturbation of the metric (Lemma 5.6(ii)), we have, with obvious notation, $d_k(T_k(x), \text{cut}_k(x)) > \sigma/2$ for $k$ large enough.

It is not clear if the tangent injectivity loci $\text{TIL}_k(x)$ in $(M, m_k)$ will be uniformly convex, but at least for $k$ large enough the sets $(\exp_{k,x})^{-1}(O_k)$, where $O_k(x) = \{d_k(y, \text{cut}_k(x)) > \sigma/2\}$, will be contained in a uniformly convex subset of $\text{TIL}_k(x)$.

Recall that $(M, m)$ satisfies $\text{MTW}(K_0, C_0)$, which is a condition involving fourth-order derivatives of the distance. Away from the cut locus, the distance depends smoothly on the Riemannian metric, so for $k$ large enough $(M, m_k)$ will satisfy $\text{MTW}(K_0/2, 2C_0)$ at all $(x, y)$ such that $y \in O_k(x)$.

To summarize: In the transport problem from $\mu_k$ to $\nu_k$ with cost $c_k$, the transport takes place in a domain where a strict form of the Ma–Trudinger–Wang condition holds; and the transport stays away from the cut locus; and everything occurs in a uniformly convex portion of the tangent injectivity loci. This, together with [5, Lemma 1], makes it possible to re-do the proof of Theorem 7.1 in [15]. So $\psi_k$ satisfies $\|\psi_k\|_{C^{1,\alpha}(M, m_k)} \leq C$, where $C$ depends continuously on $m_k$, and also depends on $\sigma, A$ and $a$. This contradicts our initial assumption, and concludes the proof of Theorem 5.5. \[ \square \]

6. Approximation of strong MTW condition and regularity at mesoscopic scale

If we relax the $C^4$ closeness appearing in Theorem 5.5 into $C^2$ closeness, then we do not a priori expect regularity (because the MTW condition has no reason to stay true, even in just a neighborhood of the transport domain). However, as we shall see in this section, there is still some “mesoscopic regularity”. As usual, the geodesic distance $d$, the gradient $\nabla$, the volume $\text{vol}$, the exponential map $\text{exp}$ are implicitly defined for the metric $m$.

**Theorem 6.1.** Let $(M, m)$ be a uniformly regular Riemannian manifold with $\delta(M) > 0$. Then for any constants $A, a > 0$ there are positive constants $C = C(M, m, A, a)$ and $\sigma = \sigma(M, m, A, a) > 0$ with the following property: For any $\varepsilon > 0$ there is $\eta > 0$ such that if $m$ is another Riemannian metric on $M$, satisfying $\|m - m\|_{C^2} \leq \eta$, and $\mu, \nu$ are two probability measures on $M$ satisfying $\mu \leq A \text{vol}$ and $\nu \geq a \text{vol}$, then the
optimal transport map $T$ between $\mu$ and $\nu$, for the cost $c = d^2/2$, satisfies

\[
\begin{aligned}
   &\left\{\begin{array}{l}
   d(T(x), \text{cut}(x)) \geq \sigma \\
   d(T(x), T(y)) \leq C \left[ d(x, y) \vee \varepsilon \right]^{\alpha},
   \end{array}\right. \\
   \alpha = \frac{1}{4n - 1},
\end{aligned}
\]

whenever $(x, T(x))$ and $(y, T(y))$ belong to the support of the optimal transport plan.

**Remark 6.2.** Estimate (5.5) holds true for all $x, y$, even outside the support of $\mu$; but estimate (6.1) holds only on the projection of the support of the optimal plan, which is a set of full $\mu$-measure. Allowing $T$ to be multi-valued on a set of zero measure, we can always modify $T$ in such a way that (6.1) holds throughout the support of $\mu$ (for some choice of $T(x)$ and $T(y)$).

The short version of Theorem 6.1 is: choose your favorite “mesoscopic” scale $\varepsilon$, then the Hölder continuity estimate which holds true on $M$ (Theorem 5.2), also holds true on a $C^2$ perturbation of $M$, not necessarily at all scales, but at least at scales larger than $\varepsilon$. Such a result is meaningful even in a perspective of numerical analysis, where there is always a minimum resolution. Note carefully that $(M, m)$ is not required to satisfy any sort of Ma–Trudinger–Wang condition, so $T$ in (6.1) is not expected to be continuous; still the constant $C$ is independent of the scale $\varepsilon$ (otherwise the theorem would be completely trivial). Further note that the size of the admissible perturbation does depend on the probability densities $f$ and $g$, but in a very weak way, since only upper and lower pointwise bounds, respectively, are needed.

It turns out that Theorem 6.1 can be deduced from Theorem 5.3:

**Proof of Theorem 6.1.** By Theorem 5.2 there is $\overline{C} > 0$ such that (with obvious notation) on $M = (M, \overline{m})$, any optimal transport $T$ between $\mu$ and $\nu$, with $\mu \leq A\text{vol}$, $\nu \geq a\text{vol}$, satisfies

\[
   d(T(x), T(y)) \leq \overline{C} d(x, y)^{\alpha}.
\]

Let $C = \overline{C} + 1$.

Assume that the conclusion of Theorem 6.1 is false, then there are $\varepsilon > 0$ and (with obvious notation) sequences $m_k$ converging to $m$ in $C^2(M)$, $\mu_k \leq A\text{vol}_k$, $\nu_k \geq a\text{vol}_k$, $\psi_k$ such that $T_k = \exp_k(\nabla_k \psi_k)$ is the optimal transport between $\mu_k$ and $\nu_k$, and $x_k, y_k$ in $M$, $X_k \in \partial_c \psi_k(x_k), Y_k \in \partial_c \psi_k(y_k)$, such that

\[
   d(X_k, Y_k) \geq C \left[ d(x_k, y_k) \vee \varepsilon \right]^{\alpha}.
\]
Without loss of generality, \( \mu_k \rightarrow \mu, \nu_k \rightarrow \nu \), where \( \mu \leq A \text{vol}, \nu \geq a \text{vol} \); and \( \psi_k \rightarrow \psi, T_k \rightarrow T \). By Theorem 5.3, the convergence is uniform, so \( d(T(x), T(y)) \geq C [d(x, y) \vee \varepsilon]^\alpha \). Combining this with (6.2) and the definition of \( C \) we have
\[
(C + 1) [d(x, y) \vee \varepsilon]^\alpha \leq \overline{C} d(x, y)^\alpha.
\]
Whether \( x \neq y \) or \( x = y \), this is impossible. This contradiction proves the theorem.

The above proof looks rather neat, and the reader may be surprised to hear that I will now present a considerably more involved argument for the same result. The reason is the following: In the previous proof the constant \( \overline{C} \) is well controlled, but the amplitude \( \eta \) of the admissible perturbation is determined implicitly by a compactness argument involving not only the Riemannian measure \( m \), but also the enormous space of all probability measures on \( M \). On the contrary, the forthcoming argument will only involve compactness over a space of Riemannian metrics, then the rest of the treatment is constructive. If one is working with approximations living in a finite-dimensional space, the proof only involves a compactness argument over a finite-dimensional space, and becomes nearly constructive in some sense.

The key concept for this alternative proof is a notion which can be informally described as “uniformly regular manifold at scale \( \delta \)”, and is close in spirit to the concept of “positively curved length space at scale \( \delta \)” (Gromov’s CAT\(_\delta\)(K) spaces [8, 9]). It is possible that one may relax the assumption of \( C^2 \) approximation to allow for, say, noncollapsing Gromov–Hausdorff approximations with uniform upper and lower bounds on sectional curvatures, see [6] for information about these notions.

Before going on with the proof, we need some simple geometric results. As usual the objects \( d, \exp, \nabla, \text{vol} \) are implicitly associated with the Riemannian metric \( m \) appearing in the corresponding statement, and the \( C^r \) norms on the set of metrics are taken with respect to some fixed system of local charts.

**Lemma 6.3.** Let \((M, \overline{m})\) be a compact \( n \)-dimensional Riemannian manifold, and let \( m \) be another Riemannian metric with \( \|m - \overline{m}\|_{C^2} \leq \eta \) (\( \eta > 0 \)). Then

(i) There is a constant \( C > 0 \) with the following property: If \( \eta \) is small enough then for any \( \overline{x} \in M \), \( v_0, v_1 \in \text{TIL}(\overline{x}) \),
\[
d(\exp_{\overline{x}} v_0, \exp_{\overline{x}} v_1) \leq C |v_0 - v_1|.
\]

(ii) For any \( \sigma > 0 \) there is a constant \( C > 0 \) with the following property: If \( \eta \) is small enough then for any \( \overline{x} \in M \), \( v_0, v_1 \in \text{TIL}(\overline{x}) \), if \( y_i = \exp_{\overline{x}} v_i \) (\( i = 0, 1 \)),
\[
d(y_i, \text{cut}(\overline{x})) \geq \sigma \implies |v_0 - v_1| \leq C d(y_0, y_1).
\]
In particular, for any measurable set \( D \subset \text{TIL}(x) \) such that \( d(\exp_x D, \text{cut}(x)) \geq \sigma \), one has \( \text{vol}[D] \leq C^n \text{vol}[\exp_x(D)] \), where \( \text{vol} \) in the left-hand side stands for the Lebesgue measure in \( T_x M \).

(iii) For any \( \sigma > 0 \) there are \( \rho, C > 0 \) such that if \( \eta \) is small enough then for any \( x, y, \pi \in M \),

\[
d(\tilde{y}, \text{cut}(\tilde{x})) \geq \sigma, \quad d(\tilde{x}, \pi) \leq \rho \quad \implies \quad \frac{d(\tilde{x}, \tilde{y})^2}{2} \geq \frac{d(\pi, \tilde{y})^2}{2} - \langle \tilde{v}, \xi \rangle - C d(\pi, \tilde{x})^2,
\]

where \( \xi = (\exp_{\pi})^{-1}(\tilde{x}), \tilde{v} = (\exp_{\pi})^{-1}(\tilde{y}) \).

The previous lemma holds for any compact Riemannian manifold; on the contrary, the next one will need the conditions of uniform regularity of \( M \), and positivity of \( \delta(M) \). The parameters \( \tau \) appearing in this lemma will reflect the choice of a mesoscopic scale.

**Lemma 6.4.** Let \((M, \overline{m})\) be a uniformly regular Riemannian manifold with \( \overline{\delta} := \delta(M, \overline{m}) > 0 \), and let \( m \) be another Riemannian metric on \( M \), such that \( \|m - \overline{m}\|_{C^2} \leq \eta \) (\( \eta > 0 \)). Then

(i) For any \( \tau > 0 \), if \( \eta \) is small enough then for any \( \overline{\pi} \in M \) and any \( v_0 \in \text{TCL}(\overline{\pi}) \) there is \( v_1 \in \text{TIL}(\overline{\pi}) \) such that \( |v_0 - v_1| \geq \overline{\delta}/2 \) and \( d(\exp_{\overline{\pi}} v_0, \exp_{\overline{\pi}} v_1) \leq \tau \).

(ii) For any \( \delta > 0 \) there are \( \beta, \lambda > 0 \) with the following property: If \( \tau \) is given then for \( \eta \) small enough, for any \( \overline{\pi}, x \in M' \), any \( v_0, v_1 \in \text{TIL}(\overline{\pi}) \), any \( v = \mu((1-t)v_0 + tv_1) \) \((0 \leq t \leq 1, \mu \geq 0)\), \( y_0 = \exp_{\overline{\pi}} v_0, y_1 = \exp_{\overline{\pi}} v_1, y = \exp_{\overline{\pi}} v \),

\[
|v_0 - v_1| \geq \delta, \quad 1 \geq \mu \geq 1 - \beta d(\overline{\pi}, x) |v_0 - v_1|^2, \quad 1/4 \leq t \leq 3/4
\]

\[
(1-t)v_0 + tv_1 \in \text{TIL}(\overline{\pi})
\]

\[
d(\exp_{\overline{\pi}}(1-t)v_0 + tv_1), \text{cut}(\overline{\pi}) \geq \sigma
\]

\[
d(x, y)^2 - d(\overline{\pi}, y)^2 \geq \min(d(x, y_0)^2 - d(\overline{\pi}, y_0)^2, d(x, y_1)^2 - d(\overline{\pi}, y_1)^2)
\]

\[+ \lambda d(\overline{\pi}, x)^2 |v_0 - v_1|^2 - \tau.
\]

(iii) Same as (ii), but with \( \delta \) replaced by \( \sigma \) and the condition \( |v_0 - v_1| \geq \delta \) replaced by \( d(y_i, \text{cut}(\overline{\pi})) \geq \sigma \) \((i = 0, 1)\).
Lemmas 6.3 and 6.4 are established by basic tools of Riemannian geometry and standard contradiction arguments.

Let us consider for instance Lemma 6.3(ii). If the conclusion is false, then we can find \( v_{0,k}, v_{1,k} \) in \( \text{TIL}(\mathbf{x}_k) \) such that (with obvious notation) \( d(y_{i,k}, \text{cut}(\mathbf{x})) \geq \sigma \) and
\[
d_k(y_{0,k}, y_{1,k}) \leq C^{-1} |v_{0,k} - v_{1,k}|.
\]
Without loss of generality \( \mathbf{x}_k \rightarrow \mathbf{x}, y_{i,k} \rightarrow y_i, v_{i,k} \rightarrow v_i \). By Lemma 5.6, \( y_i \notin \text{cut}(\mathbf{x}) \), so \( v_i \in \text{TIL}(\mathbf{x}) \), then there is a smooth connected open set \( U \) such that \( \overline{U} \) is contained in \( \text{TIL}(\mathbf{x}) \) and \( U \) contains \( v_0, v_1 \), and the inverse of the exponential map is locally \( \overline{C} \)-Lipschitz in \( U \), so the inverse of \( \exp_k \) is locally \((\overline{C} + 1)\)-Lipschitz for \( k \) large enough (recall that the exponential map converges in \( C^1 \) if the metric converges in \( C^2 \)). Since \( U \) is smooth, \( \exp_x \) is globally \( L \)-Lipschitz for some suitable constant \( L \), contradicting (6.3) if \( C > L \).

Now we can proceed to the new proof of Theorem 6.1, in two steps: First show that the optimal transport stays a positive distance away from the cut locus, then prove the rough Hölder continuity. In the sequel I shall implicitly assume that \( \eta \) is so small that the conclusions of Lemmas 6.3 and 6.4 hold true whenever I need them.

**Alternative proof of Theorem 6.1. Step 1: Stay-away property.** The goal is to prove the existence of \( \sigma > 0 \) such that \( d(T(x), \text{cut}(x)) \geq \sigma \) for all \( x \) such that \( x, T(x) \) belongs to the support of the optimal transport plan \( \pi \) between \( \mu \) and \( \nu \). For this I shall adapt the second (constructive) proof of Theorem 6.1 in [15].

Let \( \mathbf{x} \in M \) and \( y_0 = T(\mathbf{x}) \); assume that \( (\mathbf{x}, y_0) \in \text{Spt} \pi \). Let further \( v_0 \in \text{TIL}(\mathbf{x}) \) such that \( \exp_{\mathbf{x}} v_0 = y_0 \), and let \( y_c = \exp_{\mathbf{x}} v_c \) be the cut point of \( \mathbf{x} \) along the geodesic \( (\exp_{\mathbf{x}} t v_0)_{t \geq 0} \). By Lemma 6.4(i) there is \( v_1 \in \text{TIL}(\mathbf{x}) \) such that \( d(y_c, y_1) \leq \tau \), where \( y_1 = \exp_{\mathbf{x}} v_1 \), and \( |v_c - v_1| \geq \delta \). Since the injectivity radius of \( M \) is bounded below, by an elementary geometric argument there is \( \delta_1 > 0 \) such that \(|v_0 - v_1| \geq \delta_1 \).

Let \( \beta, \lambda \) be defined by Lemma 6.4(ii) (with \( \delta = \delta_1 \)); for any \( r > 0 \) let
\[
D_r = \left\{ \mu ((1 - t)v_0 + tv_1) \in T\mathbf{x}M; \quad 1 - \beta d(\mathbf{x}, x)|v_0 - v_1|^2 \leq \mu \leq 1; \quad d(\mathbf{x}, x) \geq r; \quad 1/4 \leq t \leq 3/4 \right\}.
\]
As \( r \rightarrow 0 \) we have \( \text{vol}[D_r] \geq b r^{n-1} \) for some constant \( b > 0 \).

By Lemma 6.4(ii) we have \( d(\exp_{\mathbf{x}} (1 - t)v_0 + tv_1), \text{cut}(\mathbf{x})) \geq \sigma_0 \), for some \( \sigma_0 > 0 \); so \( D_r \) lies entirely in \( \{ y; d(y, \text{cut}(\mathbf{x})) \geq \sigma_0 \} \), and \( \text{vol}[D_r] \leq B \text{vol}[Y_r] \) for some suitable constant \( B \), where \( Y_r = \exp_{\mathbf{x}}(D_r) \). Combining this with our assumption on \( \nu \), we
conclude that
\[(6.4)\quad \nu[Y_r] \geq a \operatorname{vol}[Y_r] \geq a' r^{n-1},\]
for some $a' > 0$.

Next, with $B_r(y)$ standing for the open ball of radius $r$ and center $y$, our assumptions imply
\[(6.5)\quad \mu[B_r(x)] \leq A \operatorname{vol}[B_r(x)] \leq A' r^n,\]
where the second inequality follows for instance from the uniform bounds on the sectional curvature and on the injectivity radius. Choosing $r = a'/(4A')$, we obtain
\[(6.6)\quad \nu[Y_r] > \mu[B_r(x)].\]

Let now $\psi$ be a $c$-convex function such that $\psi$ solves the dual Kantorovich problem [21, Chapter 5] and $T(x) = \exp_x(\nabla \psi(x))$. Since $(\pi, y_0) \in \operatorname{Spt} \pi$, we have
\[(6.7)\quad \psi(x) + \frac{d(x, y_0)^2}{2} \leq \psi(x) + \frac{d(x, y_0)^2}{2} \quad \forall x \in M.\]

Since $d^2/2$ is uniformly Lipschitz, we also have
\[(6.8)\quad \frac{d(x, y_1)^2}{2} - \frac{d(x, y_0)^2}{2} \geq \frac{d(x, y_0)^2}{2} - \frac{d(x, y_0)^2}{2} - K d(y_0, y_1) \]
\[\geq \psi(x) - \psi(x) - K d(y_0, y_1).\]

On the other hand, by (6.6) the optimal transport has to send some mass from $M \setminus B_r(x)$ to $Y_r$. Let $x \in M \setminus B_r(x)$ be such that $(x, T(x)) \in \operatorname{Spt} \pi$ and $y = T(x) \in Y_r$, then
\[(6.9)\quad \psi(x) + \frac{d(x, y)^2}{2} \leq \psi(x) + \frac{d(x, y)^2}{2}.\]

Combining this with (6.7), (6.8) and Lemma 6.4(ii) again, we get
\[
\frac{d(x, y)^2}{2} - \frac{d(x, y)^2}{2} \geq \frac{d(x, y_0)^2}{2} - \frac{d(x, y_0)^2}{2} + \frac{d(x, y_0)^2}{2} + \lambda r^2 |v_0 - v_1|^2 - \tau \]
\[\geq \psi(x) - \psi(x) - K d(y_0, y_1) + \lambda r^2 |v_0 - v_1|^2 - \tau\]
\[\geq \frac{d(x, y_0)^2}{2} - \frac{d(x, y_1)^2}{2} - K d(y_0, y_1) + \lambda r^2 \delta^2_1 - \tau,\]
so $d(y_0, y_1) \geq K^{-1}(\lambda r^2 \delta^2_1 - \tau)$. The desired bound follows if $\tau$ is chosen so small that $\tau \leq \lambda r^2 \delta^2_1/2$. 
Step 2: Hölder continuity. The proof is an adaptation of [15, Section 7]. Let $\overline{x}, \tilde{x}$ in $M$ and let $\overline{y} = T(\overline{x}), \tilde{y} = T(\tilde{x})$, such that $(\overline{x}, \overline{y})$ and $(\tilde{x}, \tilde{y})$ both belong to $\text{Spt } \pi$. By optimal transport theory [21, Chapter 5], we have

$$
\left\{ \begin{array}{l}
\psi(\overline{x}) + \frac{d(\overline{x}, \overline{y})^2}{2} = \inf_{x \in M} \left( \psi(x) + \frac{d(x, \overline{y})^2}{2} \right) \\
\psi(\tilde{x}) + \frac{d(\tilde{x}, \tilde{y})^2}{2} = \inf_{x \in M} \left( \psi(x) + \frac{d(x, \tilde{y})^2}{2} \right).
\end{array} \right.
$$

(6.10)

Let $\overline{v}, \tilde{v} \in \overline{\text{TIL}}(x)$ be such that $\exp_\overline{x} \overline{v} = \overline{y}$, $\exp_\overline{x} \tilde{v} = \tilde{y}$. Define the sets $D_r$ and $Y_r$ as before, with $v_0 = \overline{v}$ and $v_1 = \tilde{v}$. The stay-away property makes it possible to apply Lemma 6.4(iii); then if $y \in Y_r$ and $x \notin B_r(\overline{x})$ we have

$$
\frac{d(x, y)^2}{2} - \frac{d(\overline{x}, y)^2}{2} \\
\geq \min \left( \frac{d(x, \overline{y})^2}{2} - \frac{d(\overline{x}, \overline{y})^2}{2}, \frac{d(x, \tilde{y})^2}{2} - \frac{d(\overline{x}, \tilde{y})^2}{2} \right) + \lambda r^2 |\overline{v} - \tilde{v}|^2 - \tau_1 \\
\geq \min \left( \psi(\overline{x}) - \psi(x), \psi(\tilde{x}) - \psi(x) + \frac{d(\tilde{x}, \tilde{y})^2}{2} - \frac{d(\overline{x}, \overline{y})^2}{2} \right) + \lambda r^2 |\overline{v} - \tilde{v}|^2 - \tau_1,
$$

and the latter quantity is strictly greater than $\psi(\overline{x}) - \psi(x)$ as soon as

$$
\psi(\tilde{x}) - \psi(\overline{x}) + \frac{d(\tilde{x}, \tilde{y})^2}{2} - \frac{d(\overline{x}, \overline{y})^2}{2} + \lambda r^2 |\overline{v} - \tilde{v}|^2 - \tau_1 > 0.
$$

(6.11)

When condition (6.11) is satisfied, $(x, y)$ cannot belong to the support of $\pi$, and since $x$ and $y$ were arbitrarily chosen in $(M \setminus B_r(\overline{x})) \times Y_r$, all the mass sent into $Y_r$ by the optimal transport has to come from $B_r(\overline{x})$, so necessarily $\text{vol } [Y_r] \leq (A/a) \text{vol } [B_r(\overline{x})]$. The set $D_r$ has diameter $O(|\overline{v} - \tilde{v}|)$ and its other dimensions are of the order of $r |\overline{v} - \tilde{v}|^2$; its image under $\exp_\overline{x}$ stays away from the cut locus by Lemma 6.4(iii); then by Lemma 6.3(ii) we have

$$
b r^{n-1} |\overline{v} - \tilde{v}|^{2(n-1)+1} \leq \text{vol } [D_r] \leq B \text{ vol } [Y_r],
$$

where $b$ and $B$ are positive constants. The conclusion is that

$$
|\overline{v} - \tilde{v}|^{2(n-1)+1} \leq B' r,
$$

(6.12)
as soon as (6.11) holds true.

Let us analyze condition (6.11). Let $h(x) = \psi(x) + d(x, \overline{y})^2/2$. The function $\psi$, being $d^2/2$-convex, is semiconvex with a uniform quadratic modulus (depending on sectional curvature bounds); moreover $\overline{v} - \tilde{v}$ is a subgradient of $\psi$ at $\overline{x}$ (see [21,
Chapter 10]). As for the function \( d(\cdot, \tilde{y})^2/2 \), it is of course not semiconvex on the whole of \( M \); but since \( d(\tilde{y}, \text{cut}(\tilde{x})) \geq \sigma \), Lemma 6.3(iii) applies if \( d(\tilde{x}, \tilde{x}) \) is smaller than some small constant \( \rho \), and then we deduce

\[
h(\tilde{x}) \geq h(\tilde{x}) + \langle \tilde{\nu} - \tilde{\nu}, \xi \rangle - K d(\tilde{x}, \tilde{x})^2,
\]

where \( K \) is a positive constant and \( \xi = (\exp_{\tilde{x}})^{-1}(\tilde{x}) \). In particular,

(6.13) \[
h(\tilde{x}) \geq h(\tilde{x}) - |\tilde{\nu} - \tilde{\nu}| d(\tilde{x}, \tilde{x}) - K d(\tilde{x}, \tilde{x})^2.
\]

If \( |\tilde{\nu} - \tilde{\nu}| \leq K d(\tilde{x}, \tilde{x}) \) then the conclusion is easy; so let us focus on the other case when \( |\tilde{\nu} - \tilde{\nu}| > K d(\tilde{x}, \tilde{x}) \). Then (6.13) implies

\[
h(\tilde{x}) > h(\tilde{x}) - 2|\tilde{\nu} - \tilde{\nu}| d(\tilde{x}, \tilde{x}),
\]

and (6.11) holds true as soon as

\[
\tau_1 + 2 |\tilde{\nu} - \tilde{\nu}| d(\tilde{x}, \tilde{x}) \leq \lambda r^2 |\tilde{\nu} - \tilde{\nu}|^2,
\]

or equivalently

(6.14) \[
r \geq \lambda^{-1/2} \left( \frac{\tau_1}{|\tilde{\nu} - \tilde{\nu}|^2} + \frac{2 d(\tilde{x}, \tilde{x})}{|\tilde{\nu} - \tilde{\nu}|} \right)^{1/2}.
\]

The previous estimates also require that \( r \) be smaller than some fixed quantity \( r_0 \).

Finally, Lemma 6.3(i)-(ii) show that

(6.15) \[
K^{-1} |\tilde{\nu} - \tilde{\nu}| \leq d(\tilde{y}, \tilde{y}) \leq K |\tilde{\nu} - \tilde{\nu}|
\]

for some suitable constant \( K \).

Thus (6.12) and (6.14) lead to the implication

(6.16) \[
r_0 \geq r \geq \Lambda \left( \frac{\tau_1}{d(\tilde{y}, \tilde{y})^2} + \frac{d(\tilde{x}, \tilde{x})}{d(\tilde{y}, \tilde{y})} \right)^{1/2} \implies d(\tilde{y}, \tilde{y})^{2n-1} \leq M r,
\]

where \( r_0, \Lambda \) and \( M \) are positive constant. The game is to show that if we choose \( \tau_1 \) small enough (as a function of \( \epsilon \)), then (6.16) implies the desired conclusion \( d(\tilde{y}, \tilde{y}) \leq C [d(\tilde{x}, \tilde{x}) \vee \epsilon]^{\alpha} \), \( \alpha = (4n - 1)^{-1} \).

- If \( \frac{r_0^2}{2 \Lambda^2} \geq \frac{d(\tilde{x}, \tilde{x})}{d(\tilde{y}, \tilde{y})} \geq \frac{\tau_1}{d(\tilde{y}, \tilde{y})^2} \) we choose \( r = \Lambda \sqrt{2 d(\tilde{x}, \tilde{x}) d(\tilde{y}, \tilde{y})} \); then from (6.16) we get

\[
d(\tilde{y}, \tilde{y})^{2n-1} \leq M r \leq \sqrt{2} M \Lambda d(\tilde{x}, \tilde{x})^{1/2} / d(\tilde{y}, \tilde{y})^{1/2},
\]

which does imply \( d(\tilde{y}, \tilde{y}) \leq C_1 d(\tilde{x}, \tilde{x})^{\alpha} \).
Next if \( \frac{d(x, \tilde{x})}{d(y, \tilde{y})} \geq \frac{r_0^2}{2\Lambda^2} \), then \( d(y, \tilde{y}) \leq (2\Lambda^2/r_0^2) d(x, \tilde{x}) \leq C_3 d(x, \tilde{x})^\alpha \) for \( C_3 = (2\Lambda^2/r_0^2) (\text{diam } M)^{1-\alpha} \).

Finally, if \( \frac{\tau_1}{d(y, \tilde{y})^2} \geq \frac{d(x, \tilde{x})}{d(y, \tilde{y})} \) then we choose \( r = \varepsilon^\beta \) for \( \beta = \alpha(2n - 1) \); this is admissible if \( \varepsilon \) is small enough and \( \tau_1 \leq \varepsilon^\beta (\text{diam } M)^2 \). Then (6.16) implies

\[
d(y, \tilde{y})^{2n-1} \leq M' \varepsilon^\beta,
\]

so

\[
d(y, \tilde{y}) \leq M'' \varepsilon^{\frac{\beta}{2n-1}} = M'' \varepsilon^\alpha.
\]

This concludes the proof of Theorem 6.1.  

\[\square\]

**Remark 6.5.** We have seen in the beginning of this section that the perturbative regularity result (Theorem 6.1) follows from the pointwise stability result (Theorem 5.5). But conversely, the perturbative regularity result implies the pointwise stability result by a variant of Ascoli’s theorem. This time, the driving mechanism for the pointwise convergence is the “mesoscopic Hölder regularity”, whereas before it was the smoothness of the limit and the additional compactness based on convexity.

**References**


**Cédric Villani**

ENS Lyon & Institut Universitaire de France  
UMPA, UMR CNRS 5669  
46 allée d’Italie  
69364 Lyon Cedex 07  
FRANCE

E-MAIL: cvillani@umpa.ens-lyon.fr