

April 26, 2008

AN APPROXIMATION LEMMA ABOUT THE CUT LOCUS,
WITH APPLICATIONS IN OPTIMAL TRANSPORT THEORY

A. FIGALLI AND C. VILLANI

ABSTRACT. A path in a Riemannian manifold can be approximated by a path meeting only finitely many times the cut locus of a given point. The proof of this property uses recent works of Itoh–Tanaka and Li–Nirenberg about the differential structure of the cut locus. We present applications in the regularity theory of optimal transport.

1. MOTIVATIONS

Various authors have investigated lately the regularity of optimal transport for non-Euclidean cost functions [5, 6, 8, 10, 11, 12, 13, 14, 15]. In most of these papers, the cost function is defined on $\Omega \times \Lambda$, where Ω, Λ are bounded open subsets of \mathbb{R}^n ; but some of these papers deal with genuinely curved geometries [8, 11, 12]. To do this, two main strategies have been proposed:

(a) localize the transport using charts, and then apply the a priori estimates in \mathbb{R}^n . This is how Loeper [11] proves Hölder regularity for optimal transport on the sphere; his proof strongly uses the fact that in this geometry there is an independent argument to prove that the optimal transport map stays a positive distance away from the cut locus.

(b) establish regularity estimates directly on the solution. This was done by Loeper and the second author in [12].

Strategy (b) has the interest to provide more information “in the large”. Strategy (a) seems to be easier, but the reasoning for the localization is quite nontrivial if one does not know a priori the continuity of optimal transport [11]. (Once continuity has been established, localizing becomes very easy, and then one can also apply the higher-order Hölder regularity estimates from [13].) Anyway there is motivation for a direct non-Euclidean approach to the regularity theory.

So what’s the big deal? The reader’s first guess might be that one should try to repeat the Euclidean proofs with minor changes. The main obstacle arising in the proof is the *cut locus* issue: a typical cost function arising in a geometric context, involving the geodesic distance, will always present singularities. For instance, the

argument in [8, 10, 12] is based on a maximum-type principle for the function

$$(1) \quad [0, 1] \ni t \mapsto c(x, y_t) - c(\bar{x}, y_t),$$

where $t \mapsto \nabla_x c(\bar{x}, y_t)$ is (a perturbation of) a straight segment as t varies in $[0, 1]$. (In the language of the above-mentioned works, (y_t) is a c -segment.) Even if we know a priori that y_t remains within the domain of smoothness of $c(\bar{x}, \cdot)$, there will be values of x such that $c(x, y_t)$ is not a smooth function of t .

Kim and McCann [8] solve this problem in the particular case of the squared geodesic distance on the sphere, noting that if a c -segment (y_t) with base \bar{x} goes through $\text{cut}(x) = \{-x\}$, one can still perturb just a bit the point x into \hat{x} , in such a way as to avoid the path (y_t) . Then all inequalities which are established for \hat{x} in place of x , also hold for x by passing to the limit.

But the argument used by Kim and McCann is not robust: In general the cut locus of a point in an n -dimensional Riemannian manifold is much larger than a point, in fact it may have dimension as large as $n - 1$. So we expect that for generic manifolds (even for an ellipsoid) the event of a given one-dimensional curve meeting the cut locus is stable under perturbation.

In [12] a different density argument was proposed for the squared geodesic distance on a Riemannian manifold whose cut locus is *nonfocal*, i.e. such that for any (x, ξ) in the unit tangent bundle, $d_{t\xi} \exp_x$ is invertible at $t = t_C(x, \xi)$, the cut time of the geodesic starting at x with unit velocity ξ . Instead of avoiding the cut locus, one seeks to cross it in the most controllable way, namely on a discrete set of times. In [12] it is shown that, given a path (y_t) and a point x in a Riemannian with a nonfocal cut locus, it is possible to perturb (y_t) into another path (\hat{y}_t) , crossing the cut locus of x *transversally*.

The construction of [12] is strongly based on the nonfocality assumption. In the present paper we shall prove a similar approximation lemma by a completely different and more “probabilistic” argument, which will apply in a more general context. Our main assumption is that the cut locus has locally finite $(n - 1)$ -dimensional Hausdorff measure. That this assumption is satisfied by the squared distance on a Riemannian manifold is guaranteed by recent and rather subtle results about the structure of the cut locus, proven independently by Itoh and Tanaka [7] and Li and Nirenberg [9] (see [9, Theorem 1.3]).

The plan of this paper is the following: In Section 2 we present our main result, called “transversal approximation” lemma. In Section 3 we show how to use this approximation lemma in practical situations, and we shall quote some applications.

Some of these results were included, in a less complete form, in [16, Chapter 12].

This paper is dedicated to Neil Trudinger, as a tribute to his lasting and constantly renewed influence on the theory of fully nonlinear elliptic equations.

2. TRANSVERSAL APPROXIMATION LEMMA

For simplicity we shall restrict to the case when the cost function c is the squared geodesic distance on a Riemannian manifold. By convention, all Riemannian manifolds will be implicitly assumed to be smooth, complete and connected.

Lemma 1. *Let M be a Riemannian manifold. For any x in M , any C^2 path $(y_t)_{0 \leq t \leq 1}$ drawn in M can be approximated in C^2 topology by a path $(\hat{y}_t)_{0 \leq t \leq 1}$ such that $\{t \in (0, 1); \hat{y}_t \in \text{cut}(x)\}$ is discrete.*

Proof of Lemma 1. According to independent results by Itoh and Tanaka [7] and Li and Nirenberg [9],

$$(2) \quad \mathcal{H}^{n-1}[K \cap \text{cut}(x)] < +\infty,$$

where $K \subset M$ is any compact set, and \mathcal{H}^{n-1} is the $(n-1)$ -dimensional Hausdorff measure; see for instance [9, Corollary 1.3].

Since $\text{cut}(x)$ has empty interior, for any fixed $t_0 \in [0, 1]$ we can perturb the path (y_t) in C^2 topology into a path (\tilde{y}_t) , such that $\tilde{y}_{t_0} \notin \text{cut}(x)$. Repeating this operation finitely many times, we can ensure that \tilde{y}_{t_j} lies outside $\text{cut}(x)$ for each $t_j = j/2^k$, where $k \in \mathbb{N}$ and $j \in \{0, \dots, 2^k\}$. If k is large enough, then for each j the path $(\tilde{y}_t)_{t_j \leq t \leq t_{j+1}}$ can be written as a straight line segment, in some well-chosen local chart. Moreover, since $\text{cut}(x)$ is closed, for any j there will be $\varepsilon_j > 0$ and $r_j > 0$ such that on the interval $[t_j - \varepsilon_j, t_j + \varepsilon_j]$ the path \tilde{y}_t is entirely contained in the small ball $B_j = B(\tilde{y}_{t_j}, r_j)$, and the larger ball $2B_j = B(\tilde{y}_{t_j}, 2r_j)$ does not meet $\text{cut}(x)$.

If we prove that the path (\tilde{y}_t) can be approximated on each interval $[t_{j-1} + \varepsilon_{j-1}, t_j - \varepsilon_j]$ by a path (\hat{y}_t) meeting $\text{cut}(x)$ at most finitely many times, then we can “patch together” these pieces by smooth paths defined on the intervals $[t_j - \varepsilon_j, t_j + \varepsilon_j]$ and staying within $2B_j$. Obviously the resulting approximation will meet $\text{cut}(x)$ at most finitely many times.

All this shows that we just have to treat the case when (y_t) takes values in a small open subset U of \mathbb{R}^n and is a straight line. In these coordinates, $\Sigma := \text{cut}(x) \cap U$ has finite \mathcal{H}^{n-1} measure. Without loss of generality, we can assume that U is the cylinder $B(0, \sigma) \times (-\tau, \tau)$ for some $\sigma, \tau > 0$, and $y_t = t e_n$ for $t \in (-\tau, \tau)$ (where (e_1, \dots, e_n) is an orthonormal basis of \mathbb{R}^n).

For any $z \in B(0, \sigma) \subset \mathbb{R}^{n-1}$, let $y_t^z = (z, t)$. The goal is to show that $\mathcal{H}^{n-1}(dz)$ -almost surely, y_t^z intersects Σ in at most finitely many points. To do this one can

apply the co-area formula in the following form (see [3, p. 109] and [4, Sections 2.10.25 and 2.10.26]): let $f : (z, t) \mapsto z$ (defined on U), then

$$\mathcal{H}^{n-1}[\Sigma] \geq \int_{f(\Sigma)} \mathcal{H}^0[\Sigma \cap f^{-1}(z)] \mathcal{H}^{n-1}(dz).$$

By assumption the left-hand side is finite, and the right-hand side is exactly $\int \#\{t; y_t^z \in \Sigma\} \mathcal{H}^{n-1}(dz)$; so the integrand is finite for almost all z , and in particular there is a sequence $z_k \rightarrow 0$ such that each $(y_t^{z_k})$ intersects Σ finitely many often. \square

Remark 2. The conclusion of Lemma 1 also applies for much more general cost functions, with a suitable definition of cut locus (as the complement of the domain of smoothness of the cost function), under the assumption that the cut locus of any point has locally finite $(n-1)$ -dimensional Hausdorff measure. See Assumption **(Cut $^{n-1}$)** in [16, Chapter 12] for a precise statement.

3. APPLICATION TO REGULARITY THEORY

Let μ, ν be two probability measures on a compact Riemannian manifold M equipped with its geodesic distance d ; assume that μ gives zero mass to countably $(n-1)$ -rectifiable sets. By [16, Theorem 10.28 and Remark 10.33], there is a map T solving the Monge–Kantorovich problem with initial measure μ , final measure ν , and cost function $c = d^2/2$; it is uniquely determined up to modification on a μ -negligible set. Moreover, $\mu(dx)$ -almost surely, $T(x) \notin \text{cut}(x)$. The problem is to study the smoothness of T .

Works by Loeper [10], Ma, Trudinger and Wang [13] and others suggest that the three basic conditions to develop a regularity theory for T are

(a) an estimate that the transport map T does not approach the cut locus:

$$(3) \quad \exists \sigma > 0; \quad \forall x \in M, \quad d(T(x), \text{cut}(x)) \geq \sigma;$$

(b) a c -convexity condition on the domain of smoothness of the cost function:

$$(4) \quad \forall x \in M, \quad \nabla_x c(x, M \setminus \text{cut}(x)) \text{ is convex}$$

(or strictly convex in some appropriate sense);

(c) a suitable version of the Ma–Trudinger–Wang condition, for instance the MTW(K_0, C_0) condition used in [12]:

$$(5) \quad -(3/2) \frac{\partial^2}{\partial p_\eta^2} \frac{\partial^2}{\partial x_\xi^2} c(x, y) \geq K_0 |\xi|_x^2 |\tilde{\eta}|_x^2 - C_0 \langle \xi, \tilde{\eta} \rangle_x |\xi|_x |\tilde{\eta}|_x,$$

where $\tilde{\eta} := -\nabla_{xy}^2 c(x, y) \cdot \eta = (d_v \exp_x)^{-1} \eta$, and it is assumed at least that $K_0 \geq 0$. The meaning of the left-hand side in (5) is the following: first freeze y and differentiate $c(x, y)$ twice with respect to x in the direction $\xi \in T_x M$. Then, considering the result as a function of y , parameterize y by $p = -\nabla_x c(x, y)$, and differentiate twice with respect to p in the direction $\eta \in T_y M$. When we say that (5) is satisfied, we really mean that it is satisfied as soon as $y \notin \text{cut}(x)$ (otherwise the partial derivatives may be meaningless).

Assumption (c) can be considered as given a priori. It is shown in [12] that it sometimes automatically implies assumption (b). But the least satisfactory assumption is certainly (a) (“stay-away property”), which one really would like to deduce from (b) and (c), together with some suitable geometric conditions.

This seems to be a difficult problem. Some particular cases (spherical or nearly spherical geometry in some sense) are well-understood [1, 2], but there are very few general results establishing this stay-away property. One such result is established in [12, Theorem 6.1]: it is sufficient that M has no “focalizing uniquely minimizing geodesic” (i.e. no geodesic which is the unique minimizing curve between its endpoints, but whose endpoints are conjugate), and that M is “uniformly regular”, in the sense of the following definition. We shall denote by d the geodesic distance on a Riemannian manifold M ; by $\text{TIL}(x)$ the tangent injectivity locus, which is the largest open domain of injectivity of the exponential map \exp_x ; and by $\text{TCL}(x)$ the tangent cut locus, which is the boundary of $\text{TIL}(x)$. In case of need, see [12] for more information.

Definition 3 (Uniform regularity). With the above notation, the cost $c = d^2/2$ is said to be *uniformly regular* if there are $\varepsilon_0, \kappa, \lambda > 0$ and $C > 0$ such that

(a) for all $x \in M$, $\text{TIL}(x)$ is κ -uniformly convex;

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

$$(6) \quad \forall t \in (0, 1), \quad |\ddot{p}_t| \leq \varepsilon_0 d(\bar{x}, x) |\dot{p}_t|^2, \quad |\dot{p}_t| \leq C,$$

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

$$(7) \quad d(x, y_t)^2 - d(\bar{x}, y_t)^2 \geq \min \left(d(x, y_0)^2 - d(\bar{x}, y_0)^2, d(x, y_1)^2 - d(\bar{x}, y_1)^2 \right) \\ + 2 \lambda t(1-t) d(\bar{x}, x)^2 |p_1 - p_0|^2.$$

Uniform regularity is an “integral” manifestation of the Ma–Trudinger–Wang condition. If everything is smooth, one can deduce it from the $\text{MTW}(K_0, C_0)$ condition with an adaptation of the Kim–McCann method; see [12]. But the left-hand side

of (7) is not a smooth function of t because of cut locus issues. We shall now show how to use the approximation result from Lemma 1 to bypass this regularity problem.

Theorem 4. *Let M be a compact Riemannian manifold, and let $c = d^2/2$. Assume that there are $\kappa > 0$, $K_0 > 0$, $C_0 < \infty$ such that*

- (i) *TIL(x) is κ -uniformly convex, for all x in M ;*
- (ii) *c satisfies MTW(K_0, C_0).*

Then c is uniformly regular.

The proof of this theorem is a variant of the proof of Theorem 3.1 in [12]; the main difference is that it relies on Lemma 1 rather than [12, Lemma 4.4].

Proof of Theorem 4. Let \bar{x} , x , $(p_t)_{0 \leq t \leq 1}$ and $(y_t)_{0 \leq t \leq 1}$ be as in Definition 3. By density and Lemma 1, we may assume that $y_0, y_1 \notin \text{cut}(x)$ and that y_t meets $\text{cut}(x)$ only at finitely many times t_1, \dots, t_{N-1} , all the other conditions in Definition 3 being unchanged (apart from a slight increase in ε_0 , but we can slightly reduce it from the beginning). We may also assume $x \neq \bar{x}$ and $p_0 \neq p_1$, otherwise everything is trivial. We let $t_0 = 0$, $t_N = 1$, and we define $h(t) = -c(x, y_t) + c(\bar{x}, y_t) + \delta t(1 - t)$, where $\delta > 0$ will be chosen later.

On each time-interval (t_j, t_{j+1}) we have $y_t \notin \text{cut}(x)$, so $q_t = -\nabla_y c(x, y_t) = (\exp_{y_t})^{-1}(x)$ is well-defined, h is a smooth function of t , and by convexity of TIL(y_t) we have $[\bar{q}_t, q_t] \subset \text{TIL}(y_t)$, where $\bar{q}_t = -\nabla_y c(\bar{x}, y_t) = (\exp_{y_t})^{-1}(\bar{x})$. As in the proof of [12, Theorem 3.1], one checks that if $\varepsilon_0 \leq K_0/8$ and $\delta = \lambda d(\bar{x}, x)^2 |p_1 - p_0|^2$, with

$$\lambda := \frac{K_0 (1 - 2\sigma \eta_0 \text{diam}(M))_+^2}{12 C_0 (1 + \sigma \text{diam}(M))}, \quad \sigma := \max |\dot{p}|, \quad \eta_0 \geq \sup_{0 \leq t \leq 1} \left(\frac{|\ddot{p}_t|}{|q_t - \bar{q}_t| |\dot{p}_t|^2} \right),$$

then the identity $\dot{h}(t) = 0$ gives $\ddot{h}(t) > 0$, for any $t \in (t_j, t_{j+1})$. This implies that the function h cannot have any maximum on (t_j, t_{j+1}) .

Since h is continuous on $[0, 1]$, it achieves its maximum at one of the times t_j ($0 \leq j \leq N$). The goal is to show that necessarily $j = 0$ or $j = N$. Indeed, let $j \in \{1, \dots, N-1\}$. If \dot{h} is continuous at t_j and $\dot{h}(t_j) \neq 0$, clearly t_j cannot be a maximum of h . The same is true if \dot{h} is discontinuous at t_j , because by semiconvexity of $t \rightarrow -c(x, y_t)$, necessarily $\dot{h}(t_j^+) > \dot{h}(t_j^-)$. Finally, if \dot{h} is continuous at t_j and $\dot{h}(t_j) = 0$, the same computations as before show that $\ddot{h}(t)$ is strictly positive when t is close to (but different from) t_j , then the continuity of \dot{h} implies that h is strictly convex around t_j , so it cannot have a maximum at t_j .

The only possibility left out for h is to achieve its maximum at $t_0 = 0$ or $t_N = 1$, and we obtain (7). \square

Combining the theorem above with [12, Theorem 5.1], we obtain the continuity of the optimal transport map under some weak assumptions on the measures:

Corollary 5. *Let M be a compact Riemannian manifold, and let $c = d^2/2$. Assume that there are $\kappa > 0$, $K_0 > 0$, $C_0 < \infty$, $c_0 > 0$ such that*

(i) *TIL(x) is κ -uniformly convex, for all x in M ;*

(ii) *c satisfies $\text{MTW}(K_0, C_0)$;*

(iii) $\lim_{\varepsilon \rightarrow 0} \left(\sup_{x \in M} \frac{\mu[B_\varepsilon(x)]}{\varepsilon^{n-1}} \right) = 0$;

(iv) *there exists $c_0 > 0$ such that $\nu[A] \geq c_0 \text{vol}[A]$ for all $A \subset M$ Borel.*

Then the optimal transport map between μ and ν , for the cost $d^2/2$, takes the form $T(x) = \exp_x \nabla \psi(x)$, where ψ is a C^1 semiconvex function. In particular T is continuous.

Remark 6. The reasoning used in the proof of Theorem 4 was also used to prove Theorem 12.42 in [16] (dealing with general cost functions) and Proposition 2.5 in [17]. These two results, closely related to Theorem 4, establish the equivalence between the Ma–Trudinger–Wang condition and an “integral” inequality in the style of the one appearing in Definition 3. The same strategy is used to prove Theorem 5.5 in [17], a perturbative smoothness result for optimal transport on an approximation of a uniformly regular manifold.

REFERENCES

- [1] DELANOË, P., AND LOEPER, G. Gradient estimates for potentials of invertible gradient-mappings on the sphere. *Calc. Var. Partial Differential Equations* 26, 3 (2006), 297–311.
- [2] DELANOË, P., AND GE, Y. Regularity of optimal transportation maps on compact, locally nearly spherical, manifolds. In preparation.
- [3] EVANS, L. C., AND GARIEPY, R. *Measure theory and fine properties of functions*. CRC Press, Boca Raton, FL, 1992.
- [4] FEDERER, H. *Geometric measure theory*. Die Grundlehren der mathematischen Wissenschaften, Band 153. Springer-Verlag New York Inc., New York, 1969.
- [5] FIGALLI, A., KIM, Y.-H., AND MCCANN, R. J. Work in progress.
- [6] FIGALLI, A., AND LOEPER, G. C^1 regularity of solutions of the Monge–Ampère equation for optimal transport in dimension two. Preprint, 2008.
- [7] ITOH, J.-I., AND TANAKA, M. The Lipschitz continuity of the distance function to the cut locus. *Trans. Amer. Math. Soc.* 353, 1 (2001), 21–40.

- [8] KIM, Y.-H., AND MCCANN, R. J. Continuity, curvature, and the general covariance of optimal transportation. Preprint, 2007. Archived online at arxiv.org/abs/0706.1266.
- [9] LI, Y., AND NIRENBERG, L. The distance function to the boundary, Finsler geometry, and the singular set of viscosity solutions of some Hamilton–Jacobi equations. *Comm. Pure Appl. Math.* 58 (2004), 85–146.
- [10] LOEPER, G. On the regularity of maps solutions of optimal transportation problems. To appear in *Acta Math.*
- [11] LOEPER, G. On the regularity of maps solutions of optimal transportation problems II. Preprint, 2007.
- [12] LOEPER, G., AND VILLANI, C. Regularity of optimal transport in curved geometry: the nonfocal case. Preprint, 2007. Available at www.umpa.ens-lyon.fr/~cvillani.
- [13] MA, X.-N., TRUDINGER, N. S., AND WANG, X.-J. Regularity of potential functions of the optimal transportation problem. *Arch. Ration. Mech. Anal.* 177, 2 (2005), 151–183.
- [14] TRUDINGER, N. S., AND WANG, X.-J. On strict convexity and continuous differentiability of potential functions in optimal transportation. Preprint, 2006.
- [15] TRUDINGER, N. S., AND WANG, X.-J. On the second boundary value problem for Monge–Ampère type equations and optimal transportation. Preprint, 2006. Archived online at arxiv.org/abs/math.AP/0601086.
- [16] VILLANI, C. Optimal transport, old and new. Notes for the 2005 Saint-Flour summer school. To appear in the Springer series *Grundlehren der mathematischen Wissenschaften*. Preliminary version available at www.umpa.ens-lyon.fr/~cvillani.
- [17] VILLANI, C. Stability of a 4th-order curvature condition arising in optimal transport theory. Preprint, 2008.

ALESSIO FIGALLI

UNIVERSITÉ DE NICE SOPHIA-ANTIPOLIS
 LABORATOIRE J.-A. DIEUDONNÉ
 CNRS UMR 6621
 PARC VALROSE
 06108 NICE CEDEX 02
 FRANCE

E-MAIL: figalli@unice.fr

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