

# NEARLY ROUND SPHERES LOOK CONVEX

A. FIGALLI, L. RIFFORD, AND C. VILLANI

ABSTRACT. We prove that a Riemannian manifold  $(M, g)$ , close enough to the round sphere in the  $C^4$  topology, has uniformly convex injectivity domains — so  $M$  appears uniformly convex in any exponential chart. The proof is based on the Ma–Trudinger–Wang nonlocal curvature tensor, which originates from the regularity theory of optimal transport.

## CONTENTS

1. Introduction	1
2. Discussion	3
3. Variations of Jacobi fields	10
4. Convexity of nonfocal domains	12
5. Stability of MTW condition	14
6. Convexity of injectivity domains	24
References	32

## 1. INTRODUCTION

The main result of this paper states, in short, that the round sphere is “robustly intrinsically convex”, or “robustly log convex”, in a sense which we shall now explain.

Let  $(M, g)$  be a  $C^\infty$  compact Riemannian manifold of dimension  $n \geq 2$ , and  $\exp$  the associated Riemannian exponential. For any  $x \in M$ ,  $v \in T_x M \setminus \{0\}$ , we define

$$\begin{aligned} t_C(x, v) &= \textit{cut time of } (x, v) \\ &= \max \left\{ t \geq 0; (\exp_x(sv))_{0 \leq s \leq t} \text{ is a minimizing geodesic} \right\}. \end{aligned}$$

Then for any  $x \in M$  we let

$$\begin{aligned} \text{TCL}(x) &= \textit{tangent cut locus of } x \\ &= \{t_C(x, v)v; v \in T_x M \setminus \{0\}\}; \\ \text{I}(x) &= \textit{injectivity domain of the exponential map at } x \\ &= \left\{tv; 0 \leq t < t_C(x, v), v \in T_x M \setminus \{0\}\right\}. \end{aligned}$$

We write  $\exp^{-1}$  or  $\log$  for the inverse of the exponential map: by convention  $\log_x(y) = \exp_x^{-1}(y)$  is the set of *minimizing* velocities  $v$  such that  $\exp_x v = y$ . In particular  $\text{TCL}(x) = \log_x(\text{cut}(x))$ , and  $\text{I}(x) = \log_x(M \setminus \text{cut}(x))$ . So the injectivity domain at  $x$  is the parameterization of  $M$  (after “cutting out the cut locus”) in the maximal exponential chart centered at  $x$ .

Ever since its introduction by Poincaré [31], the cut locus has retained some mystery; see Berger [2, Subsection 6.5.4] for a review. Around 1960 it played a key role in Klingenberg’s proof of the topological sphere theorem, as exposed e.g. in [10, Chapter 13]. Since then not so much has been found, except for explicit computations in particular geometries,<sup>1</sup> and local properties, such as the Lipschitz continuity and  $(n - 1)$ -dimensional rectifiability [6, 21, 25, 30]. The second-order behavior is still open: for instance it is not known whether  $\text{I}(x)$  is a semiconvex set [28, Appendix B], or if  $\text{TCL}(x)$  is an Alexandrov space [21, Problem 3.4]. A genuinely nonsmooth object, the cut locus is in general not triangulable [32] and in high enough dimension does not depend smoothly on the metric, even for generic manifolds [3, 4].

In the present paper we prove a *global*, perturbative geometric result of a new nature on the injectivity domain, or equivalently on the tangent cut locus:

**Theorem 1.1.** *Let  $(M, g)$  be a  $C^4$  perturbation of the round sphere  $\mathbb{S}^n$ . Then all injectivity domains of  $M$  are uniformly convex.*

An informal way to paraphrase the conclusion is to say that  $M$  appears convex from any of its points. Here are some first comments.

**Remarks 1.2.** (1) In Theorem 1.1, “ $C^4$  perturbation” means that  $M$  is  $\mathbb{S}^n$  equipped with a metric  $g$ , such that  $\|g - g^0\|_{C^4} \leq \delta$ , where  $g^0$  is the round metric,  $\delta = \delta(n)$  is small enough, and the  $C^4$  norm is computed in a choice of local charts. This implies that the exponential map is a  $C^3$  perturbation

---

<sup>1</sup>Shockingly, the cut locus of the ellipsoid was rigorously described only a few years ago [20].

of the “round” exponential (recall that the geodesic equations involve the Christoffel symbols, which are derivatives of the metric).

- (2) One may think that the result still holds under the more intrinsic assumption that the Riemannian curvature of  $g$  is  $C^2$ -close to the Riemannian curvature of  $g^0$ , in the sense used by Delanoë & Ge [8, 9]. We shall not however pursue in this direction.
- (3) Theorem 1.1 was established for  $n = 2$  in [14], except for the fact that only strict convexity was proven.

In the next section we shall provide much more comments, discuss the main difficulties and ingredients behind Theorem 1.1, and set up some preliminaries for the proof.

## 2. DISCUSSION

**2.1. Focalization.** The major difficulty in the proof of Theorem 1.1, as in the vast majority of studies on the cut locus, is the *focalization* phenomenon. Let us introduce some more notation:

$$\begin{aligned}
 t_F(x, v) &= \text{focalization time of } (x, v) \\
 &= \inf \left\{ t \geq 0; \det(d_{tv} \exp_x) = 0 \right\}; \\
 \text{TFL}(x) &= \text{tangent focal locus of } x \\
 &= \{t_F(x, v)v; v \in T_x M \setminus \{0\}\}; \\
 \text{NF}(x) &= \text{nonfocal domain at } x \\
 &= \{tv; 0 \leq t < t_F(x, v), v \in T_x M \setminus \{0\}\}.
 \end{aligned}$$

It is a classical result that  $I(x) \subset \text{NF}(x)$ , see e.g. [18, Corollary 3.77] or [33, Problem 8.8]. In negative sectional curvature, there is no focalization and TFL is empty; conversely, in positive curvature there is focalization in all directions, and the tangent cut locus is “surrounded” by the tangent focal locus (remember that  $M$  is assumed to be compact).

The tangent focal locus is much better understood than the tangent cut locus. For instance, for any complete Riemannian manifold  $(M, g)$  the tangent focal locus is included in a countable union of smooth hypersurfaces; and nonfocal domains are semiconvex [6]. However, the focal locus is also the source for most of the major

difficulties in the analysis of the cut locus. In fact the “bad set” is the **tangent focal cut locus**, defined by

$$(2.1) \quad \text{TFCL}(x) = \text{TCL}(x) \cap \text{TFL}(x).$$

To illustrate this, let us consider  $M = \mathbb{R}\mathbb{P}^n = \mathbb{S}^n / \{\pm \text{Id}\}$ ; then  $\text{TFCL} = \emptyset$ . This nonfocality property makes it possible to locally describe  $\text{TCL}(x)$ , for a perturbation of  $\mathbb{R}\mathbb{P}^n$ , by the Implicit Function Theorem [28, Appendix C]. Then it is easy to prove the convexity of injectivity domains as soon as  $(M, g)$  is just a  $C^3$  perturbation of  $\mathbb{R}\mathbb{P}^n$ . In other words, if in Theorem 1.1 we replace  $\mathbb{S}^n$  by  $\mathbb{R}\mathbb{P}^n$  the result becomes just an exercise, and the conclusion can be improved.

More generally, the nonfocality property is also true for any manifold with sufficiently pinched positive curvature and nontrivial topology. Indeed, if all sectional curvatures are close to 1 and  $M$  is not homeomorphic to the sphere, then by the Grove–Shiohama theorem [19] [1, Theorem 1.9], there is  $\varepsilon > 0$  small such that the diameter  $D$  satisfies  $D \leq \pi/2 + \varepsilon$ , in particular  $t_C \leq \pi/2 + \varepsilon$  throughout the whole unit tangent bundle; while by classical comparison theorems [10, Chapter 10]  $t_F \simeq \pi$ . We are not aware of any general result of convexity of injectivity domains on quotients of the round sphere, but if such a property holds then it will survive  $C^3$  perturbations.

But on  $\mathbb{S}^n$  things are made much more tricky by the focalization (in this case  $\text{TFCL} = \text{TCL}$ , i.e. the whole tangent cut locus is focal). If  $M$  is an perturbation of  $\mathbb{S}^n$ , then the tangent focal locus of  $M$  is locally defined by the equation  $\det(d_v \exp_x) = 0$ , so the convexity of the nonfocal domains is guaranteed only if  $d \exp$  is a  $C^2$  perturbation of the “round”  $d \exp$ ; this means that  $g$  should be  $C^4$ -close to the round metric. (The sufficiency of the  $C^2$ -perturbation of  $d \exp$  is easy on  $\mathbb{S}^2$  because we can apply the Implicit Function Theorem. In dimension  $n$  a more subtle reasoning is required, see Section 4 below.)

And focalization is not a rare event for simply connected manifolds: according to a classical result by Klingenberg [24], amplified by Weinstein [35], if  $M$  is a simply connected Riemannian manifold with strictly  $1/4$ -pinched positive sectional curvatures, then the injectivity radius coincides with the conjugate radius, therefore the tangent focal locus and tangent cut locus intersect ( $\text{TFCL} \neq \emptyset$ ). In even dimension, the pinching condition is not necessary and positive curvatures are sufficient. (For surfaces, such results go back to Poincaré himself.)

**Conclusion:** What makes Theorem 1.1 nontrivial is the fact that the sphere is simply connected, allowing focalization at the cut locus; and it is for the same reason that the exponent 4 (rather than 3) is natural, and possibly optimal.

**2.2. The Ma–Trudinger–Wang tensor.** Although the conclusion of Theorem 1.1 is natural and simple, our proof is quite indirect, since it is based on the **Ma–Trudinger–Wang curvature tensor** (MTW tensor), introduced in [29] and further studied in [8, 14, 16, 22, 23, 26, 28, 33, 34], in relation with the regularity theory of optimal transport; see [13] or [33, Chapter 12] for a presentation and survey. Below is a precise definition, with the same conventions as in [33]; we write  $\nabla_x^2 F$  for the Hessian of  $F$  at  $x$ , and  $d$  for the geodesic distance.

**Definition 2.1** (MTW tensor). Let  $(x, y) \in (M \times M) \setminus \text{cut}(M)$ , and  $(\xi, \zeta) \in T_x M \times T_y M$ . Then

- the *pseudo-scalar product* of  $\xi$  and  $\zeta$  is defined by

$$\langle \xi, \zeta \rangle_{x,y} = \langle \xi, \eta \rangle_x, \quad v = (\exp_x)^{-1}(y), \quad \eta = (d_v \exp_x)^{-1}(\zeta);$$

- the MTW *tensor* evaluated on  $(\xi, \zeta)$  is

$$\begin{aligned} \mathfrak{S}_{(x,y)}(\xi, \zeta) &= -\frac{3}{2} \frac{d^2}{ds^2} \Big|_{s=0} \frac{d^2}{dt^2} \Big|_{t=0} \frac{d^2}{2} \left( \exp_x(t\xi), \exp_x(v + s\eta) \right) \\ &= -\frac{3}{2} \frac{d^2}{ds^2} \Big|_{s=0} \left\langle \nabla_x^2 \frac{d^2}{2} \left( \cdot, \exp_x(v + s\eta) \right) \cdot \xi, \xi \right\rangle_x. \end{aligned}$$

The MTW tensor is a nonlocal generalization of *sectional curvature*. Indeed, if  $P$  is a tangent plane included in  $T_x M$ , with orthonormal basis  $\{\xi, \eta\}$ , then  $\mathfrak{S}_{(x,x)}(\xi, \eta)$  is the sectional curvature at  $x$  along  $P$  [33, Particular Case 12.30]. (Kim and McCann [22] reinterpret this tensor as a sectional curvature on the pseudo-Riemannian manifold  $(M \times M, h)$ , where  $h$  is the mixed second-order derivative of the squared distance.)

The **Ma–Trudinger–Wang condition** requires that for all  $(x, y) \in (M \times M) \setminus \text{cut}(M)$ , and  $(\xi, \zeta) \in T_x M \times T_y M$ ,

$$(2.2) \quad \langle \xi, \zeta \rangle_{x,y} = 0 \implies \mathfrak{S}_{(x,y)}(\xi, \zeta) \geq 0.$$

This condition implies nonnegative sectional curvature. It may or may not be satisfied by  $M$ ; and if it is not, this has dramatic consequences on the regularity theory of optimal transport [33, Theorem 12.44].

Various conditions reinforcing the MTW condition (extended or not) have been introduced and studied; they can be thought of as nonlocal variants of the condition of positive lower bound on the sectional curvature. Away from focalization, all these conditions are equivalent, but it is not so in presence of focalization. To state the condition used in this paper, we need first an extended notion of the MTW tensor, and secondly a bit of background in Jacobi fields analysis.

**2.3. The extended Ma–Trudinger–Wang tensor.** In Definition 2.1 the exponential map induces a one-to-one correspondence between  $v \in I(x)$  and  $y \in M \setminus \text{cut}(x)$ , and its differential induces a one-to-one correspondence between  $\eta \in T_x M$  and  $\zeta \in T_y M$ ; so it makes sense to abuse notation by writing, say  $\mathfrak{S}_{(x,y)}(\xi, \zeta) = \mathfrak{S}_{(x,v)}(\xi, \eta)$ . Then the latter object may be extended by letting  $v$  vary in the whole nonfocal domain rather than in the injectivity domain.

To define this extension, we let  $x \in M$ ,  $v \in \text{NF}(x)$ , and  $(\xi, \eta) \in T_x M \times T_x M$ . Since  $y := \exp_x v$  is not conjugate to  $x$ , by the Inverse Function Theorem there are an open neighborhood  $\mathcal{V}$  of  $(x, v)$  in  $TM$ , and an open neighborhood  $\mathcal{W}$  of  $(x, y)$  in  $M \times M$ , such that

$$\begin{aligned} \Psi_{(x,v)} : \mathcal{V} \subset TM &\longrightarrow \mathcal{W} \subset M \times M \\ (x', v') &\longmapsto (x', \exp_{x'}(v')) \end{aligned}$$

is a smooth diffeomorphism from  $\mathcal{V}$  to  $\mathcal{W}$ . Then we may define  $\widehat{c}_{(x,v)} : \mathcal{W} \rightarrow \mathbb{R}$  by

$$(2.3) \quad \widehat{c}_{(x,v)}(x', y') := \frac{1}{2} |\Psi_{(x,v)}^{-1}(x', y')|_{x'}^2 \quad \forall (x', y') \in \mathcal{W}.$$

If  $v \in I(x)$  then for  $y'$  close to  $\exp_x v$  and  $x'$  close to  $x$  we have  $\widehat{c}_{(x,v)}(x', y') = c(x', y') := d(x', y')^2/2$ .

**Definition 2.2** (extended MTW tensor). Let  $x \in M$ ,  $v \in \text{NF}(x)$ , and  $(\xi, \eta) \in T_x M \times T_x M$ . Then the *extended MTW tensor* at  $(x, v)$ , evaluated on  $(\xi, \eta)$ , is

$$\begin{aligned} \overline{\mathfrak{S}}_{(x,v)}(\xi, \eta) &= -\frac{3}{2} \left. \frac{d^2}{ds^2} \right|_{s=0} \left. \frac{d^2}{dt^2} \right|_{t=0} \widehat{c}_{(x,v)}(\exp_x(t\xi), \exp_x(v + s\eta)) \\ &= -\frac{3}{2} \left. \frac{d^2}{ds^2} \right|_{s=0} \left\langle \nabla_x^2 \widehat{c}_{(x,v)}(\cdot, \exp_x(v + s\eta)) \cdot \xi, \xi \right\rangle_x \\ &= -\frac{3}{2} \left. \frac{d^2}{ds^2} \right|_{s=0} \left\langle \nabla_x^2 \widehat{c}_{(x,v+s\eta)}(\cdot, \exp_x(v + s\eta)) \cdot \xi, \xi \right\rangle_x. \end{aligned}$$

We note that  $\nabla_x^2 \widehat{c}_{(x,v)}(x, \exp_x v)$  blows up as  $v$  approaches  $\text{TFL}(x)$ .<sup>2</sup> In contrast, all the  $x$ -derivatives of  $c(x, \exp_x v)$  remain bounded (but not continuous) if  $v$  approaches a nonfocal cut velocity.

---

<sup>2</sup>Beware of confusions:  $\nabla_x^2 \widehat{c}_{(x,v)}(x, \exp_x v)$  means  $\nabla_{x'}^2 \widehat{c}_{(x,v)}(x', y')$  evaluated at  $x' = x$ ,  $y' = \exp_x v$ .

**2.4. Jacobi fields.** Jacobi fields are variations of geodesics [10, Chapter 5]. Given a geodesic  $\gamma$  and a moving reference frame along  $\gamma$ , all Jacobi fields along  $\gamma$  can be reconstructed from two “elementary” matrix-valued functions, which we denote by  $J_0$  and  $J_1$ . In the next statement we use dots for derivatives with respect to the  $t$  variable, and write  $I_n$  for the  $n \times n$  identity matrix,  $0_n$  for the  $n \times n$  zero matrix.

**Definition 2.3** (elementary Jacobi fields). Let  $(x, v) \in TM$ ,  $v \neq 0$ . Let  $(e_1, \dots, e_n)$  be an orthonormal basis of  $T_x M$  with  $e_1 = v/|v|_x$ . For  $t \geq 0$  let  $\gamma(t) = \exp_x(tv)$ , and let  $(e_1(t), \dots, e_n(t))$  be the orthonormal basis of  $T_{\gamma(t)}M$  obtained by parallel transport of  $(e_1, \dots, e_n)$  along  $\gamma$ . Let further, for  $t \geq 0$ ,

$$(2.4) \quad R_{ij}(t) = \left\langle \text{Riem}_{\gamma(t)}(\dot{\gamma}(t), e_i(t))\dot{\gamma}(t), e_j(t) \right\rangle_{\gamma(t)}, \quad 1 \leq i, j \leq n,$$

where  $\text{Riem}$  stands for the Riemann curvature tensor. We define  $J_0(t)$ ,  $J_1(t)$ , implicitly depending on  $x$  and  $v$ , as the matrix-valued solutions of

$$(2.5) \quad \begin{cases} \ddot{J}_i(t) + R(t)J_i(t) = 0, & i = 0, 1, \\ J_0(0) = 0_n, & \dot{J}_0(0) = I_n, \\ J_1(0) = I_n, & \dot{J}_1(0) = 0_n. \end{cases}$$

The **Hessian of the squared distance** can be expressed in terms of  $J_0$  and  $J_1$ :

**Proposition 2.4.** *With the same notation as in Definition 2.3, for  $0 \leq t < t_F(x, v)$ , let  $S(x, v, t)$  be the linear operator  $T_x M \rightarrow T_x M$  whose matrix in the orthonormal basis  $(e_1, \dots, e_n)$  is given by  $tJ_0(t)^{-1}J_1(t)$ ; then this operator is symmetric. If  $v \in I(x)$ , then for any  $\xi \in T_x M$ ,*

$$\begin{aligned} \langle S(x, v, 1)\xi, \xi \rangle_x &= \left\langle \left( \nabla_x^2 \frac{d(\cdot, y)^2}{2} \right) \cdot \xi, \xi \right\rangle_x \\ &= \frac{d^2}{d\tau^2} \Big|_{\tau=0} \frac{d(\exp_x(\tau\xi), y)^2}{2}, \quad y = \exp_x v. \end{aligned}$$

The proof can be found in [33, Chapter 14], see in particular p. 414. This statement is also implicit in [7, Section 2] or [14, Section 2].

For any  $x \in M$ ,  $v \in \text{NF}(x) \setminus \{0\}$ , we write

$$(2.6) \quad S_{(x,v)} = S(x, v, 1) = S\left(x, \frac{v}{|v|_x}, |v|_x\right),$$

so that if  $v \in I(x)$ , then  $S_{(x,v)}$  coincides with  $\nabla_x^2 d(\cdot, \exp_x v)^2/2$ . By mimicking the proof of Proposition 2.4, one easily obtains the following result: if  $v \in \text{NF}(x)$ , then

for any  $\xi \in T_x M$ ,

$$\begin{aligned} \langle S_{(x,v)} \xi, \xi \rangle_x &= \langle \nabla_x^2 \widehat{c}_{(x,v)}(\cdot, y) \cdot \xi, \xi \rangle_x \\ &= \frac{d^2}{d\tau^2} \Big|_{\tau=0} \widehat{c}_{(x,v)}(\exp_x(\tau\xi), y), \quad y = \exp_x v. \end{aligned}$$

In particular, the extended MTW tensor can be computed as follows:

$$(2.7) \quad \overline{\mathfrak{S}}_{(x,v)}(\xi, \eta) = -\frac{3}{2} \frac{d^2}{ds^2} \Big|_{s=0} \langle S_{(x,v+s\eta)} \xi, \xi \rangle_x.$$

In the sequel, we will always use this formula to compute the extended MTW tensor on perturbations of the sphere, see Section 5. Let us further observe that, since the above formula involves only derivatives of Jacobi fields, the (extended) MTW tensor depends on the metric only through its Riemannian curvature. (Compare with Remark 1.2(2).)

Modulo identification,  $J_1(t)$  sends  $T_x M$  to  $T_{\gamma(t)} M$ , then  $J_0(t)^{-1}$  does the reverse; so  $S$  is an endomorphism of  $T_x M$ . Accordingly, we shall never need to consider the moving basis  $(e_1(t), \dots, e_n(t))$ , but only the fixed basis  $(e_1, \dots, e_n)$  which we identify with the canonical basis of  $\mathbb{R}^n$ .

The symmetric matrix  $S$  has an eigenvalue 1 on  $\mathbb{R}e_1$  (the extended squared distance grows quadratically along the geodesic). For the round sphere, all other eigenvalues of  $S^{-1}$  vanish at focalization ( $t = \pi$ ). If the metric  $g$  is close to the round metric  $g^0$  and  $t$  is close to, but strictly less than the focalization time  $t_F(v/|v|_x)$ , we may define  $\Lambda$  by

$$(2.8) \quad S = \begin{bmatrix} 1 & 0 \\ 0 & -\Lambda^{-1} \end{bmatrix}.$$

(More intrinsically,  $\Lambda$  is the inverse of the restriction of  $-S$  to  $(\mathbb{R}e_1)^\perp$ .) Then, for any  $\varepsilon > 0$ , if  $\|g - g^0\|_{C^2} \leq \delta$  and  $t_F - \delta \leq t < t_F$ ,  $\delta = \delta(\varepsilon)$  small enough, we have

$$(2.9) \quad 0 < \Lambda \leq \varepsilon I_{n-1},$$

where  $I_{n-1}$  can be thought of as the identity on  $(\mathbb{R}e_1)^\perp$ . The operator  $\Lambda$  is smooth even at focalization, where its determinant vanishes.

In the sequel we shall abuse notation by writing  $\Lambda^{-1}$  for the operator  $\xi \mapsto \Lambda^{-1} \tilde{\xi}$ , where  $\tilde{\xi}$  is the orthogonal projection of  $\xi$  on  $e_1^\perp$ . Note carefully that while  $\Lambda$  is defined only near the focalization time,  $\Lambda^{-1}$  is defined for any  $v \in \text{NF}(x)$  as the restriction of  $-S$  to  $(\mathbb{R}e_1)^\perp$ .



**2.5. Scheme of proof.** Theorem 1.1 is obtained by concatenating three results of independent interest:

**Theorem 2.5.** *If  $(M, g)$  is a  $C^4$  perturbation of the round sphere, then all its non-focal domains are uniformly convex.*

**Theorem 2.6.** *If  $(M, g)$  is a  $C^4$  perturbation of the round sphere, then it satisfies an extended uniform Ma–Trudinger–Wang condition of the form*

$$(2.10) \quad \forall x \in M, \forall v \in \text{NF}(x) \setminus \{0\}, \\ \mathfrak{S}_{(x,v)}(\xi, \eta) \geq \kappa(|\xi|_x^2 + |v|_x^2 |\Lambda^{-1} \xi|_x^2) |\eta|_x^2 - c \langle \xi, \eta \rangle_x^2,$$

where  $\kappa, c$  are positive constants, and we used the notation defined in Subsection 2.4.

**Theorem 2.7.** *If  $(M, g)$  is a  $C^\infty$  compact Riemannian manifold satisfying (2.10), and all its nonfocal domains are uniformly convex, then all its injectivity domains are also uniformly convex.*

Theorem 2.7 may look surprising and calls for comments. The convexity of the nonfocal domains is a “pseudo-local” property, in the sense that it only depends on the behavior of the metric in the neighborhood of an arbitrary geodesic (before focalization time). The same is true for the positivity of the extended MTW tensor, for instance in the form of (2.10). However, the combination of these two properties will imply a property about injectivity domains, which is of “completely global” nature. So our results can be compared to other theorems relating local curvature conditions to global properties, the most classical being the Bonnet–Myers theorem.

We also note that the positivity of the MTW tensor carries more information than just the convexity of tangent injectivity domains, since it implies continuity results for the solution of optimal transport problems (see [14] for the two-dimensional case, and [16] for the general case).

Theorems 2.5, 2.6 and 2.7 will be proven respectively in Sections 4, 5 and 6. As a preliminary step, in Section 3 we shall establish useful integral representations for variations of Jacobi fields.

A somewhat mysterious step in the proof will be an explicit computation, performed in Subsection 5.4, in which all the potentially dangerous terms in a certain inequality will combine for no apparent reason to form an exact square. This might be the indication of some deeper unexplored structure.

**2.6. Bibliographical notes.** For the convenience of the reader, we shall present self-contained proofs; but our work builds on a number of earlier conceptual contributions. Here is a short account.

The uniform convexity of nonfocal domains on an perturbation of the sphere was already proven by Castelpietra and Rifford [14] with a “symplectic” approach. In this paper we shall provide a more direct Riemannian approach.

After the works of Ma, Trudinger and Wang [29] and Loeper [26], it was known that the positivity of the MTW tensor, together with the convexity of tangent injectivity domains, were sufficient conditions for the regularity of optimal transport on Riemannian manifolds, when the cost function is the squared distance, and cut locus issues are avoided (we refer to the above-mentioned references for precise statements). Then Loeper [27] discovered the relation with sectional curvature. He further showed that the round sphere satisfies a strict form of the MTW condition; this result was improved by Kim and McCann [22], and Figalli and Rifford [14, Appendix].

Loeper and Villani [28] conjectured a general relation between the MTW condition and the shape of the tangent cut locus, and proved that the positivity of the MTW curvature implies the convexity of injectivity domains, under a stringent technical restriction of nonfocality.

The focalization problem was the motivation for several progress: a “probabilistic” perturbation lemma for paths crossing the tangent cut locus [17], and more importantly the introduction of the extended Ma–Trudinger–Wang tensor by Figalli and Rifford [14]. Then in [14] the stability of the extended MTW condition around the round two-dimensional sphere was established, and from there the strict convexity of tangent injectivity domains was deduced. In the present work we shall work in higher dimension, and improve the conclusion from strict to uniform convexity.

Many of the above-mentioned works use an inequality of maximum principle type, introduced by Loeper [26], simplified by Kim and McCann [22], later simplified again and modified in [33, Theorem 12.36] [14, Lemma 3.3] [28, Theorem 3.1]. Another variant of this inequality will be used in Section 6 below.

### 3. VARIATIONS OF JACOBI FIELDS

It is well-known that if the matrix-valued function  $J(t)$  satisfies the Jacobi equation  $\ddot{J} + RJ = 0$ , then  $U = \dot{J}J^{-1}$  solves the Riccati equation  $\dot{U} + U^2 + R = 0$ . In this section, we shall establish equations of a related spirit bearing on first-order variations of Jacobi fields.

**Lemma 3.1.** *With the notation from Definition 2.3,*

$$(a) \quad J_0 J_1^* = J_1 J_0^*.$$

$$(b) \quad \dot{J}_0 \dot{J}_1^* = \dot{J}_1 \dot{J}_0^*.$$

$$(c) \dot{J}_0 J_1^* - \dot{J}_1 J_0^* = I_n.$$

**Lemma 3.2.** *The general solution of the matrix-valued inhomogeneous Jacobi equation*

$$(3.1) \quad \ddot{\mathcal{J}}(t) + R(t) \mathcal{J}(t) = M(t)$$

is given by the formula

$$(3.2) \quad \mathcal{J}(t) = J_1(t) \mathcal{J}(0) + J_0(t) \dot{\mathcal{J}}(0) + J_0(t) \int_0^t J_1(s)^* M(s) ds - J_1(t) \int_0^t J_0(s)^* M(s) ds.$$

*Proof of Lemma 3.1.* Property (a) is equivalent to the symmetry of  $J_1^{-1} J_0$ , which follows from Proposition 2.4. Then by time-differentiation of (a) we get

$$\dot{J}_0 J_1^* + J_0 \dot{J}_1^* = \dot{J}_1 J_0^* + J_1 \dot{J}_0^*,$$

which means that  $\dot{J}_0 J_1^* - \dot{J}_1 J_0^*$  is symmetric. By time-differentiation again, and use of the Jacobi equation and of (a), we obtain

$$\begin{aligned} \frac{d}{dt} (\dot{J}_0 J_1^* - \dot{J}_1 J_0^*) &= \dot{J}_0 \dot{J}_1^* - \dot{J}_1 \dot{J}_0^* + R (J_1 J_0^* - J_0 J_1^*) \\ &= \dot{J}_0 \dot{J}_1^* - \dot{J}_1 \dot{J}_0^*. \end{aligned}$$

This matrix is obviously antisymmetric, but it is also symmetric as the time-derivative of a symmetric matrix; so it vanishes identically, which proves (b). Then  $\dot{J}_0 J_1^* - \dot{J}_1 J_0^*$  is time-independent, and therefore constantly equal to its value at  $t = 0$ , which yields (c).  $\square$

*Proof of Lemma 3.2.* Both sides of (3.2) have the same conditions at  $t = 0$ , so it is sufficient to check that  $\ddot{\mathcal{G}} + R\mathcal{G} = M$ , where  $\mathcal{G}$  is the right-hand side of (3.2). Thanks to the Jacobi equation and Lemma 3.1(a)(c), we have

$$\begin{aligned} \ddot{\mathcal{G}} + R\mathcal{G} &= 2 \left( \dot{J}_0(t) J_1(t)^* - \dot{J}_1(t) J_0(t)^* \right) M(t) \\ &\quad + \left( J_0(t) \dot{J}_1(t)^* - J_1(t) \dot{J}_0(t)^* \right) M(t) \\ &\quad + \left( J_1(t) J_0(t)^* - J_0(t) J_1(t)^* \right) \dot{M}(t) \\ &= M(t) \end{aligned}$$

(observe that  $J_0 \dot{J}_1^* - J_1 \dot{J}_0^* = -(\dot{J}_0 J_1^* - \dot{J}_1 J_0^*)^* = -I_n$ ), and the proof is complete.  $\square$

Now comes the main result of this section:

**Proposition 3.3.** *If  $J_0(\alpha, t)$ ,  $J_1(\alpha, t)$  are a family of Jacobi fields, defined by (2.5), and depending smoothly on an extra parameter  $\alpha$ , then  $J'_i = \partial J_i / \partial \alpha$  and  $\dot{J}_i = \partial J_i / \partial t$  satisfy, whenever  $J_1$  is invertible,*

$$(3.3) \quad J_1^{-1} J'_0 = A_0 - KC, \quad J_1^{-1} J'_1 = C^* - KA_1,$$

$$(3.4) \quad J_1^{-1} \dot{J}_0 = I + T_0 - KD, \quad J_1^{-1} \dot{J}_1 = D^* - KT_1 - KR^0,$$

where

$$(3.5) \quad K(t) = J_1(t)^{-1} J_0(t), \quad I = I_n, \quad R^0 = R(0),$$

$$(3.6) \quad A_0(t) = \int_0^t J_0(s)^* R'(s) J_0(s) ds, \quad A_1(t) = \int_0^t J_1(s)^* R'(s) J_1(s) ds,$$

$$(3.7) \quad T_0(t) = \int_0^t J_0(s)^* \dot{R}(s) J_0(s) ds, \quad T_1(t) = \int_0^t J_1(s)^* \dot{R}(s) J_1(s) ds,$$

$$(3.8) \quad C(t) = \int_0^t J_1(s)^* R'(s) J_0(s) ds, \quad D(t) = \int_0^t J_1(s)^* \dot{R}(s) J_0(s) ds.$$

In particular,  $A_0$ ,  $A_1$ ,  $T_0$ ,  $T_1$ ,  $C$ ,  $D$  have vanishing first row and first column; moreover  $K$ ,  $R^0$ ,  $A_0$ ,  $A_1$ ,  $T_0$ , and  $T_1$  are symmetric.

*Proof of Proposition 3.3.* The fields  $\mathcal{J}_i = J'_i$  and  $\mathcal{G}_i = \dot{J}_i$  satisfy

$$\ddot{\mathcal{J}}_i + R\mathcal{J}_i = -R'J_i, \quad \ddot{\mathcal{G}}_i + R\mathcal{G}_i = -\dot{R}J_i.$$

Then the conclusion follows in a straightforward way from Lemma 3.2.  $\square$

#### 4. CONVEXITY OF NONFOCAL DOMAINS

In this section we prove Theorem 2.5, referring to [28, Appendix B] for some basic properties of uniformly convex sets.

We first note that for the round metric  $g^0$ , we have  $t_F(x, \sigma) = \pi$  for all  $\sigma \in U_x M \simeq \mathbb{S}^{n-1}$  (the space of unit tangent vectors at  $x$ ). The goal is to prove that for any given  $\varepsilon > 0$ , if  $g$  is close enough to  $g^0$  then for all  $\sigma \in U_x M$ ,

$$(4.1) \quad |t_F(x, \sigma) - \pi| \leq \varepsilon, \quad |\nabla_\sigma t_F(x, \sigma)| \leq \varepsilon, \quad \nabla_\sigma^2 t_F(x, \sigma) \leq \varepsilon \text{Id}.$$

The last condition should be interpreted in the weak sense of support functions: to prove (4.1) it is sufficient to show that for every  $\bar{\sigma}$  there is a  $C^2$  function  $\tau = \tau(\sigma)$ , defined in a neighborhood of  $\bar{\sigma}$ , such that

$$\tau(\bar{\sigma}) = t_F(x, \bar{\sigma}), \quad \tau(\sigma) \geq t_F(x, \sigma),$$

and

$$|\tau(\sigma) - \pi| \leq \varepsilon, \quad |\nabla_{\sigma}\tau(\sigma)| \leq \varepsilon, \quad |\nabla_{\sigma}^2\tau(\sigma)| \leq \varepsilon.$$

The inequalities (4.1) will imply that  $\mathbb{I}_{\text{TFL}}(v) \geq 1 - \gamma$  (the second fundamental form of the tangent focal locus, evaluated at  $v$ , defined in weak sense, is bounded below by  $1 - \gamma$ ), where  $\gamma = \gamma(\varepsilon) \rightarrow 0$ . In other words, the tangent focal locus is uniformly convex, and the uniform convexity constant approaches the constant of the round sphere.

For any unit tangent vector  $\sigma$ , we can define  $S(x, \sigma, t)$  as in Proposition 2.4, and  $\Lambda = \Lambda(x, \sigma, t)$  as the restriction of  $-S^{-1}(x, \sigma, t)$  to  $\sigma^{\perp}$  (as in (2.8)). The operator  $\Lambda$  is intrinsically defined, independently of any choice of orthonormal basis  $(e_1 = \sigma, e_2, \dots, e_n)$ , and makes sense near focalization. For the round sphere, by explicit calculation it is equal to  $\Lambda^0 = -(\tan t/t) \text{Id}_{\sigma^{\perp}}$ .

Then it follows from the definition of  $t_F$  that

$$(4.2) \quad t_F(x, \sigma) = \inf \left\{ t_F(x, \sigma; h); h \in U_x M, \langle h, \sigma \rangle_x = 0 \right\},$$

where

$$(4.3) \quad t_F(x, \sigma; h) = \inf \left\{ t \in [3\pi/4, 5\pi/4]; \langle \Lambda(x, \sigma, t)h, h \rangle_x = 0 \right\}.$$

(In the above formula, we need to restrict the set of times to  $t \in [3\pi/4, 5\pi/4]$  to ensure that  $\Lambda$  is well-defined for  $g$  close to  $g^0$  in  $C^2$  topology.) By what we said above, to prove (4.1) it is sufficient to establish that, for any fixed  $h$  in  $T_x M$ ,  $h \perp \sigma$ ,

$$(4.4) \quad |t_F(x, \sigma; h) - \pi| \leq \varepsilon, \quad |\nabla_{\sigma} t_F(x, \sigma; h)| \leq \varepsilon, \quad |\nabla_{\sigma}^2 t_F(x, \sigma; h)| \leq \varepsilon.$$

So let us fix  $\sigma$  and  $h$ , two unit orthogonal tangent vectors, and let

$$q(\sigma, t) = \langle \Lambda(x, \sigma, t)h, h \rangle_x.$$

For the round sphere, this is equal to  $q^0(\sigma, t) = -(\tan t)/t$ , so

$$\left. \frac{\partial q^0}{\partial t} \right|_{t=\pi} = -\frac{1}{\pi} \neq 0.$$

It follows by the Implicit Function Theorem that  $\theta(\sigma) = t_F(x, \sigma; h)$  is well-defined by the implicit equation

$$(4.5) \quad q(\sigma, \theta(\sigma)) = 0$$

in a neighborhood of  $\theta = \pi$  and for  $g$  close to  $g^0$  in  $C^3$  topology. (Since  $q$  depends on second derivatives of  $g$ , this assumption implies that  $q$  is close to  $q^0$  in  $C^1$  topology).

Moreover, differentiating (4.5) we find

$$\frac{\partial\theta}{\partial\sigma} = - \left( \frac{\partial q}{\partial t} \right)^{-1} \frac{\partial q}{\partial\sigma},$$

$$\begin{aligned} \frac{\partial^2\theta}{\partial\sigma^2} &= - \left( \frac{\partial q}{\partial t} \right)^{-1} \left( \frac{\partial^2 q}{\partial\sigma^2} \right) \\ &\quad + \left( \frac{\partial q}{\partial t} \right)^{-2} \left( \frac{\partial^2 q}{\partial\sigma \partial t} \otimes \frac{\partial q}{\partial\sigma} + \frac{\partial q}{\partial\sigma} \otimes \frac{\partial^2 q}{\partial\sigma \partial t} \right) - \left( \frac{\partial q}{\partial t} \right)^{-3} \frac{\partial^2 q}{\partial t^2} \left( \frac{\partial q}{\partial\sigma} \otimes \frac{\partial q}{\partial\sigma} \right). \end{aligned}$$

Since  $\partial q/\partial\sigma$ ,  $\partial^2 q/\partial\sigma^2$  and  $\partial^2 q/\partial t \partial\sigma$  vanish for the round metric, we conclude that  $|\partial\theta/\partial\sigma|$  and  $|\partial^2\theta/\partial\sigma^2|$  are bounded above by  $\varepsilon$  for  $g$  close enough to  $g^0$  in  $C^4$  topology (so  $q$  is close to  $q^0$  in  $C^2$  topology). This concludes the proof.

## 5. STABILITY OF MTW CONDITION

In this section we use Equation (2.7) to compute the extended MTW tensor for the sphere and its perturbations, in order to prove our stability result. Let us remark that, being  $\overline{\mathfrak{S}}_{(x,v)}$  bilinear in both  $\xi$  and  $\eta$ , it suffices to prove the estimate for  $\xi, \eta$  unit tangent vectors at  $x$ . So in this whole section we will always assume that  $|\xi|_x = |\eta|_x = 1$ .

**5.1. Computations.** Let us fix a geodesic  $(\overline{\gamma}(t))_{0 \leq t \leq T}$  with  $\overline{\gamma}(0) = x$ ,  $\overline{\gamma}(1) = \overline{y}$ ,  $\dot{\overline{\gamma}}(0) = \sigma$ ,  $|\sigma| = 1$ . We assume that  $T \leq t_F(\sigma)$ , so that the geodesic is nonfocal except maybe at its final point. We pick up  $(e_1 = \sigma, e_2, \dots, e_n)$  an orthonormal basis of  $T_x M$ , and identify tangent vectors at  $x$  with their coordinates in this basis. Under this identification the metric  $g_x$  is given by the canonical scalar product of  $\mathbb{R}^n$ .

Next, we let  $(\gamma_\alpha(\tau))_{\tau \geq 0}$  be the geodesic starting at  $x$  with initial velocity  $\sigma_\alpha = (\cos \alpha, \sin \alpha, 0, \dots, 0)$ . We further define  $\sigma_\alpha^\perp = (-\sin \alpha, \cos \alpha, 0, \dots, 0)$ .

For any  $\alpha \in \mathbb{R}$ ,  $|\alpha|$  small, and  $\tau \geq 0$ , we solve the Jacobi equation (2.5) with  $e_1(\alpha, 0) = \sigma_\alpha$ ,  $e_2(\alpha, 0) = \sigma_\alpha^\perp$ ,  $e_i(\alpha, 0) = e_i$  for  $i \geq 3$ , and  $R(\alpha, \tau)$  defined by (2.4) evaluated along the geodesic  $(\gamma_\alpha(\tau))_{\tau \geq 0}$ . If  $w = \tau \sigma_\alpha$  with  $\tau < t_F(\sigma_\alpha)$ , then the matrix of  $S_{(x,w)}$  in the orthonormal basis  $(\sigma_\alpha, \sigma_\alpha^\perp, e_3, \dots, e_n)$  is

$$S(\alpha, \tau) = \tau J_0(\alpha, \tau)^{-1} J_1(\alpha, \tau).$$

Let

$$Q(\alpha) = \begin{bmatrix} \cos \alpha & \sin \alpha & 0 & 0 & \dots & 0 \\ -\sin \alpha & \cos \alpha & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 \end{bmatrix}.$$

Then the matrix of  $S_{(x,w)}$  in the original basis of  $\mathbb{R}^n$  is  $Q(-\alpha) S(\alpha, \tau) Q(\alpha)$ ; in other words,

$$(5.1) \quad \langle S_{(x,\tau\sigma_\alpha)} \xi, \xi \rangle = \langle S(\alpha, \tau) Q(\alpha) \xi, Q(\alpha) \xi \rangle.$$

Let now  $v = (t, 0, \dots, 0)$ ,  $\eta = (\eta_1, \eta_2, 0, \dots, 0) \in \mathbb{R}^n \simeq T_x M$ . For any  $s \in \mathbb{R}$ , we can write  $v + s\eta = \tau\sigma_\alpha$ , where

$$(5.2) \quad \tau = |v + s\eta|_x = \sqrt{(t + s\eta_1)^2 + (s\eta_2)^2}, \quad \alpha = \tan^{-1} \left( \frac{s\eta_2}{t + s\eta_1} \right).$$

(Here we used the orthonormality of the basis.) Let us differentiate (5.1) twice with respect to  $s$ :

$$\begin{aligned} \frac{d}{ds} \langle S_{(x,\tau\sigma_\alpha)} \xi, \xi \rangle &= \left[ \left\langle \left( \frac{\partial S}{\partial \alpha} \right) Q\xi, Q\xi \right\rangle + 2 \left\langle S Q\xi, \left( \frac{\partial Q}{\partial \alpha} \right) \xi \right\rangle \right] \left( \frac{\partial \alpha}{\partial s} \right) \\ &\quad + \left\langle \left( \frac{\partial S}{\partial \tau} \right) Q\xi, Q\xi \right\rangle \left( \frac{\partial \tau}{\partial s} \right); \end{aligned}$$

$$\begin{aligned}
(5.3) \quad \frac{d^2}{ds^2} \langle S_{(x, \tau \sigma_\alpha)} \xi, \xi \rangle = & \\
& \left[ \left\langle \left( \frac{\partial^2 S}{\partial \alpha^2} \right) Q\xi, Q\xi \right\rangle + 4 \left\langle \left( \frac{\partial S}{\partial \alpha} \right) \left( \frac{\partial Q}{\partial \alpha} \right) \xi, Q\xi \right\rangle \right. \\
& \quad \left. + 2 \left\langle S \left( \frac{\partial Q}{\partial \alpha} \right) \xi, \left( \frac{\partial Q}{\partial \alpha} \right) \xi \right\rangle + 2 \left\langle S Q\xi, \left( \frac{\partial^2 Q}{\partial \alpha^2} \right) \xi \right\rangle \right] \left( \frac{\partial \alpha}{\partial s} \right)^2 \\
& + \left[ 2 \left\langle \left( \frac{\partial^2 S}{\partial \alpha \partial \tau} \right) Q\xi, Q\xi \right\rangle + 4 \left\langle \left( \frac{\partial S}{\partial \tau} \right) Q\xi, \left( \frac{\partial Q}{\partial \alpha} \right) \xi \right\rangle \right] \left( \frac{\partial \tau}{\partial s} \right) \left( \frac{\partial \alpha}{\partial s} \right) \\
& + \left\langle \left( \frac{\partial^2 S}{\partial \tau^2} \right) Q\xi, Q\xi \right\rangle \left( \frac{\partial \tau}{\partial s} \right)^2 \\
& + \left[ \left\langle \left( \frac{\partial S}{\partial \alpha} \right) Q\xi, Q\xi \right\rangle + 2 \left\langle S Q\xi, \left( \frac{\partial Q}{\partial \alpha} \right) \xi \right\rangle \right] \left( \frac{\partial^2 \alpha}{\partial s^2} \right) \\
& + \left\langle \left( \frac{\partial S}{\partial \tau} \right) Q\xi, Q\xi \right\rangle \left( \frac{\partial^2 \tau}{\partial s^2} \right).
\end{aligned}$$

Moreover, by direct computation, at  $s = 0$  we have

$$(5.4) \quad \tau = t, \quad \frac{\partial \tau}{\partial s} = \eta_1, \quad \frac{\partial^2 \tau}{\partial s^2} = \frac{\eta_2^2}{t}, \quad \alpha = 0, \quad \frac{\partial \alpha}{\partial s} = \frac{\eta_2}{t}, \quad \frac{\partial^2 \alpha}{\partial s^2} = -\frac{2\eta_1 \eta_2}{t^2}.$$

Combining (5.3) with (5.4) we arrive at our final expression for the MTW tensor: writing

$$(5.5) \quad P\xi = (\xi_1, \xi_2, 0, \dots, 0), \quad (P\xi)^\perp = (\xi_2, -\xi_1, 0, \dots, 0),$$



we have (note the minus sign)

$$\begin{aligned}
(5.6) \quad -\frac{2}{3}\overline{\mathfrak{G}}_{(x,v)}(\xi, \eta) &= \left\langle \left( \frac{\partial^2 S}{\partial \tau^2} \right) \xi, \xi \right\rangle \eta_1^2 \\
&+ \left[ \frac{1}{t^2} \left\langle \left( \frac{\partial^2 S}{\partial \alpha^2} \right) \xi, \xi \right\rangle + \frac{1}{t} \left\langle \left( \frac{\partial S}{\partial \tau} \right) \xi, \xi \right\rangle + \frac{4}{t^2} \left\langle \left( \frac{\partial S}{\partial \alpha} \right) (P\xi)^\perp, \xi \right\rangle \right. \\
&\quad \left. + \frac{2}{t^2} \left\langle S (P\xi)^\perp, (P\xi)^\perp \right\rangle - \frac{2}{t^2} \left\langle S\xi, P\xi \right\rangle \right] \eta_2^2 \\
&+ \left[ \frac{2}{t} \left\langle \left( \frac{\partial^2 S}{\partial \alpha \partial \tau} \right) \xi, \xi \right\rangle - \frac{2}{t^2} \left\langle \left( \frac{\partial S}{\partial \alpha} \right) \xi, \xi \right\rangle \right. \\
&\quad \left. + \frac{4}{t} \left\langle \left( \frac{\partial S}{\partial \tau} \right) \xi, (P\xi)^\perp \right\rangle - \frac{4}{t^2} \left\langle S\xi, (P\xi)^\perp \right\rangle \right] \eta_1 \eta_2.
\end{aligned}$$

In the rest of this paper, we shall systematically use a dot to designate a derivative with respect to  $\tau$  (“time”), and a prime to designate a derivative with respect to  $\alpha$ :  $\dot{S} = \partial S / \partial \tau$ ,  $S' = \partial S / \partial \alpha$ , etc.

**5.2. The round sphere.** In this subsection we establish a strict form of the MTW condition for the round sphere  $\mathbb{S}^n$ . We shall not try to get the best possible estimates near focalization; this will be examined more in detail in the next subsection.

If the metric is the round metric, then  $S(\tau, \alpha)$  does not depend on  $\alpha$ , and is equal to

$$S(\tau) = \begin{bmatrix} 1 & 0 \\ 0 & \left( \frac{\tau \cos \tau}{\sin \tau} \right) I_{n-1} \end{bmatrix}.$$

Without loss of generality, we may choose the orthonormal basis  $(e_1, \dots, e_n)$  in such a way that  $v = t e_1$ ,  $\eta = \eta_1 e_1 + \eta_2 e_2$ ,  $\xi = \xi_1 e_1 + \xi_2 e_2 + \xi_3 e_3$ . Then from (5.6),

$$\begin{aligned}
(5.7) \quad \frac{2}{3}\overline{\mathfrak{G}}_{(x,v)}(\xi, \eta) &= 2 \left( \frac{1}{t^2} - \frac{\cos t}{t \sin t} \right) \xi_1^2 \eta_2^2 + 4 \left( \frac{1}{t^2} - \frac{1}{\sin^2 t} \right) \xi_1 \xi_2 \eta_1 \eta_2 \\
&+ 2 \left( \frac{1}{\sin^2 t} - \frac{t \cos t}{\sin^3 t} \right) \xi_2^2 \eta_1^2 + \left( \frac{1}{\sin^2 t} + \frac{\cos t}{t \sin t} - \frac{2}{t^2} \right) \xi_2^2 \eta_2^2 \\
&+ 2 \left( \frac{1}{\sin^2 t} - \frac{t \cos t}{\sin^3 t} \right) \xi_3^2 \eta_1^2 + \left( \frac{1}{\sin^2 t} - \frac{\cos t}{t \sin t} \right) \xi_3^2 \eta_2^2.
\end{aligned}$$

The following elementary inequalities are established in [14, Appendix]: for all  $t \in [0, \pi]$ ,

$$(5.8) \quad \frac{1}{\sin^2 t} - \frac{\cos t}{t \sin t} \geq \frac{2}{3}, \quad \frac{1}{\sin^2 t} - \frac{t \cos t}{\sin^3 t} \geq \frac{1}{3}, \quad \frac{1}{\sin^2 t} + \frac{\cos t}{t \sin t} - \frac{2}{t^2} \geq 0,$$

$$(5.9) \quad \sqrt{\left(\frac{1}{t^2} - \frac{\cos t}{t \sin t} - \alpha\right) \left(\frac{1}{\sin^2 t} - \frac{t \cos t}{\sin^3 t} - \alpha\right)} \geq \frac{1}{\sin^2 t} - \frac{1}{t^2} - \alpha \geq 0,$$

where  $\alpha > 0$  is independent of  $t \in (0, \pi)$ . Moreover, a slightly more refined analysis than the one in [14, Proof of Lemma A.3] allows to show that the third inequality in (5.8) can be improved: there exists  $\beta > 0$  such that

$$(5.10) \quad \frac{1}{\sin^2 t} + \frac{\cos t}{t \sin t} - \frac{2}{t^2} \geq \beta t^2.$$

Hence, combining (5.7)–(5.10), we deduce

$$\begin{aligned} \frac{2}{3} \overline{\mathfrak{G}}_{(x,v)}(\xi, \eta) &\geq 2 \left( \frac{1}{t^2} - \frac{\cos t}{t \sin t} - \alpha \right) \xi_1^2 \eta_2^2 + 4 \left( \frac{1}{t^2} - \frac{1}{\sin^2 t} \right) \xi_1 \xi_2 \eta_1 \eta_2 \\ &\quad + 2 \left( \frac{1}{\sin^2 t} - \frac{t \cos t}{\sin^3 t} - \alpha \right) \xi_2^2 \eta_1^2 + 2\alpha (\xi_1^2 \eta_2^2 + \xi_2^2 \eta_1^2) \\ &\quad + \beta t^2 \xi_2^2 \eta_2^2 + \frac{2}{3} \xi_3^2 (\eta_1^2 + \eta_2^2) \\ &\geq 2 \left( \sqrt{\frac{1}{t^2} - \frac{\cos t}{t \sin t} - \alpha} |\xi_1 \eta_2| - \sqrt{\frac{1}{\sin^2 t} - \frac{t \cos t}{\sin^3 t} - \alpha} |\xi_2 \eta_1| \right)^2 \\ &\quad + 4 \left[ \sqrt{\left(\frac{1}{t^2} - \frac{\cos t}{t \sin t} - \alpha\right) \left(\frac{1}{\sin^2 t} - \frac{t \cos t}{\sin^3 t} - \alpha\right)} \right. \\ &\quad \quad \left. - \left( \frac{1}{\sin^2 t} - \frac{1}{t^2} - \alpha \right) \right] |\xi_1 \xi_2 \eta_1 \eta_2| \\ &\quad + 2\alpha (\xi_1^2 \eta_2^2 + \xi_2^2 \eta_1^2 - 2 \xi_1 \xi_2 \eta_1 \eta_2) + \beta t^2 \xi_2^2 \eta_2^2 + \frac{2}{3} \xi_3^2 (\eta_1^2 + \eta_2^2) \\ &\geq \kappa \left( \xi_1^2 \eta_2^2 + \xi_2^2 \eta_1^2 + \xi_3^2 \eta_1^2 + \xi_3^2 \eta_2^2 - 2 \xi_1 \xi_2 \eta_1 \eta_2 \right) \\ &\quad + \kappa (t^2 \xi_2^2 \eta_1^2 + t^2 \xi_2^2 \eta_2^2 + \xi_3^2 \eta_1^2 + \xi_3^2 \eta_2^2) \\ &= \kappa \left( |\xi|^2 |\eta|^2 - \langle \xi, \eta \rangle^2 \right) + \kappa t^2 |\tilde{\xi}|^2 |\eta|^2, \end{aligned}$$

where  $\tilde{\xi} = \xi_2 e_2 + \xi_3 e_3$ , and  $\kappa > 0$  is a small constant.

**5.3. Computations again.** Now we go back to (5.6), assume the metric to be close to the round metric, and we work near focalization. Before studying the asymptotic behavior of (5.6) we shall rewrite  $\dot{S}$ ,  $\ddot{S}$ ,  $S'$ ,  $S''$  and  $\dot{S}'$  in a suitable way, by means of Section 3.

As a first illustration, let us first take care of  $\dot{S}$ :

$$\begin{aligned}
(5.11) \quad \dot{S} &= -\tau J_0^{-1} \dot{J}_0 J_0^{-1} J_1 + \tau J_0^{-1} \dot{J}_1 + J_0^{-1} J_1 \\
&= -\tau (J_0^{-1} J_1) (J_1^{-1} \dot{J}_0) (J_0^{-1} J_1) + \tau (J_0^{-1} J_1) (J_1^{-1} \dot{J}_1) + J_0^{-1} J_1 \\
&= -\tau K^{-1} (I + T_0 - KD) K^{-1} + \tau K^{-1} (D^* - KT_1 - KR^0) + K^{-1} \\
&= -\tau K^{-1} (I + T_0) K^{-1} + \tau (DK^{-1} + K^{-1} D^*) - \tau (T_1 + R^0) + K^{-1} \\
&= -\tau K^{-1} (I + T_0) K^{-1} + \tau (DK^{-1} + K^{-1} D^*) - \tau (R^0 + T_1) + K^{-1}.
\end{aligned}$$

Of course this expression is symmetric. A similar computation yields

$$\begin{aligned}
(5.12) \quad S' &= -\tau J_0^{-1} J'_0 J_0^{-1} J_1 + \tau J_0^{-1} J'_1 \\
&= -\tau K^{-1} A_0 K^{-1} + \tau (CK^{-1} + K^{-1} C^*) - \tau A_1.
\end{aligned}$$

Then we can iterate the process and derive expressions for second-order variations: thus, using the Jacobi equation (2.5), we get after a bit of algebra:

$$\begin{aligned}
(5.13) \quad \ddot{S} &= -2J_0^{-1} \dot{J}_0 J_0^{-1} J_1 + 2J_0^{-1} \dot{J}_1 + 2\tau \left( J_0^{-1} \dot{J}_0 J_0^{-1} \dot{J}_0 J_0^{-1} J_1 - J_0^{-1} \dot{J}_0 J_0^{-1} \dot{J}_1 \right) \\
&= 2\tau K^{-1} (I + T_0) K^{-1} (I + T_0) K^{-1} \\
&\quad - 2\tau \left( DK^{-1} (I + T_0) K^{-1} + K^{-1} (I + T_0) K^{-1} D^* \right) \\
&\quad - 2K^{-1} (I + T_0) K^{-1} - 2\tau K^{-1} (I + T_0 - KD) (D - T_1 K - R^0 K) K^{-1} \\
&\quad + 2(DK^{-1} + K^{-1} D^*) + 2\tau DK^{-1} D^* - 2(T_1 + R^0).
\end{aligned}$$

Now the symmetry is not obvious, but comes from the identity

$$(I + T_0 - KD)(D - T_1 K - R^0 K) = J_1^{-1} \dot{J}_0 J_1^* (J_1^{-1})^*$$

and Lemma 3.1(b). The other second-order variations will not be “obviously” symmetric either:

(5.14)

$$\begin{aligned} S'' &= 2\tau J_0^{-1} J_0' J_0^{-1} J_0' J_0^{-1} J_1 - \tau J_0^{-1} J_0'' J_0^{-1} J_1 - 2\tau J_0^{-1} J_0' J_0^{-1} J_1' + \tau J_0^{-1} J_1'' \\ &= 2\tau K^{-1} A_0 K^{-1} A_0 K^{-1} - 2\tau (CK^{-1} A_0 K^{-1} + K^{-1} A_0 K^{-1} C^*) + 2\tau CK^{-1} C^* \\ &\quad + \tau K^{-1} \left( J_1^{-1} J_1'' K + 2A_0 A_1 K + 2KC^2 - 2A_0 C - J_1^{-1} J_0'' - 2KCA_1 K \right) K^{-1}. \end{aligned}$$

(5.15)

$$\begin{aligned} \dot{S}' &= -J_0^{-1} J_0' J_0^{-1} J_1 + J_0^{-1} J_1' + \tau J_0^{-1} \dot{J}_0 J_0^{-1} J_0' J_0^{-1} J_1 - \tau J_0^{-1} \dot{J}_0' J_0^{-1} J_1 \\ &\quad + \tau J_0^{-1} J_0' J_0^{-1} \dot{J}_0 J_0^{-1} J_1 - \tau J_0^{-1} J_0' J_0^{-1} \dot{J}_1 - \tau J_0^{-1} \dot{J}_0 J_0^{-1} J_1' + \tau J_0^{-1} \dot{J}_1' \\ &= \tau \left( K^{-1} (I + T_0) K^{-1} A_0 K^{-1} + K^{-1} A_0 K^{-1} (I + T_0) K^{-1} \right) \\ &\quad - \tau (DK^{-1} A_0 K^{-1} + K^{-1} A_0 K^{-1} D^*) \\ &\quad - \tau \left( CK^{-1} (I + T_0) K^{-1} + K^{-1} (I + T_0) K^{-1} C^* \right) + \tau (DK^{-1} C^* + CK^{-1} D^*) \\ &\quad - K^{-1} \left[ \tau (I + T_0) C + \tau J_1^{-1} \dot{J}_0 + \tau A_0 D + A_0 - KC - C^* K \right. \\ &\quad \left. - \tau K (CD + DC) - \tau (A_0 (T_1 + R^0) + (I + T_0) A_1 + J_1^{-1} \dot{J}_1) K \right. \\ &\quad \left. + K (A_1 + \tau C (T_1 + R^0) + \tau D A_1) K \right] K^{-1}. \end{aligned}$$

**5.4. Behavior near focalization.** Let us rewrite (5.6) in the form

$$(5.16) \quad \frac{2}{3} \overline{\Theta}_{(x,v)}(\xi, \eta) = a_{11} \eta_1^2 + a_{22} \eta_2^2 + a_{12} \eta_1 \eta_2,$$

and compute the coefficients  $a_{ij}$ . After some computation, we find

$$(5.17) \quad \begin{aligned} a_{11} &= -2\tau \left\langle K^{-1} (I + T_0) K^{-1} \xi, (I + T_0) K^{-1} \xi \right\rangle - 2\tau \langle K^{-1} D^* \xi, D^* \xi \rangle \\ &\quad + 4\tau \langle K^{-1} (I + T_0) K^{-1} \xi, D^* \xi \rangle + 2 \langle K^{-1} \xi, K^{-1} \xi \rangle + \langle ZK^{-1} \xi, K^{-1} \xi \rangle, \end{aligned}$$

where

$$Z = 2T_0 + 2\tau (I + T_0 - KD)(D - T_1 K - R^0 K) - 4D^* K + 2K(T_1 + R^0)K.$$

Recall from Proposition 3.3 that  $K, A_0, A_1, T_0, T_1, C, D, C^*, D^*, R^0$  all admit  $e_1$  as an eigenvector, and apart from  $K$  the associated eigenvalue is 0. It follows that  $Z$  has vanishing first row and first column. Moreover,  $Z(\pi) = 0$  for the round metric

$g^0$ , so  $Z(\alpha, \tau)$  is very small (say  $|Z| \leq \varepsilon$ ) if  $\tau \simeq \pi$  and  $g \simeq g^0$ . In the sequel, we shall use  $Z$  as a generic symbol for a matrix-valued function satisfying these two properties (vanishing of the first row and column, and smallness as  $\tau \simeq \pi$ ,  $g \simeq g^0$ ).

Similarly,

$$\begin{aligned}
(5.18) \quad a_{22} &= \langle K^{-1}\xi, K^{-1}\xi \rangle - \frac{2}{\tau} \langle K^{-1}A_0K^{-1}\xi, A_0K^{-1}\xi \rangle + \frac{4}{\tau} \langle K^{-1}C^*\xi, A_0K^{-1}\xi \rangle \\
&\quad - \frac{2}{\tau} \langle K^{-1}C^*\xi, C^*\xi \rangle - \frac{2}{\tau} \langle K^{-1}(P\xi)^\perp, (P\xi)^\perp \rangle + \frac{2}{\tau} \langle K^{-1}\xi, P\xi \rangle \\
&\quad - \frac{1}{\tau} \langle K^{-1}\xi, \xi \rangle + \frac{4}{\tau} \langle K^{-1}A_0K^{-1}(P\xi)^\perp, \xi \rangle - \frac{4}{\tau} \langle C^*(P\xi)^\perp, K^{-1}\xi \rangle \\
&\quad - \frac{4}{\tau} \langle (P\xi)^\perp, K^{-1}C^*\xi \rangle + \frac{4}{\tau} \langle A_1(P\xi)^\perp, \xi \rangle + \langle ZK^{-1}\xi, K^{-1}\xi \rangle,
\end{aligned}$$

where

$$\begin{aligned}
Z &= -2D^*K + K(R^0 + T_1)K + T_0 \\
&\quad - \frac{1}{\tau} \left( J_1^{-1}J_1''K + 2A_0A_1K + 2KC^2 - 2A_0C - J_1^{-1}J_0'' - 2KCA_1K \right);
\end{aligned}$$

and

$$\begin{aligned}
(5.19) \quad a_{12} &= 4 \langle (I + T_0)K^{-1}\xi, K^{-1}(P\xi)^\perp \rangle - 4 \langle K^{-1}D^*\xi, (P\xi)^\perp \rangle - 4 \langle DK^{-1}\xi, (P\xi)^\perp \rangle \\
&\quad + 4 \langle (R^0 + T_1)\xi, (P\xi)^\perp \rangle - 4 \langle K^{-1}A_0K^{-1}\xi, (I + T_0)K^{-1}\xi \rangle \\
&\quad + 4 \langle K^{-1}A_0K^{-1}\xi, D^*\xi \rangle + 4 \langle K^{-1}(I + T_0)K^{-1}\xi, C^*\xi \rangle \\
&\quad - 4 \langle K^{-1}C^*\xi, D^*\xi \rangle + \langle ZK^{-1}\xi, K^{-1}\xi \rangle.
\end{aligned}$$

where

$$\begin{aligned}
Z &= \frac{2}{\tau} \left[ \tau(I + T_0)C + \tau J_1^{-1}J_0' + \tau A_0D - KC + C^*K - \tau K(CD + DC) \right. \\
&\quad \left. - \tau \left( A_0(T_1 + R^0) + (I + T_0)A_1 + J_1^{-1}J_1' \right) K + \tau K \left( C(T_1 + R^0) + DA_1 \right) K \right].
\end{aligned}$$

Fifteen (!) of the terms in (5.17)–(5.19) combine in (5.16) to form a “perfect square”:

$$-\frac{2}{\tau} \langle K^{-1}w, w \rangle,$$

where

$$w = \left( -(I + T_0)K^{-1}\xi + D^*\xi \right) \tau \eta_1 + \left( -A_0K^{-1}\xi + C^*\xi + (P\xi)^\perp \right) \eta_2.$$

Recalling (2.8), separating the first component from the rest, we write

$$\begin{aligned} -\frac{2}{\tau}\langle K^{-1}w, w \rangle &= -\frac{2}{\tau^2}\xi_1^2\eta_1^2 - \frac{2}{\tau^2}\xi_2^2\eta_2^2 + \frac{4}{\tau^2}\xi_1\xi_2\eta_1\eta_2 \\ &\quad + \frac{2}{\tau^2}\left|\Lambda^{-\frac{1}{2}}\left(\left((I+T_0)\Lambda^{-1}\xi + \tau D^*\xi\right)\eta_1\right.\right. \\ &\quad \left.\left.+ \left(\left(\tau^{-1}A_0\Lambda^{-1}\xi + C^*\xi\right) + (P\xi)^\perp\right)\eta_2\right)\right|^2, \end{aligned}$$

and

$$\langle K^{-1}\xi, K^{-1}\xi \rangle = \frac{1}{\tau^2}\xi_1^2 + \frac{1}{\tau^2}|\Lambda^{-1}\xi|^2$$

. Thus, we deduce

$$\begin{aligned} \frac{2}{3}\overline{\mathfrak{G}}_{(x,v)}(\xi, \eta) &= \frac{2}{\tau^2}\left|\Lambda^{-\frac{1}{2}}\left(\left((I+T_0)\Lambda^{-1}\xi + \tau D^*\xi\right)\eta_1\right.\right. \\ (5.20) \quad &\quad \left.\left.+ \left(\left(\tau^{-1}A_0\Lambda^{-1}\xi + C^*\xi\right) + (P\xi)^\perp\right)\eta_2\right)\right|^2 \\ &\quad + \frac{1}{\tau^2}|\Lambda^{-1}\xi|^2(2\eta_1^2 + \eta_2^2) + \frac{2}{\tau^2}\xi_1^2\eta_2^2 - \frac{2}{\tau^2}\xi_2^2\eta_2^2 + \frac{4}{\tau^2}\xi_1\xi_2\eta_1\eta_2 \\ &\quad + \left(\langle ZK^{-1}\xi, K^{-1}\xi \rangle + \langle \tilde{Z}K^{-1}\xi, (P\xi)^\perp \rangle\right), \end{aligned}$$

where

$$\tilde{Z} = \frac{4}{\tau}(A_1K - C)\eta_2^2 + 4((R^0 + T_1)K - D)\eta_1\eta_2.$$

(Recall that  $|\eta|_x = 1$ , so  $\tilde{Z}$  is small.) Let us observe that, since  $\tilde{Z}$  has vanishing first row and first column, we have  $|\langle \tilde{Z}\Lambda^{-1}\xi, (P\xi)^\perp \rangle| = |\langle \tilde{Z}\Lambda^{-1}\xi, \xi_1 e_2 \rangle|$ . Furthermore, the two terms coming from  $\tilde{Z}$  can be respectively bounded by

$$c\varepsilon|\Lambda^{-1}\xi||\xi_1|\eta_2^2 \leq c\varepsilon(|\Lambda^{-1}\xi|^2 + \xi_1^2)\eta_2^2$$

and

$$c\varepsilon|\Lambda^{-1}\xi||\xi_1||\eta_1||\eta_2| \leq c\varepsilon(|\Lambda^{-1}\xi|^2\eta_1^2 + \xi_1^2\eta_2^2).$$

Hence, we can control the “dangerous” terms in (5.20):

$$\begin{aligned} |\langle ZK^{-1}\xi, K^{-1}\xi \rangle| &\leq c\varepsilon|\Lambda^{-1}\xi|^2, \\ |\langle \tilde{Z}\Lambda^{-1}\xi, (P\xi)^\perp \rangle| &\leq c\varepsilon|\Lambda^{-1}\xi|^2 + \varepsilon\xi_1^2\eta_2^2, \\ \xi_2^2\eta_2^2 &\leq c\varepsilon|\Lambda^{-1}\xi|^2\eta_2^2, \end{aligned}$$

$$(5.21) \quad 2|\xi_1 \xi_2 \eta_1 \eta_2| \leq \delta^{-1} \xi_2^2 \eta_1^2 + \delta \xi_1^2 \eta_2^2 \leq c \delta^{-1} \varepsilon^2 |\Lambda^{-1} \xi|^2 \eta_1^2 + \delta \xi_1^2 \eta_2^2,$$

where  $\delta$  is small,  $c$  is a positive constant, and we choose  $\varepsilon$  much smaller than  $\delta$  and  $c^{-1}$ . With these bounds we conclude that if  $g$  is close enough to  $g^0$  and  $t$  close enough to  $\pi$ ,

$$(5.22) \quad \frac{2}{3} \overline{\mathfrak{S}}_{(x,v)}(\xi, \eta) \geq \kappa \left( |\Lambda^{-1} \xi|^2 (\eta_1^2 + \eta_2^2) + \xi_1^2 \eta_2^2 + |\Lambda^{-1/2} \omega|^2 \right),$$

where  $\kappa$  is a positive constant, and

$$\omega = ((I + T_0)\Lambda^{-1}\xi + \tau D^*\xi)\eta_1 + \left( (\tau^{-1}A_0\Lambda^{-1}\xi + C^*\xi) + (P\xi)^\perp \right)\eta_2.$$

Let us write  $\widehat{\xi} = (0, 0, \xi_3, \dots, \xi_n)$ . Since  $|\Lambda^{-1}\xi|^2$  controls  $\xi_2^2 + |\widehat{\xi}|^2$ , and thanks to (5.21), up to slightly reducing the value of  $\kappa$  we deduce from (5.22)

$$(5.23) \quad \begin{aligned} \overline{\mathfrak{S}}_{(x,v)}(\xi, \eta) &\geq \kappa \left[ |\Lambda^{-1}\xi|^2 (\eta_1^2 + \eta_2^2) \right. \\ &\quad \left. + \left( \xi_1^2 \eta_2^2 + \xi_2^2 \eta_1^2 - 2\xi_1 \xi_2 \eta_1 \eta_2 + |\widehat{\xi}|^2 (\eta_1^2 + \eta_2^2) \right) + |\Lambda^{-1/2}\omega|^2 \right] \\ &= \kappa \left( |\Lambda^{-1}\xi|^2 |\eta|^2 + (|\xi|^2 |\eta|^2 - \langle \xi, \eta \rangle^2) + |\Lambda^{-1/2}\omega|^2 \right). \end{aligned}$$

**Remark 5.1.** These computations use the fact that all the eigenvalues of  $K$  in  $(\mathbb{R}e_1)^\perp$  vanish simultaneously for the round sphere, so these eigenvalues are still very small for the perturbed sphere. This simultaneous vanishing is of course very particular, but should also be the most degenerate situation. Apart from that, the above arguments do not really take advantage of the closeness to the round metric; we shall see in [15] that, in dimension 2, similar inequalities hold as soon as the nonfocal domains are uniformly convex near the tangent focal cut locus.

**5.5. Improved inequality on the sphere.** Before going on, let us notice that the results of Subsections 5.2 and 5.4 imply a very strong nonlocal curvature inequality on the round sphere:

$$(5.24) \quad \overline{\mathfrak{S}}_{(x,v)}(\xi, \eta) \geq \kappa (|\xi|^2 |\eta|^2 - \langle \xi, \eta \rangle^2) + \kappa |v|^2 (|\widetilde{\xi}|^2 + |\Lambda^{-1}\xi|^2) |\eta|^2,$$

where  $\widetilde{\xi}$  denotes the orthogonal projection of  $\xi$  on  $v^\perp$ .

**Remark 5.2.** The expression in (5.24) is strictly positive as soon as  $v \neq 0$  and  $\xi, \eta$  are not both parallel to  $v$ . In contrast, for  $v = 0$  the right-hand side of (5.24) vanishes as soon as  $\xi$  and  $\eta$  are parallel; this was expected since in this case the MTW curvature reduces to sectional curvature. An informal way to state this conclusion is that *nonlocality improves the curvature of the sphere*. This improvement is all the

more dramatic as we approach the cut locus ( $|v| \rightarrow \pi$ ), since then all the eigenvalues of  $\Lambda^{-1}$  diverge.

**5.6. Stability.** Back to the study of perturbations of the round metric, we can now prove Theorem 2.6.

Let  $\delta > 0$ . We define

$$\Theta_\delta = \left\{ (x, v, \xi, \eta); x \in M, v \in \text{NF}(x), \text{dist}(v, \text{TFL}(x)) \geq \delta, \right. \\ \left. \xi, \eta \in T_x M; |\xi|_x = |\eta|_x = 1, \langle \xi, \eta \rangle_x = 0 \right\}.$$

From (5.6) we see that  $\overline{\mathfrak{S}}_{(x,v)} = \overline{\mathfrak{S}}_{(x,v)}^g$  is a smooth function of the metric  $g$  as  $(x, v, \xi, \eta)$  varies in the compact set  $\Theta_\delta$ . It follows from (5.24) that this function is always positive for the round sphere; so it also has a positive lower bound for  $g$  close enough to  $g^0$  in  $C^4$  topology. We now observe that, if a smooth function of  $(z_1, z')$  is bounded below by  $\kappa$  on  $z_1 = 0$ , then it is bounded below by  $\kappa - c|z_1|$  on a compact set. As a consequence, there exist  $\kappa > 0$  and  $c > 0$  such that for all  $x \in M$ ,  $v \in \text{NF}(x)$  with  $\text{dist}(v, \text{TFL}(x)) \geq \delta$ , and  $\xi, \eta$  unit tangent vectors at  $x$ ,

$$\overline{\mathfrak{S}}_{(x,v)}^g(\xi, \eta) \geq \kappa - c|\langle \xi, \eta \rangle_x|.$$

Hence, since  $|\langle \xi, \eta \rangle_x| \leq 1$ , we obtain

$$\overline{\mathfrak{S}}_{(x,v)}^g(\xi, \eta) \geq \frac{\kappa^2 - c^2 \langle \xi, \eta \rangle_x^2}{\kappa + c|\langle \xi, \eta \rangle_x|} \geq \frac{\kappa^2}{\kappa + c} - \frac{c^2}{\kappa} \langle \xi, \eta \rangle_x^2.$$

Observing that  $\Lambda^{-1}$  is uniformly bounded away from  $\text{TFL}(x)$ , we deduce that (2.10) is satisfied away from the focal locus for a perturbation of the round metric.

On the other hand, by Subsection 5.4, the inequality is true also near the focal locus (again, for a perturbation of the round metric), and Theorem 2.6 follows.

**Remark 5.3.** One may ask whether the stronger inequality (5.23) is also stable under perturbation. Informally, this amounts to asking whether the MTW tensor on the perturbed sphere is positive even when evaluated on non-orthogonal tangent vectors  $\xi, \eta$ . According to Delanoë and Ge [9], the answer is positive in dimension 2.

## 6. CONVEXITY OF INJECTIVITY DOMAINS

In this section we prove Theorem 2.7. We shall need some preparations before we start the core of the proof.



**6.1. Preliminaries.** In this subsection we recall some facts from Riemannian geometry. The first one is the **formula of first variation** [18, Paragraph 3.31]: if  $y \notin \text{cut}(x)$  then

$$(6.1) \quad d_x \left( \frac{d(x, y)^2}{2} \right) \cdot \zeta = -\langle (\exp_x)^{-1}(y), \zeta \rangle_x.$$

With the notation  $c = d(x, y)^2/2$ , this can be reformulated as

$$(6.2) \quad d_x c = -gv, \quad v = (\exp_x)^{-1}(y),$$

where  $(gv)\zeta = g(v, \zeta)$ .

Next, the map

$$(6.3) \quad \phi : (x, v) \longmapsto (\exp_x v, -(d_v \exp_x)v)$$

is an involution between nonfocal tangent vectors. (If  $\gamma(t) = \exp_x(tv)$ , it maps  $(\gamma(0), \dot{\gamma}(0))$  to  $(\gamma(1), -\dot{\gamma}(1))$ .) Then we have the useful formula

$$(6.4) \quad \langle (d_v \exp_x)\xi, \eta \rangle_y = \langle \xi, (d_w \exp_y)\eta \rangle_x, \quad (y, w) = \phi(x, v).$$

Let us briefly recall the proof of (6.4). Let  $\gamma(t) = \exp_x(tv)$ , and let  $X, Y$  be Jacobi fields along  $\gamma$  defined by  $X(0) = 0$ ,  $\dot{X}(0) = \xi$ ,  $Y(1) = 0$ ,  $\dot{Y}(1) = -\eta$ . From the properties of Jacobi fields we have

$$\frac{d}{dt} \langle X(t), \dot{Y}(t) \rangle_{\gamma(t)} = \langle X(t), \ddot{Y}(t) \rangle_{\gamma(t)} + \langle \dot{X}(t), \dot{Y}(t) \rangle_{\gamma(t)} = -\langle X, \text{Riem}(Y, \dot{\gamma})\dot{\gamma} \rangle + \langle \dot{X}, \dot{Y} \rangle,$$

where  $\text{Riem}$  is the Riemann curvature tensor. This quantity being symmetric in  $X$  and  $Y$ ,

$$\frac{d}{dt} \left( \langle X(t), \dot{Y}(t) \rangle_{\gamma(t)} - \langle \dot{X}(t), Y(t) \rangle_{\gamma(t)} \right) = 0,$$

so  $\langle X(t), \dot{Y}(t) \rangle_{\gamma(t)} - \langle \dot{X}(t), Y(t) \rangle_{\gamma(t)}$  is independent of  $t$ . Therefore

$$\langle X(1), \dot{Y}(1) \rangle_{\gamma(1)} - \langle X(0), \dot{Y}(0) \rangle_{\gamma(0)} = \langle Y(1), \dot{X}(1) \rangle_{\gamma(1)} - \langle Y(0), \dot{X}(0) \rangle_{\gamma(0)},$$

so that  $\langle X(1), \dot{Y}(1) \rangle_{\gamma(1)} = -\langle Y(0), \dot{X}(0) \rangle_{\gamma(0)}$ , which is the same as (6.4).

Now we recall a key result about the size of the cut locus. By [25, Corollary 1.3] (see also [21] and [6]), for any  $x \in M$  we have

$$(6.5) \quad \mathcal{H}^{n-1}[K \cap \text{cut}(x)] < +\infty,$$

where  $K \subset M$  is any compact set, and  $\mathcal{H}^{n-1}$  is the  $(n-1)$ -dimensional Hausdorff measure.

As a final preparation, we give a partially coordinate-wise expression for the MTW curvature: if we pick up a coordinate system  $(x_i)_{1 \leq i \leq n}$  around  $x$ , and write  $u_{ij} = \partial^2 u / \partial x^i \partial x^j$ , then

$$(6.6) \quad \overline{\mathfrak{S}}_{(x,v)}(\xi, \eta) = -\frac{3}{2} \left. \frac{d^2}{ds^2} \right|_{s=0} \widehat{c}_{ij}(x, \exp_x(v + s\eta)) \xi^i \xi^j,$$

where  $\widehat{c}(x, \exp_x v)$  should be understood as  $\widehat{c}_{(x,v)}(x, \exp_x v) = |v|_x^2/2$ .

The key observation is that (6.6) is an intrinsic expression, independent of any choice of coordinates (e.g. geodesic), although  $c_{ij} \xi_i \xi_j$  in itself does not make sense unless we specify a choice of coordinates.

To prove (6.6), it will be sufficient to prove this intrinsic property and recall Definition 2.2. But a change of coordinates in the right-hand side of (6.6) induces the replacement of  $\widehat{c}_{ij}(x, \exp_x(v + s\eta))$  by  $\widehat{c}_{ij}(x, \exp_x(v + s\eta)) + \Gamma_{ij}^\ell(x) \widehat{c}_\ell(x, \exp_x(v + s\eta))$ , where  $c_\ell = \partial c / \partial x^\ell$  and  $\Gamma_{ij}^\ell$  are smooth functions. According to (6.2), the extra terms  $\Gamma_{ij}^\ell(x) \widehat{c}_\ell(x, \exp_x(v + s\eta))$  are linear in  $v + s\eta$ , and thus disappear under the action of  $d^2/ds^2$  in (6.6).

**6.2. Main technical ingredients.** The next Proposition is the key to the use of the MTW tensor. It is extracted from [14]; precursors appeared in [22] and [28].

**Proposition 6.1.** *Let  $\bar{x}, x \in M$ , and let  $(p_t)_{t_0 < t < t_1}$  be a  $C^2$  curve valued in  $\text{NF}(\bar{x})$ , such that*

$$\forall t \in (t_0, t_1), \quad \exp_{\bar{x}} p_t \notin \text{cut}(x).$$

Let

$$h(t) = \frac{|p_t|_{\bar{x}}^2}{2} - \frac{d(x, \exp_{\bar{x}} p_t)^2}{2}.$$

Then, with the notation

$$y_t = \exp_{\bar{x}} p_t, \quad q_t = (\exp_{y_t})^{-1}(x), \quad \bar{q}_t = -(d_{p_t} \exp_{\bar{x}})(p_t),$$

if  $[\bar{q}_t, q_t] \subset \text{NF}(y_t)$  for all  $t \in (t_0, t_1)$  then we have

$$(6.7) \quad \dot{h}(t) = \langle q_t - \bar{q}_t, \dot{y}_t \rangle_{y_t},$$

$$(6.8) \quad \ddot{h}(t) = \frac{2}{3} \int_0^1 (1-s) \overline{\mathfrak{S}}_{(y_t, (1-s)\bar{q}_t + sq_t)}(\dot{y}_t, q_t - \bar{q}_t) ds + \langle (d_{\bar{q}_t} \exp_{y_t})(q_t - \bar{q}_t), \ddot{p}_t \rangle_{\bar{x}}.$$

**Remark 6.2.** We shall only use Proposition 6.1 in the case  $x = \bar{x}$ ; then  $h(t)$  measures the difference between the “real” cost  $c = d^2/2$ , and the “extended” cost

$\widehat{c} = \widehat{c}_{(\bar{x}, p_t)}$ , defined in (2.3), evaluated at  $\bar{x}$ . In particular  $h(t) = 0$  if and only if  $p_t \in \overline{I(\bar{x})}$ .

*Proof of Proposition 6.1.* Since  $(y_t)$  does not cross  $\text{cut}(x)$ , the function  $y \mapsto d(x, y)^2$  is smooth, and by (6.1)

$$(6.9) \quad \frac{d}{dt} \left( \frac{d(x, y_t)^2}{2} \right) = -\langle q_t, \dot{y}_t \rangle_{y_t}.$$

On the other hand, since  $p_t = (\exp_{\bar{x}})^{-1}(y_t)$ , recalling (6.4) we have

$$\frac{d}{dt} \frac{|p_t|_{\bar{x}}^2}{2} = \langle p_t, (d_{p_t} \exp_{\bar{x}})^{-1}(\dot{y}_t) \rangle_{\bar{x}} = \langle (d_{\bar{q}_t} \exp_{y_t})^{-1}(p_t), \dot{y}_t \rangle_{y_t} = -\langle \bar{q}_t, \dot{y}_t \rangle_{y_t}.$$

This combined with (6.9) implies (6.7).

To prove (6.8) it will be convenient to introduce coordinates, since intrinsic notation becomes rather cumbersome. So let  $(x_i)_{1 \leq i \leq n}$ ,  $(y_j)_{1 \leq j \leq n}$  be coordinate systems around  $x$  and  $y_t$  respectively. Writing  $\widehat{c} = \widehat{c}_{(\bar{x}, p)}$ , we have

$$p_i = -\widehat{c}_i(\bar{x}, y), \quad \bar{q}_j = -\widehat{c}_j(\bar{x}, y),$$

where  $\widehat{c}_i = \partial \widehat{c} / \partial x^i$  and  $\widehat{c}_j = \partial \widehat{c} / \partial y^j$ . After differentiating and applying the inverse  $\widehat{c}^{i,j}$  of  $\widehat{c}_{i,j} = \partial^2 \widehat{c} / \partial x^i \partial y^j$ , we get, using Einstein's summation convention,

$$\dot{y}^i = -\widehat{c}^{i,j}(\bar{x}, y) \dot{p}_j,$$

$$\ddot{y}^i = -\widehat{c}^{i,k}(\bar{x}, y) \widehat{c}_{k,\ell j}(\bar{x}, y) \dot{y}^\ell \dot{y}^j - \widehat{c}^{i,j}(\bar{x}, y) \ddot{p}_j,$$

where of course  $\widehat{c}_{k,\ell j} = \partial^3 \widehat{c} / \partial x^k \partial y^\ell \partial y^j$ .

Since (6.7) can be rewritten as

$$\dot{h}(t) = (q_i - \bar{q}_i) \dot{y}^i = [\widehat{c}_i(\bar{x}, y) - c_{,i}(x, y)] \dot{y}^i,$$

we deduce, with obvious notation,

$$\begin{aligned} \ddot{h}(t) &= [\widehat{c}_{,ij}(\bar{x}, y) - c_{,ij}(x, y)] \dot{y}^i \dot{y}^j + [\widehat{c}_{,i}(\bar{x}, y) - c_{,i}(x, y)] \ddot{y}^i \\ &= \left[ (\widehat{c}_{,ij}(\bar{x}, y) - c_{,ij}(x, y)) - \widehat{c}_{k,ij} \widehat{c}^{\ell,k}(\bar{x}, y) (q_\ell - \bar{q}_\ell) \right] \dot{y}^i \dot{y}^j + \widehat{c}^{i,j}(\bar{x}, y) (\bar{q}_i - q_i) \ddot{p}_j. \end{aligned}$$

Noting that  $c(x, y_t) = \widehat{c}_{(y_t, q_t)}(y_t, \exp_{y_t} q_t)$ , we see that the expression inside square brackets coincides with

$$\begin{aligned} & - \left[ \widehat{c}_{ij}(y_t, \exp_{y_t} q_t) - \widehat{c}_{ij}(y_t, \exp_{y_t} \bar{q}_t) - d_{\bar{q}_t} \widehat{c}_{ij}(y_t, \exp_{y_t} \bar{q}_t) \cdot (q_t - \bar{q}_t) \right] \\ & = - \int_0^1 (1-s) \frac{\partial^2}{\partial s^2} \widehat{c}_{ij}(y_t, \exp_{y_t} [(1-s)\bar{q}_t + sq_t]) ds. \end{aligned}$$

Then we apply (6.6) to conclude that

$$\ddot{h}(t) = \frac{2}{3} \int_0^1 (1-s) \overline{\mathfrak{S}}_{(y_t, (1-s)\bar{q}_t + sq_t)}(\dot{y}_t, q_t - \bar{q}_t) ds + \widehat{c}^{i,j}(\bar{x}, y) (\bar{q}_i - q_i) \ddot{p}_j,$$

and (6.8) follows.  $\square$

The next perturbation lemma is borrowed from [17].

**Lemma 6.3.** *Let  $\bar{x}, x \in M$ , and let  $(p_t)_{0 \leq t \leq 1}$  be a smooth path valued in  $\text{NF}(\bar{x})$ . Then for any  $\varepsilon > 0$  there is a path  $(\tilde{p}_t)_{0 \leq t \leq 1}$  such that*

$$\|p - \tilde{p}\|_{C^2(0,1)} \leq \varepsilon$$

and

$$\left\{ t; \exp_{\bar{x}} \tilde{p}_t \in \text{cut}(x) \right\} \quad \text{is finite.}$$

*Proof of Lemma 6.3.* First, since  $(p_t)$  is valued in  $\text{NF}(\bar{x})$ , it is equivalent to perturb  $p_t$  in the  $C^2$  topology, or to perturb  $y_t := \exp_{\bar{x}} p_t$  in the  $C^2$  topology. So in the sequel we shall construct  $\tilde{y}_t = \exp_{\bar{x}} \tilde{p}_t$ .

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  $(\widehat{y}_t)$ , such that  $\widehat{y}_{t_0} \notin \text{cut}(x)$ . Repeating this operation finitely many times, we can ensure that  $\widehat{y}_{t_j}$  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  $(\widehat{y}_t)_{t_j \leq t \leq t_{j+1}}$  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  $\widehat{y}_t$  is entirely contained in the small ball  $B_j = B(\widehat{y}_{t_j}, r_j)$ , and the larger ball  $2B_j = B(\widehat{y}_{t_j}, 2r_j)$  does not meet  $\text{cut}(x)$ .

If we prove that the path  $(\widehat{y}_t)$  can be approximated on each interval  $[t_{j-1} + \varepsilon_{j-1}, t_j - \varepsilon_j]$  by a path  $(\tilde{y}_t)$  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 perturbation 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 by (6.5). 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  $\Sigma$  in at most finitely many points. To do this one can

apply the co-area formula in the following form (see [11, p. 109] and [12, Sections 2.10.25 and 2.10.26]): let  $f : (z, t) \mapsto z$  (defined on  $U$ ), then

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

Thanks to (6.5) 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  $\Sigma$  finitely many often.  $\square$

Now comes a maximum principle type lemma, borrowed from [33].

**Lemma 6.4.** *Let  $f : [0, 1] \rightarrow \mathbb{R}$  be a semiconcave function. Assume that there are  $0 = t_0 < t_1 < \dots < t_N = 1$  such that for any  $j \in \{0, \dots, N-1\}$ ,  $f$  is twice continuously differentiable on  $(t_j, t_{j+1})$  and satisfies*

$$(6.10) \quad \ddot{f} \leq C |\dot{f}(t)|$$

for some constant  $C \geq 0$ . Then

$$(6.11) \quad \forall t \in [0, 1], \quad f(t) \geq \min(f(0), f(1)).$$

*Proof of Lemma 6.4.* Let  $\varepsilon > 0$  and  $f_\varepsilon(t) = f(t) - \varepsilon t^k$ , where  $k \in \mathbb{N}$  is such that  $k > C + 2$ . Then

$$\begin{aligned} \ddot{f}_\varepsilon(t) &= -\varepsilon k(k-1)t^{k-2} + \ddot{f}(t) \leq -\varepsilon k(k-1)t^{k-2} + C |\dot{f}(t)| \\ &\leq -\varepsilon k(k-1)t^{k-2} + C |\dot{f}_\varepsilon(t)| + C \varepsilon k t^{k-1} = -\varepsilon k(k-1-Ct)t^{k-2} + C |\dot{f}_\varepsilon(t)|. \end{aligned}$$

So

$$(6.12) \quad \ddot{f}_\varepsilon(t) \leq -\varepsilon k t^{k-2} + C |\dot{f}_\varepsilon(t)|.$$

Let now  $t_* \in [0, 1]$  be such that  $f_\varepsilon$  is minimum at  $t_*$ . If  $t_* \in (t_j, t_{j+1})$  for some  $j \in \{0, \dots, N-1\}$  then  $\dot{f}_\varepsilon(t_*) = 0$ , so by (6.12)  $\ddot{f}_\varepsilon(t_*) < 0$ , which is impossible. Thus  $t_* = t_j$  for some  $j \in \{0, \dots, N\}$ .

Let us assume that  $j \in \{1, \dots, N-1\}$ . If  $\dot{f}_\varepsilon$  is discontinuous at  $t_j$ , then by semiconcavity  $\dot{f}_\varepsilon(t_j^-) > \dot{f}_\varepsilon(t_j^+)$ , which is incompatible with  $t_j$  being a minimum of  $f_\varepsilon$ . If on the contrary  $\dot{f}_\varepsilon$  is continuous at  $t_j$ , then by semiconcavity again  $f_\varepsilon$  is differentiable at  $t_j$ . Because  $t_j$  is a minimum,  $\dot{f}_\varepsilon(t_j) = 0$ , and by (6.12),  $\ddot{f}_\varepsilon < 0$  in a neighborhood of  $t_j$ , so  $\dot{f}_\varepsilon$  is positive on the left of  $t_j$  and negative on the right of  $t_j$ , which again is impossible.

We conclude that  $j \in \{0, N\}$  i.e.  $f_\varepsilon \geq \min(f_\varepsilon(0), f_\varepsilon(1))$ , and the claim follows by letting  $\varepsilon \rightarrow 0$ .  $\square$

**6.3. Proof of Theorem 2.7.** Let  $M$  satisfy the assumptions of Theorem 2.7. Let  $x \in M$  and  $p_0, p_1 \in I(x)$ . Fix  $\delta > 0$  to be chosen later, and let  $(p_t)_{0 \leq t \leq 1}$  be a path valued in  $T_x M$ , joining  $p_0$  to  $p_1$ , such that  $|\ddot{p}_t|_x \leq \delta |p_0 - p_1|_x^2$ . If we can show that  $p_t \in \overline{I(x)}$  for all  $t \in [0, 1]$ , then this will imply the uniform convexity of  $I(x)$ .

Since  $I(x) \subset \text{NF}(x)$  and the latter set is assumed uniformly convex, we know that for  $\delta$  small enough  $(p_t)$  is valued in  $\text{NF}(x)$ . By Lemma 6.3, for any  $\delta' > \delta$  and any  $\varepsilon > 0$  we may find a path  $(\tilde{p}_t)$ , also valued in  $\text{NF}(x)$ , such that  $|p_t - \tilde{p}_t|_x \leq \varepsilon$ ,  $|\ddot{\tilde{p}}_t|_x \leq \delta' |\tilde{p}_0 - \tilde{p}_1|_x^2$ ,  $\tilde{p}_0, \tilde{p}_1 \in I(x)$ , and  $\exp_x \tilde{p}_t$  meets  $\text{cut}(x)$  only for finitely many times  $t$ . If we can prove that  $(\tilde{p}_t)$  is valued in  $I(x)$  then we are done.

In the sequel, for simplicity we shall note  $\delta$  for  $\delta'$  and  $p_t$  for  $\tilde{p}_t$ .

Let

$$(6.13) \quad \ell(t) = \frac{d(x, y_t)^2}{2} - \frac{|p_t|_x^2}{2}, \quad y_t = \exp_x p_t.$$

Let  $j \in \{0, \dots, N-1\}$ . When  $t$  varies in  $(t_j, t_{j+1})$  we may define

$$\bar{q}_t = -(d_{p_t} \exp_x) p_t, \quad q_t = (\exp_{y_t})^{-1}(x).$$

Then by Proposition 6.1 (using the convexity of  $\text{NF}(y_t)$  for all  $t$ )

$$(6.14) \quad \dot{\ell}(t) = -\langle \dot{y}_t, q_t - \bar{q}_t \rangle_{y_t},$$

$$(6.15) \quad \ddot{\ell}(t) = -\frac{2}{3} \int_0^1 (1-s) \overline{\mathfrak{S}}_{(y_t, (1-s)\bar{q}_t + sq_t)}(\dot{y}_t, q_t - \bar{q}_t) ds - \langle (d_{\bar{q}_t} \exp_{y_t})(q_t - \bar{q}_t), \ddot{p}_t \rangle_x.$$

So our curvature assumptions imply

$$(6.16) \quad \begin{aligned} \ddot{\ell}(t) \leq & -\frac{\kappa}{3} (|\dot{y}_t|_{y_t}^2 + |\Lambda^{-1} \dot{y}_t|_{y_t}^2) |q_t - \bar{q}_t|_{y_t}^2 \\ & + \frac{c}{3} \langle \dot{y}_t, q_t - \bar{q}_t \rangle_{y_t}^2 + |(d_{\bar{q}_t} \exp_{y_t})(q_t - \bar{q}_t)|_x |\ddot{p}_t|_x. \end{aligned}$$

At this point, we note that

$$\begin{aligned} |p_0 - p_1|_x & \leq \int_0^1 |\dot{p}_s|_x ds \leq |\dot{p}_t|_x + \int_0^1 |\ddot{p}_s|_x ds \\ & \leq |\dot{p}_t|_x + \delta |p_0 - p_1|_x^2 \leq |\dot{p}_t|_x + 2\delta \text{diam}(M) |p_0 - p_1|_x; \end{aligned}$$

so if  $\delta \leq (4 \operatorname{diam}(M))^{-1}$  then

$$(6.17) \quad |\dot{p}_t|_x \geq (1 - 2\delta \operatorname{diam}(M)) |p_0 - p_1|_x \geq \frac{|p_0 - p_1|_x}{2}.$$

Also, recalling the definition of  $\Lambda$  from (2.8), it is easily seen that

$$(6.18) \quad |\dot{y}_t|_{y_t}^2 + |\Lambda^{-1} \dot{y}_t|_{y_t}^2 \geq \nu |\dot{p}_t|_x^2$$

for some constant  $\nu > 0$ .

Next, by Taylor's formula, for  $|q_t - \bar{q}_t|_{y_t} \leq \alpha$  small enough, the equality  $\exp_{y_t} q_t = \exp_{y_t} \bar{q}_t$  implies

$$(6.19) \quad \left| (d_{\bar{q}_t} \exp_{y_t})(q_t - \bar{q}_t) \right|_x \leq B |q_t - \bar{q}_t|_{y_t}^2$$

for some constant  $B > 0$ ; this inequality is also obviously true for  $|q_t - \bar{q}_t|_{y_t} \geq \alpha$ .

Combining (6.17), (6.18) and (6.19) with (6.14) and (6.15), we deduce

$$(6.20) \quad \begin{aligned} \ddot{\ell}(t) &\leq -\left(\frac{\kappa\nu}{12} - B\delta\right) |q_t - \bar{q}_t|_{y_t}^2 |p_0 - p_1|_x^2 + \frac{c}{3} |\dot{\ell}(t)|^2 \\ &\leq -\left(\frac{\kappa\nu}{12} - B\delta\right) |q_t - \bar{q}_t|_{y_t}^2 |p_0 - p_1|_x^2 + C |\dot{\ell}(t)|, \end{aligned}$$

where the constant  $C$  is an upper bound for  $(c/3) \sup_t |\dot{y}_t| |q_t - \bar{q}_t|_{y_t}$  (which depends only on  $M$  if  $\delta$  is sufficiently small, and is of order  $c \operatorname{diam}(M)^2$ ).

If  $\delta$  is small enough then (6.20) implies  $\ddot{\ell}(t) \leq C |\dot{\ell}(t)|$  for  $t \in (t_j, t_{j+1})$ . Hence, since  $y \mapsto d(x, y)^2$  is semiconcave, we may apply Lemma 6.4 to deduce

$$\forall t \in (0, 1), \quad \ell(t) \geq \min(\ell(0), \ell(1)).$$

But since  $p_0, p_1 \in I(x)$  we have  $d(x, \exp_x p_i) = |p_i|_x$  for  $i = 0, 1$ ; that is,  $\ell(0) = \ell(1) = 0$ . It follows that  $\ell(t) \geq 0$  for all  $t$ , i.e.

$$|p_t|_x^2 \leq d(x, \exp_x p_t)^2.$$

The reverse inequality is obviously true, so  $p_t$  is a minimizing velocity, that is  $p_t \in \overline{I(x)}$ , and the proof is complete.

**Remark 6.5.** In this section we have shown that, if all  $\operatorname{NF}(x)$  are uniformly convex, and the strong version of the extended MTW condition given in (2.10) holds, then all  $I(x)$  are uniformly convex. It is actually possible to prove also some “weaker” versions of this result, which are important for applications to the regularity theory of optimal transport:

- If all  $\operatorname{NF}(x)$  are convex, and  $\overline{\mathfrak{S}}_{(x,v)}(\xi, \eta) \geq 0$  for all  $\xi \perp \eta$ , then all  $I(x)$  are convex.

- If all  $\text{NF}(x)$  are strictly convex, and  $\overline{\mathfrak{S}}_{(x,v)}(\xi, \eta) \geq 0$  for all  $\xi \perp \eta$  with strict inequality unless  $\xi = 0$  or  $\eta = 0$ , then all  $\text{I}(x)$  are strictly convex.

We refer to [16] for a proof of these results.

## REFERENCES

- [1] U. ABRESCH AND W.T. MEYER. Injectivity radius estimates and sphere theorems. *Comparison geometry*. Math. Sci. Res. Inst. Publ. Vol. 30, Cambridge Univ. Press, 1997, pp. 1–47.
- [2] M. BERGER. *A panoramic view of Riemannian geometry*. Springer-Verlag, Berlin, 2003.
- [3] M.A. BUCHNER. Stability of the cut locus in dimensions less than or equal to 6. *Invent. Math.* 43, 3 (1977), 199–231.
- [4] M.A. BUCHNER. The structure of the cut locus in dimension less than or equal to six. *Compositio Math.* 37, 1 (1978), 103–119.
- [5] P. CANNARSA AND C. SINISTRARI. *Semiconcave functions, Hamilton-Jacobi equations, and optimal control*. Progress in Nonlinear Differential Equations and their Applications, 58. Birkhäuser Boston Inc., Boston, 2004.
- [6] M. CASTELPIETRA AND L. RIFFORD. Regularity properties of the distance function to conjugate and cut loci for viscosity solutions of Hamilton–Jacobi equations and applications in Riemannian geometry. *ESAIM Control Optim. Calc. Var.* 16, 3 (2010), 695–718.
- [7] D. CORDERO-ERAUSQUIN, R.J. MCCANN AND M. SCHMUCKENSHLÄGER. A Riemannian interpolation inequality à la Borell, Brascamp and Lieb. *Invent. Math.* 146, 2 (2001), 219–257.
- [8] PH. DELANOË AND Y. GE. Regularity of optimal transportation maps on compact, locally nearly spherical, manifolds. *J. Reine Angew. Math.* 646 (2010), 65–115.
- [9] PH. DELANOË AND Y. GE. Positivity of  $c$ -curvature on surfaces locally near to sphere and its application.
- [10] M.P. DO CARMO. *Riemannian geometry*. Birkhäuser, Boston, MA, 1992.
- [11] L.C. EVANS AND R. GARIEPY. *Measure theory and fine properties of functions*. CRC Press, Boca Raton, FL, 1992.
- [12] H. FEDERER. *Geometric measure theory*. Die Grundlehren der mathematischen Wissenschaften, Band 153. Springer-Verlag New York Inc., New York, 1969.
- [13] A. FIGALLI. Regularity of optimal transport maps (after Ma-Trudinger-Wang and Loeper). *Séminaire Bourbaki*. Vol. 2008/2009, Exp. No. 1009.
- [14] A. FIGALLI AND L. RIFFORD. Continuity of optimal transport maps and convexity of injectivity domains on small deformations of  $\mathbb{S}^2$ . *Comm. Pure Appl. Math.* 62, (12) (2009), 1670–1706.
- [15] A. FIGALLI, L. RIFFORD AND C. VILLANI. On the Ma–Trudinger–Wang curvature tensor on surfaces. *Calc. Var. Partial Differential Equations* 39, 3-4 (2010), 307–332.
- [16] A. FIGALLI, L. RIFFORD AND C. VILLANI. Necessary and sufficient conditions for continuity of optimal transport maps on Riemannian manifold. Preprint, 2010.
- [17] A. FIGALLI AND C. VILLANI. An perturbation lemma about the cut locus, with applications in optimal transport theory. *Meth. Appl. Anal.* 15, 2 (2008), 149–154.
- [18] S. GALLOT, D. HULIN AND J. LAFONTAINE. *Riemannian geometry*, second ed. Universitext. Springer-Verlag, Berlin, 1990.



- [19] K. GROVE AND K. SHIOHAMA. A generalized sphere theorem. *Ann. of Math.* 106, 2 (1977), 201–211.
- [20] J.-I. ITOH AND K. KIYOHARA. The cut loci and the conjugate loci on ellipsoids. *Manuscripta Math.* 114, 2 (2004), 247–264.
- [21] J.-I. ITOH AND M. TANAKA. The Lipschitz continuity of the distance function to the cut locus. *Trans. Amer. Math. Soc.* 353, 1 (2001), 21–40.
- [22] Y.H. KIM AND R.J. MCCANN. Continuity, curvature, and the general covariance of optimal transportation. *J. Eur. Math. Soc. (JEMS)* 12 (2010), 1009–1040.
- [23] Y.H. KIM AND R.J. MCCANN. Towards the smoothness of optimal maps on Riemannian submersions and Riemannian products (of round spheres in particular). To appear in *J. Reine Angew. Math.*
- [24] W. KLINGENBERG. Über Riemannsche Mannigfaltigkeiten mit positiver Krümmung. *Comment. Math. Helv.* 35 (1961), 47–54.
- [25] Y. LI AND L. NIRENBERG. 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, 1 (2005), 185–146.
- [26] G. LOEPER. On the regularity of solutions of optimal transportation problems. *Acta Math.* 202 (2009), 241–283.
- [27] G. LOEPER. Regularity of optimal maps on the sphere: The quadratic cost and the reflector antenna. *Arch. Ration. Mech. Anal.* 199, 1 (2011), 269–289.
- [28] G. LOEPER AND C. VILLANI. Regularity of optimal transport in curved geometry: the non-focal case. Preprint, 2008.
- [29] X.N. MA, N.S. TRUDINGER AND X.J. WANG. Regularity of potential functions of the optimal transportation problem. *Arch. Ration. Mech. Anal.* 177, 2 (2005), 151–183.
- [30] C. MANTEGAZZA AND A.C. MENNUCCI. Hamilton–Jacobi equations and distance functions on Riemannian manifolds. *Appl. Math. Optim.* 47, 1 (2003), 1–25.
- [31] H. POINCARÉ. Sur les lignes géodésiques des surfaces convexes. *American M. S. Trans.* 6 (1905), 237–274.
- [32] D. SINGER AND H. GLUCK. The existence of nontriangulable cut loci. *Bull. Amer. Math. Soc.* 82, 4 (1976), 599–602.
- [33] C. VILLANI. Optimal transport, old and new. *Grundlehren des mathematischen Wissenschaften*, Vol. 338, Springer-Verlag, Berlin-New York, 2009.
- [34] C. VILLANI. Stability of a 4th-order curvature condition arising in optimal transport theory. *J. Funct. Anal.* 255, 9 (2008), 2683–2708.
- [35] A. WEINSTEIN. The cut locus and conjugate locus of a Riemannian manifold. *Ann. of Math.* 87 (1968), 29–41.

ALESSIO FIGALLI

LABORATOIRE DE MATHÉMATIQUES LAURENT SCHWARTZ, UMR 7640  
ÉCOLE POLYTECHNIQUE  
91128 PALAISEAU CEDEX

A. FIGALLI, L. RIFFORD, AND C. VILLANI

FRANCE

EMAIL: [figalli@math.polytechnique.fr](mailto:figalli@math.polytechnique.fr)

LUDOVIC RIFFORD

UNIVERSITÉ DE NICE–SOPHIA ANTIPOLIS  
LABO. J.-A. DIEUDONNÉ, UMR 6621  
PARC VALROSE, 06108 NICE CEDEX 02  
FRANCE

EMAIL: [rifford@unice.fr](mailto:rifford@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@math.univ-lyon1.fr](mailto:cvillani@math.univ-lyon1.fr)