SIMPLE RIEMANNIAN SURFACES ARE SCATTERING RIGID

HAOMIN WEN[+]

ABSTRACT. Scattering rigidity of a Riemannian manifold allows one to tell the metric of a manifold with boundary by looking at the directions of geodesics at the boundary. Lens rigidity allows one to tell the metric of a manifold with boundary from the same information plus the length of geodesics. There are a variety of results about lens rigidity but very little is known for scattering rigidity. We will discuss the subtle difference between these two types of rigidities and prove that they are equivalent for two-dimensional simple manifolds with boundaries. In particular, this implies that two-dimensional simple manifolds (such as the flat disk) are scattering rigid since they are lens/boundary rigid (Pestov–Uhlmann, 2005).

1. INTRODUCTION

1.1. **The invisible Eaton lens.** The invisible Eaton lens [8, 11, 13] is a gradient-index (GRIN) lens that looks like the vacuum from the outside, but has an infinite refractive index at the center. The refractive index n of the invisible Eaton lens is given by

$$\sqrt{n} = \frac{1}{nr} + \sqrt{\frac{1}{n^2 r^2} - 1}.$$

One can think the Eaton lens as a Riemannian manifold with the conformally flat metric $n^2 g_0$ on the unit disk. The metric has a singularity at the center, and the trajectories of light will be geodesics in that Riemannian manifold.

As can be seen from Figure 1, the direction of each light ray when entering the lens is the same as the direction of the light ray when leaving the lens. Hence there is no refraction visible from the outside, (even though each light ray makes a complete circuit inside the Eaton lens,) and thus this Eaton lens is invisible.
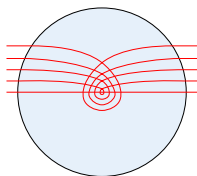

FIGURE 1. Trajectories of light in an invisible Eaton lens

Question 1.1. *Can we have an invisible lens without singularities?*

1.2. **Scattering rigidity and lens rigidity.** Question 1.1 is equivalent to asking if flat balls are scattering rigid. Simply put, a Riemannian manifold M is scattering rigid if M is determined by its scattering data (see below) up to isometries which leave the boundary fixed.

Let $\pi : \Omega M \to M$ be the unit tangent bundle of M and $\Omega_x M$ be the set of unit tangent vectors at x for any $x \in M$. Let $\partial \Omega M$ be the boundary of the unit tangent bundle of M. In other words, $\partial \Omega M = \bigcup_{x \in \partial M} \Omega_x M$. For each $x \in \partial M$, let $\nu_M(x)$ be the unit normal vector of M pointing inwards at x. Then put $\partial_+ \Omega_x M = \{X \in \Omega_x M : (X, \nu_M(x))_{g_M} > 0\}$, $\partial_0 \Omega_x M = \{X \in \Omega_x M : (X, \nu_M(x))_{g_M} = 0\}$, and $\partial_- \Omega_x M = \{X \in \Omega_x M : (X, \nu_M(x))_{g_M} < 0\}$. Also, write $\partial_+ \Omega M = \bigcup_{x \in \partial M} \partial_+ \Omega_x M$, $\partial_0 \Omega M = \bigcup_{x \in \partial M} \partial_0 \Omega_x M$, and $\partial_- \Omega M = \bigcup_{x \in \partial M} \partial_- \Omega_x M$.

For each $X \in \partial_+ \Omega M$, there is a geodesic γ_X whose initial tangent vector is X. Extend the geodesic as long as possible until it touches the boundary ∂M again. Let $\tau_X := \ell(\gamma_X)$, the length of γ_X.

If the geodesic γ_X is of finite length, call its tangent vector at the other end point $\alpha_M(X)$. (See Figure 2.) The map $\alpha_M : \partial_+ \Omega M \to \partial \Omega M$ defined above is called the *scattering relation* of M. Note that $\alpha_M(X)$ will be undefined if γ_X is of infinite length.
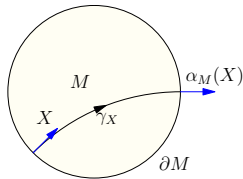

FIGURE 2. The scattering map α_M

Suppose that we have two Riemannian manifolds (M, g_M), (N, g_N) and an isometry $h : \partial M \to \partial N$ between their boundaries. Then there is a natural bundle map $\varphi : \partial \Omega M \to \partial \Omega N$ defined as

(1.1)
$$\varphi(aX + b\nu_M(x)) = ah_*(X) + b\nu_N(h(x))$$

for any unit vector X based at x tangent to ∂M and real numbers a and b such that $a^2 + b^2 = 1$. M and N are said to have the same *scattering data rel h* if $\varphi \circ \alpha_M = \alpha_N \circ \varphi$. If we also have $\ell(\gamma_X) = \ell(\gamma_{\varphi(X)})$, then we say M and N have the same *lens data rel h*.

Definition 1.2. *We say a Riemannian manifold M is* scattering rigid *(resp.* lens rigid*) if for any Riemannian manifold N which has the same scattering data as M (resp. lens data) rel h, (where $h : \partial M \to \partial N$ is an isometry,) we can always extend h to an isometry from M to N.*

We will omit "rel h" when h is clear from the context or the specific choice of h does not matter.

Question 1.3 (Equivalent to Question 1.1). *Are flat balls scattering rigid?*

Remark 1.4. Theorem 1.10 (below) shows that 2-D flat disks are scattering rigid. Flat balls (of any dimensions) are known (Gromov [10]) to be lens rigid.

1.3. Simple manifolds.

Definition 1.5. *A compact Riemannian manifold with boundary is* simple *if*

(1) *its boundary is strictly convex,*
(2) *there is a unique minimizing geodesic connecting any pair of points on the boundary,*
(3) *the manifold has no conjugate points.*

Remark 1.6. Note that simple manifolds are topological balls.

Conjecture 1.7 (Michel [14])**.** *Simple Riemannian manifolds are lens (boundary) rigid.*

Theorem 1.8 (Pestov–Uhlmann [15])**.** *Simple Riemannian surfaces are lens (boundary) rigid.*

Remark 1.9. A Riemannian manifold is *boundary rigid* if its metric is determined by the distance function between boundary points. The above statements are originally about boundary rigidity, which is equivalent to lens rigidity when the manifold is simple. Theorem 1.8 confirms the conjecture for surfaces. There are a variety of results in higher dimensions (Besson–Courtois–Gallot [2], Burago–Ivanov [3, 4], Croke–Kleiner [7], Michel [14]), but it is still largely open.

Our result extends Theorem 1.8 to scattering rigidity.

Theorem 1.10. *Simple Riemannian surfaces are scattering rigid.*

Remark 1.11. Simple Riemannian manifolds do not have trapped geodesics and trapped geodesics often make this type of rigidity problems much harder. Amazingly, the first (and the only one before this one) known result (Croke [6]) of scattering rigidity is for the flat product metric on $\mathbb{S} \times D^n$, which has trapped geodesics.

To get Theorem 1.10 from Theorem 1.8, it suffices to show that M and N have the same lens data if they have the same scattering data, assuming that M is simple. Note that this is not true in general without the assumption that M is simple. (See Figure 3.)
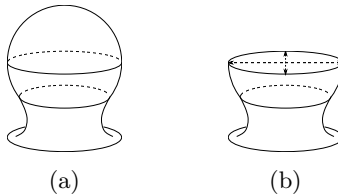

(a) (b)

FIGURE 3. 3b is obtained from 3a by removing the upper hemisphere and identifying antipodal points in the top boundary component. 3a and 3b have the same scattering data but different lens data.

By the first variation of arc length, $\ell(\gamma_{\varphi(X)}) - \ell(\gamma_X)$ is equal to a constant $L \geq 0$. If $L > 0$, then $\gamma_{\varphi(X)}$ converges to a closed geodesic of length L as X converges to a vector X_0 tangent to the boundary. (See Figure 4.) We will call this closed geodesic γ_{X_0}.

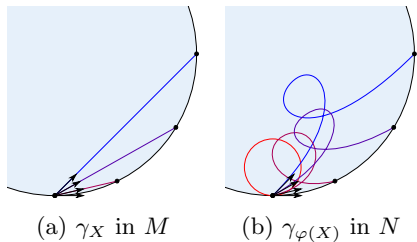(a) γ_X in M (b) $\gamma_{\varphi(X)}$ in N

FIGURE 4. Closed geodesics?

At first glance, this is very unlikely to happen, since one expects ∂N to be convex as ∂M is convex. However, the convexity of the boundary of a manifold, being a local property, is **not** determined by local scattering data as illustrated by the invisible Eaton lens. The boundary of the invisible Eaton lens is actually totally geodesic, and we have closed geodesics running along the boundary. The trickiest part of the proof is to get rid of these closed geodesics using knot theory, (which only works in dimension 2 so far).

1.4. **Scheme of the proof.** As explained in the previous section, we need to close the gap between lens rigidity and scattering rigidity, that is, to show $L = 0$. Recall that $L = \ell(\gamma_{\varphi(X)}) - \ell(\gamma_X)$, the difference between the lengths of corresponding geodesics in M and N, where M and N are two Riemannian manifolds with the same scattering data rel $h : \partial M \to \partial N$.

In section 2, we will prove that N is homeomorphic a disk.

Pick any $x \in \partial N$. If $L > 0$, then there is a closed geodesic γ_x of length L which is tangent to ∂N at x. There are two such closed geodesics for each x, but we can choose γ_x properly such that γ_x moves continuously as x moves. In this section, we will assume that γ_x has multiplicity 1. The actual proof will be more complicated due to the possibility of higher multiplicities, but the idea of the proof is the same.

The paper will study the isotopy type of the *projectivized unit tangent vector field* $P \circ \tilde{\gamma}_x : \mathbb{R}/\mathbb{Z} \to P\Omega N$ of γ_x where $\tilde{\gamma}_x : \mathbb{R}/\mathbb{Z} \to \Omega N$ is the unit tangent vector field of γ_x (see (3.1) in section 3), $P\Omega N = \Omega N/\{(x, \xi) \sim (x, -\xi)\}$ is the *projectivized unit tangent bundle* of N, and $P : \Omega N \to P\Omega N$ is the corresponding quotient map.

In section 3 we shall define a family of knot invariants for contractible knots embedded in $P\Omega N$, and then use those invariants to prove Theorem 1.12, which is interesting on its own.

Theorem 1.12. $P \circ \tilde{\gamma}$ *is an isotopically non-trivial knot in* $P\Omega N$ *for any smooth immersed curve* $\gamma : \mathbb{R}/\mathbb{Z} \to N$ *without self-tangencies.*

Remark 1.13. Theorem 1.12 is purely knot-theoretical as it involves nether scattering data nor lens data. It is a bit surprising that this simple fact was not known before even for plane curves. Actually, it would be a completely different story if the projectivization were dropped: Chmutov–Goryunov–Murakami [5] showed that every knot type in $\Omega\mathbb{R}^2$ (including the trivial type) is realized by the unit tangent vector field along an immersed plane curve.

Notice that the union of $P \circ \tilde{\gamma}_x$ for all $x \in \partial N$ is a torus immersed in $P\Omega N$. We can perturb the immersion to an embedding. Then we can prove that the torus is compressible by showing that $P \circ \tilde{\gamma}_x$ is contractible. (Actually, any embedded

torus in $P\Omega N$ is compressible.) Next, we can show that the other generator of the fundamental group of the torus is not contractible in $P \circ \tilde{\gamma}_x$. It follows that $P \circ \tilde{\gamma}_x$ bounds an embedded disk, which contradicts Theorem 1.12. Therefore, there is no such closed geodesics.

In the actual proof, we shall prove Theorem 1.10 in section 4 using a similar contradiction without the assumption on the multiplicity.

1.5. **Acknowledgements.** Many thanks to my advisor Christopher Croke for introducing me this subject and teaching me the techniques in this field. Many ideas in this paper stem from discussions with him.

2. TOPOLOGY OF N

Through out the paper (except in section 3), M and N will be two Riemannian surfaces with the same scattering data rel $h : \partial M \to \partial N$ where h is an isometry. Also, M is assumed to be simple. $\varphi : \partial \Omega M \to \partial \Omega N$ is the induced bundle map defined in (1.1). We aim to prove the following result in this section

Proposition 2.1. *N is homeomorphic to a 2-disk if M is simple.*

If $L := \ell(\gamma_{\varphi(X)}) - \ell(\gamma_X) = 0$, then M and N have the same lens data, and hence N is a 2-disk. Thus we shall assume that $L > 0$ in this section.

Pick a point $p_0 \in \partial N$, and let $\beta_1 : [0,1] \to \partial N$ be a constant speed closed curve of multiplicity 1, starting and ending at p_0. There are two such curves corresponding to different orientations but either one is fine.

Fix an orientation of ∂N and let $Y_0(x)$ be the unit vector tangent to ∂N at $x \in \partial N$ such that $Y_0(x)$ and ∂N have the same orientation. Define $\beta_x : [0,1] \to N$ as

$$\beta_x(t) = \gamma_{Y_0(x)}(Lt),$$

where $\gamma_{Y_0(x)}$ is the closed unit speed geodesic tangent to $Y_0(x)$ of length L. Write $\beta_2 = \beta_{p_0}$.

For any loop β in N based at $p \in N$, we will denote by $[\beta]_p$ the based homotopy class of β. Also, let $h : \pi_1(N,p) \to H_1(N,\mathbb{Z})$ be the abelianization map which sends based homotopy classes to corresponding homology classes. We will write $[\beta] := h([\beta]_p)$.

Proposition 2.2. $[\beta_1]_{p_0} = [\beta_2]_{p_0}^{-2}$.

Proof. We shall prove the equivalent statement

$$(2.1) \qquad\qquad [\beta_2]_{p_0} = [\beta_2]_{p_0}^{-1}[\beta_1]_{p_0}^{-1}.$$

Let $Y : [0,1]_{p_0} \to \partial_+ \Omega N$ be a smooth curve from $Y_0(x)$ to $-Y_0(x)$ such that $\gamma_{Y_t}(\tau(Y_t)) = \beta_1(t)$.

Define $H : [0,1] \times [0,1] \to M$ as

$$H_s(t) = \begin{cases} \gamma_{Y_s}(2\tau(Y_s)t) & \text{if } 0 \le t \le \frac{1}{2}, \\ \beta_1((2-2t)s) & \text{if } \frac{1}{2} \le t \le 1. \end{cases}$$

Then $[H_0]_{p_0} = [\beta_2]_{p_0}$ and $[H_1]_{p_0} = [\beta_2]_{p_0}^{-1}[\beta_1]_{p_0}^{-1}$, which implies (2.1). $\qquad\square$

Notice that β_x are all in the same homology class. Denote by g_C the homology class of β_x.

Assume that $g_C \neq 0$. Since N is a surface with boundary, it deformation retracts to a graph. (The deformation is quite simple. Take any cell structure on N. Remove a 1-cell on the boundary and a 2-cell by deformation retraction if they intersect. Repeat this process until all 2-cells are removed.) Hence $H_1(N, \mathbb{Z}) = \mathbb{Z}^n$ for some $n \in \mathbb{N}$. So $g_C = mg_0$ for some $m > 0$ and g_0 prime. Then the multiplicity of β_x is at most m since it must divide m. Let m_0 be the maximal multiplicity of β_x.

Proposition 2.3. *If $g_C \neq 0$, then $H(N, \mathbb{Z})$ is generated by g_0.*

Proof. For any $g \in \pi_1(N, p_0)$, let $\gamma_g : [0,1] \to N$ be the length minimizing representative of g that is of constant speed T_g. Since $A := \gamma_g^{-1}(N \setminus \partial N)$ is open, $A = \bigcup \mathcal{A}$ where $\mathcal{A}$ is a family of disjoint open intervals. For any $(a, b) \in \mathcal{A}$, since γ_g is length minimizing, $\gamma_g|_{[a,b]}$ has to be a geodesic segment. If $a \neq 0$, then $\gamma_g'(a)$ has to be tangent to ∂N, or γ_g will have a corner at $\gamma_g(a)$, contradicting the assumption that γ_g is length minimizing. According to the scattering data, $\gamma_g'(b)$ is also tangent to ∂N if $\gamma_g'(a)$ is tangent to ∂N. Hence $\gamma_g|_{[a,b]}$ is a closed geodesic tangent to ∂N when $a \neq 0$. Similarly, $\gamma_g|_{[a,b]}$ is closed geodesic tangent to ∂N when $b \neq 1$. Suppose that $a = 0$ and $b = 1$. If $\gamma_g'(0)$ is not tangent to ∂N, then $\gamma_g'(0)/|\gamma_g'(0)| = \varphi(X) \in \partial_+ \Omega N$ for some $X \in \partial_+ \Omega M$, and we have $\gamma_g'(1)/|\gamma_g'(1)| = \alpha_M(X)$. Since M is a simple manifold and $X \in \partial_+ \Omega M$, γ_X is a length minimizing geodesic, and thus X and $\alpha_M(X)$ have different base points. It follows that $\gamma_g'(0)$ and $\gamma_g'(1)$ also have different base points, contradicting our assumption that γ_g is a loop. Therefore, in any case, $\gamma_g|_{[a,b]}$ is a closed geodesic tangent to ∂N, and thus of length at least $\frac{L}{m_0}$. It follows that $|\mathcal{A}| \leq m_0 T_g / L < \infty$. So we can write $\mathcal{A} = \{(a_1, b_1), (a_2, b_2), \ldots, (a_{n_g}, b_{n_g})\}$ where $n_g = |\mathcal{A}|$ and $0 \leq a_1 < b_1 \leq a_2 < b_2 \leq \cdots \leq a_{n_g} < b_{n_g} \leq 1$.

Since $\gamma_g|_{[a_i, b_i]}$ is a closed geodesic, $[\gamma_g|_{[a_i, b_i]}] = g_0^{k_i}$ for some $k_i > 0$. Deleting all those closed geodesics from γ_g, we obtain a curve running around ∂N l times for some $l \in \mathbb{Z}$. Its homology class will be $[\beta_1]^{\pm l} = [\beta_2]^{\pm 2l} = g_0^{\pm 2ml}$. Therefore, $h(g) = g_0^{\pm 2m_0 l + \sum_{i=1}^{n_g} k_i}$. Since h is surjective, $H_1(N, \mathbb{Z})$ is generated by g_0. $\qquad\square$

Proposition 2.4. *N is not a Möbius strip.*

Proof. Let $\pi : N_1 \to N$ be a double over of N. Then N_1 is an annulus with two boundary components S_1 and S_2. There are $p \in S_1$ and $q \in S_2$ such that $d(p, q) = d(S_1, S_2)$. Let γ be the shortest curve from p to q, then γ is perpendicular to S_1 and S_2 at tis end points. Let ν be the unit normal vector at p. Since γ is the shortest curve from p to q, its beginning part must coincide with γ_ν. If the end point $\gamma_\nu(\tau_{N_1}(\nu))$ is on S_1, we can shorten γ by deleting γ_ν. If the end point $\gamma_\nu(\tau_{N_1}(\nu))$ is on S_2, then γ can not be any longer. Thus $\ell(\gamma) = \tau_{N_1}(\nu)$

Notice that, for any $X \in \partial_+ \Omega N_1$, $\pi \circ \gamma_X = \gamma_{\pi_*(X)}$. Let Y_t be a smooth curve in $\Omega_p N_1$ such that $Y_0 = \nu$, that $Y_t \in \partial_+ \Omega N_1$ for $t \in [0, 1)$ and that Y_1 is tangent to ∂S_1. Notice that the end point of $\pi \circ \gamma_{Y_t} = \gamma_{\pi_*(Y_t)}$ moves continuously (since N has the same scattering data as the simple surface M), and hence the end point of γ_{Y_t} moves continuously for $t \in [0, 1)$. Therefore, γ_{Y_t} connects p and S_2 for any $t \in [0, 1)$. However, we have $\ell(\gamma) = \tau_{N_1}(\nu) = \tau_N(\pi_*(\nu)) > L$, and $\lim_{t \to 1} \ell(\gamma_{Y_t}) = \lim_{t \to 1} \tau_N(\pi_*(Y_t)) = L$, which contradicts our assumption that γ is a shortest curve connecting S_1 and S_2. $\qquad\square$

Proof of Proposition 2.1. If $g_C \neq 0$, then $H_1(N, \mathbb{Z})$ is generated by g_0 by Proposition 2.3. Hence $H(N, \mathbb{Z}) = \mathbb{Z}$, which implies that N is a Möbius strip, which contradicts Proposition 2.4.

Therefore, $g_C = 0$. It follows that β_2 is contractible, and hence β_1 is contractible by Proposition 2.2. Since every contractible simple closed curve on a surface bounds a disk [9, Theorem 1.7], N is a disk. □

3. KNOT THEORY

In this chapter, N will denote a Riemannian surface, with or without boundary, orientable or not. We assume that there is a Riemannian metric on N just for convenience and all the results can be stated with only a smooth structure.

3.1. **Projectivized unit tangent vector fields.**

Definition 3.1. *The* unit tangent vector field *of a smoothly immersed curve γ on any Riemannian surface N^2 (possibly with boundary) is a smoothly immersed curve $\tilde{\gamma}$ in ΩN defined as*

$$(3.1) \qquad \tilde{\gamma}(t) = \left(\gamma(t), \frac{\gamma'(t)}{|\gamma'(t)|} \right).$$

Definition 3.2. *Let $P : \Omega N \to P\Omega N$ be the quotient map on the unit tangent bundle which identifies the opposite vectors based at the same point. For any smoothly immersed curve γ in N^2, $P \circ \tilde{\gamma}$ is called the* projectivized unit tangent vector field *(or the* tangent line field*) of γ.*

Remark 3.3. Chmutov–Goryunov–Murakami [5] showed that every knot type in $\Omega\mathbb{R}^2$ is realized by the unit tangent vector field along an immersed closed plane curve. However, Theorem 1.12 shows that it is no longer possible to realize the trivial knot after the projectivization. Figure 5 is an interesting example showing that the unit tangent vector field of the figure eight curve is an unknot while the projectivized unit tangent vector field of the figure eight curve is knotted.

(a) $\gamma : \mathbb{R}/\mathbb{Z} \to \mathbb{R}^2$ (b) $\tilde{\gamma} : \mathbb{R}/\mathbb{Z} \to \Omega\mathbb{R}^2$ (c) $P \circ \tilde{\gamma} : \mathbb{R}/\mathbb{Z} \to P\Omega\mathbb{R}^2$
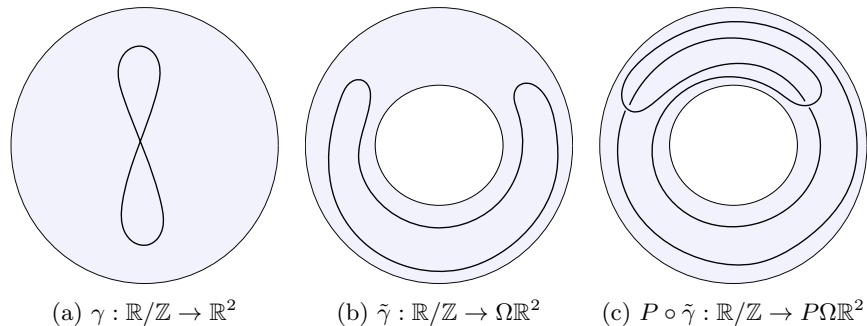
FIGURE 5. The unit tangent vector field of the figure eight curve is unknotted while the projectivized unit tangent vector field of the figure eight curve is knotted. Here the solid tori ($\Omega\mathbb{R}^2$ and $P\Omega\mathbb{R}^2$) are projected to annuli for illustration.

Proposition 3.4. *For any smoothly **embedded** closed curve $\gamma : \mathbb{R}/\mathbb{Z} \to N$ in a 2-dimensional manifold N, $\tilde{\gamma}$ is not contractible in ΩN and hence $P \circ \tilde{\gamma}$ is not contractible in $P\Omega N$.*

Proof. If γ is contractible, then γ bounds an embedded disk N_1 in N [9, Theorem 1.7].

Let $x = \tilde{\gamma}(0)$, $p = \gamma(0)$ and $F = \pi^{-1}(p)$. Denote by $[\tilde{\gamma}]_x$ the based homotopy class of $\tilde{\gamma}$. As in Figure 6, $\tilde{\gamma}$ corresponds to a vector moving along γ for a complete

(a) $\tilde{\gamma}$ (b) Moving base points towards p (c) A generator of $\pi_1(F, x)$
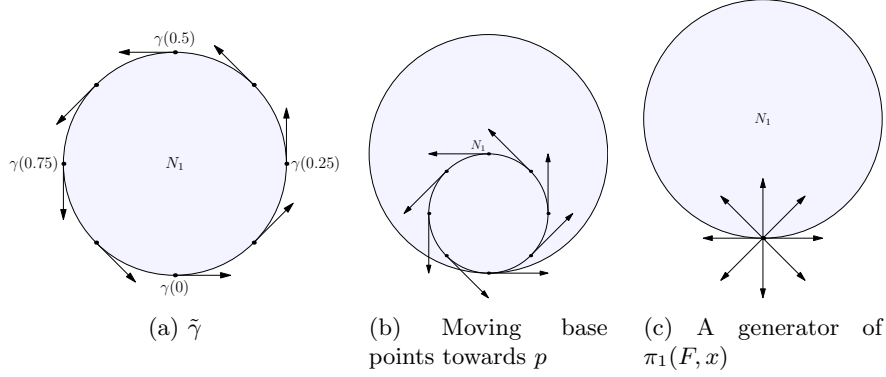
FIGURE 6. $\tilde{\gamma}$ is homotopic to a generator of $\pi_1(F, x)$

circle and being tangent to γ all the time, which is homotopic to a generator of $\pi_1(F, x)$. Denote the generator of $\pi_1(F, x)$ by g_1.

Since

$$ F \xrightarrow{\;\;i\;\;} \Omega N \xrightarrow{\;\;\pi\;\;} N $$

is a fibration, we have an exact sequence of homotopy groups

$$ \pi_2(N, p) \longrightarrow \pi_1(F, x) \xrightarrow{\;\;i_*\;\;} \pi_1(\Omega N, x) \xrightarrow{\;\;\pi_*\;\;} \pi_1(N, p). $$

Here $\pi_1(F, x) = \mathbb{Z}$ since F is a circle.

If $N = \mathbb{S}^2$, then $\pi_1(N, p) = 0$ and $\pi_1(\Omega N, p) = \pi_1(\mathbb{R}P^3, *) = \mathbb{Z}/2\mathbb{Z}$. Hence $i_*(\pi_1(F, x)) = \mathbb{Z}/2\mathbb{Z}$. In particular, $i_*(g_1) \neq 0$. A similar argument shows that $i_*(e) \neq 0$ when $N = \mathbb{R}P^2$.

If $N \neq \mathbb{S}^2$ and $N \neq \mathbb{R}P^2$, then $\pi_2(N, p) = 0$. Hence i_* is injective. In particular, $i_*(g_1) \neq 0$.

This completes the proof of Proposition 3.4 $\square$

3.2. **Knot invariants.** We shall define a family of knot invariants for contractible knots in the projectivized unit tangent bundle $P\Omega N$ and use these invariants to prove Theorem 1.12.

Let $\beta : \mathbb{R}/\mathbb{Z} \to P\Omega N$ be a contractible smooth knot in the projectivized unit tangent bundle $P\Omega N$, whose projection to the surface N^2 is a smoothly immersed curve $\gamma : \mathbb{R}/\mathbb{Z} \to N^2$ without self-tangencies.

Definition 3.5. β *has a* crossing *at* $(l, l') \in \mathbb{R}/\mathbb{Z} \times \mathbb{R}/\mathbb{Z}$ *if* $l \neq l'$ *and* $\gamma(l) = \gamma(l')$. *Note that a triple crossing will be treated as three independent crossings according to this definition.*

Since $\beta : \mathbb{R}/\mathbb{Z} \to P\Omega N$ is contractible, we can lift β to $\hat{\beta} : \mathbb{R}/\mathbb{Z} \to \Omega N$, a knot embedded in the unit tangent bundle. ($\hat{\beta}(t)$ is a unit vector at $\gamma(t)$ but not necessarily tangent to γ.)

Definition 3.6. *A crossing of β at (l, l') is* positive *if the two pairs of vectors $(\hat{\beta}(l), \hat{\beta}(l'))$ and $(\gamma'(l), \gamma'(l'))$ are of the same orientation. (See Figure 7.) A crossing will be called* negative *if it is not positive.*
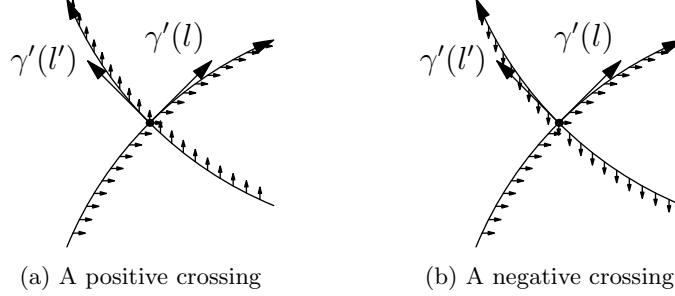

(a) A positive crossing   (b) A negative crossing

FIGURE 7. Each little arrow means a point on $\hat{\beta}$

.

Lemma 3.7. *Suppose that $\gamma : \mathbb{R}/\mathbb{Z} \to N$ is a smoothly immersed closed curve on a surface N without self-tangencies, then all crossings of $P \circ \tilde{\gamma}(t)$ are positive.*

Proof. Suppose that $P \circ \tilde{\gamma}$ has a crossing at (l, l'). Write $\beta = P \circ \tilde{\gamma}$. Then $(\hat{\beta}(l), \hat{\beta}(l')) = (\gamma'(l), \gamma'(l'))$, and hence they have the same orientation. Therefore the crossing at (l, l') is positive. $\qquad\square$

Let X be any topological space. For any two curves $\alpha_1 : [0, 1] \to X$ and $\alpha_2 : [0, 1] \to X$ such that $\alpha_1(1) = \alpha_2(0)$, denote by $\alpha_1 * \alpha_2 : [0, 1] \to X$ the curve obtained by gluing α_2 to α_1. Also, define $R(\alpha_1)$ as $R(\alpha_1)(t) := \alpha_1(1 - t)$. If α_1 and α_2 are loops based at p, we have $[\alpha_1]_p [\alpha_2]_p = [\alpha_1 * \alpha_2]_p$ and $[R(\alpha_1)]_p = [\alpha_1]_p^{-1}$.

Definition 3.8. *Two closed curves $\alpha_1 : \mathbb{R}/\mathbb{Z} \to X$ and $\alpha_2 : \mathbb{R}/\mathbb{Z} \to X$ in any topological space X are said to be in the same* unoriented free homotopy class *if γ_1 is homotopic to either γ_2 or $R(\gamma_2)$.*

When β has a crossing at (l, l'), $\hat{\beta}(l)$ and $\hat{\beta}(l')$ are two unit vectors with the same base point $x = \pi(\beta(l))$ and they are neither opposite to each other nor the same (since β is an embedding). Hence there is a unique shortest curve $\hat{\beta}_{(l,l')}$ in $\pi^{-1}(x)$ connecting $\hat{\beta}(l)$ and $\hat{\beta}(l')$. Separate $\hat{\beta}$ into two arcs by cutting at $\hat{\beta}(l)$ and $\hat{\beta}(l')$, obtaining two arcs $\hat{\beta}_1 : [0, 1] \to P\Omega N$ and $\hat{\beta}_2 : [0, 1] \to P\Omega N$ going from $\hat{\beta}(l)$ to $\hat{\beta}(l')$.

Now, let $\beta_1' = (P \circ \hat{\beta}_1) * R(P \circ \hat{\beta}_{(l,l')})$, and $\beta_2' = (P \circ \hat{\beta}_{(l,l')}) * R(P \circ \hat{\beta}_2)$. Notice that $\beta_1' * R(\beta_2')$ is homotopic to β, and hence $[\beta_1']_p [R(\beta_2')]_p = [\beta_1' * R(\beta_2')]_p = [\beta]_p = e$. Hence $[\beta_1']_p = [R(\beta_2')]_p^{-1} = [\beta_2']_p$. In other words, β_1' is homotopic to β_2', and hence β_1' and β_2' are in the same unoriented free homotopy class of $P\Omega N$.

Definition 3.9. *The unoriented free homotopy class $g_{(l,l')}$ of β_1' is called the* type *of the crossing of β at (l, l').*
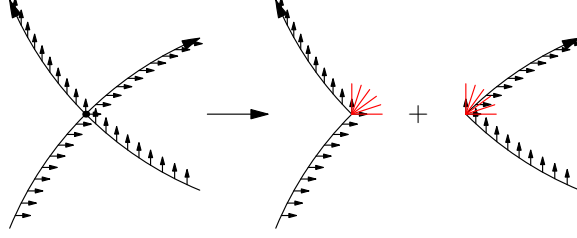
FIGURE 8. Smoothing a crossing. Here each little arrow means a point on $\hat{\beta}$ and each little bar means a point on $\hat{\beta}_{(l,l')}$.

Definition 3.10. *For each **nontrivial** unoriented free homotopy class g of closed curves in the projectivized unit tangent bundle $P\Omega N$, define*

$$W_g(\beta) = \#\{\text{positive crossings of } \beta \text{ of type } g\}$$
$$- \#\{\text{negative crossings of } \beta \text{ of type } g\}$$

3.3. W_g is a knot invariant.

Theorem 3.11. *For each non-trivial free homotopy class g, W_g can be extended to all the contractible knots embedded in $P\Omega N$ as a knot invariant.*

We will show that W_g is a knot invariant by verifying that W_g is unchanged under Reidemeister moves. A knot will gain or lose a crossing of trivial type after going through a Reidemeister move of type I. It will gain or lose a pair of crossing of the same type but opposite signs after going through a Reidemeister move of type II. Reidemeister moves of type III will not affect crossing. The proof is rather lengthy because of some technical difficulties.

We will assume that N is compact, and the general case follows automatically since any manifold is σ-compact.

Definition 3.12. *According to [1, Theorem 5], there is $r > 0$ such that there is a unique minimal geodesic segment joining $p, q \in N$ if $d(p,q) < r$. The biggest such r will be called the* injectivity radius *of N and we will denote it by $inj(N)$.*

For any two points $p, q \in P\Omega N$, let $d_h(p,q)$ be the distance between $\pi(p)$ and $\pi(q)$ on N. (So d_h is a pseudo metric on $P\Omega N$.) Notice that p is a projectivized unit tangent vector at $\pi(p)$. When $d_h(p,q) < inj(N)$, there is a unique shortest geodesic $\gamma : [0,1] \to N$ in N connecting $\pi(p)$ and $\pi(q)$. Let $X : [0,1] \to P\Omega N$ be the parallel projectivized vector field along γ such that $X(0) = p$. Similarly, let $Y : [0,1] \to P\Omega N$ be the parallel projectivized vector field along γ such that $Y(1) = q$. Notice that the angle between X and Y is constant, which is smaller or equal to $\frac{\pi}{2}$. Call this angle $d_v(p,q)$. Next, put $d^0(p,q) = \max(d_h(p,q), d_v(p,q))$. Note that d_h, d_v and $d^0(p,q)$ are all non-negative and symmetric, but they are not metrics.

Definition 3.13. *For any $p, q \in P\Omega N$ such that $d_h(p,q) < inj(N)$ and that $d_v(p,q) < \frac{\pi}{2}$, let $\gamma^0_{p,q} : [0,1] \to \Omega N$ be the curve that satisfies the following conditions.*

(1) $\gamma^0_{p,q}(0) = p$ and $\gamma^0_{p,q}(1) = q$.
(2) $\pi \circ \gamma$ is the minimal geodesic connecting $\pi(p)$ and $\pi(q)$.

(3)

(3.2)
$$\left| \frac{D}{dt} \gamma_{p,q}^0 \right| = d_v(p,q).$$

$\gamma_{p,q}^0$ *will be called the* minimal linear curve *connecting p and q. A curve will be called* linear *if it coincides with $\gamma_{p,q}^0$ for any pair of points p, q on the curve that are close enough. A curve will be called* piecewise linear *if it consists of finitely many linear curves.*

Proof of Theorem 3.11. For any $\varepsilon < \min(inj(N), \frac{\pi}{2})$ and $n \geq 4$, we will define a class of closed piecewise linear knots in $P\Omega N$ called $\mathcal{K}(n, \varepsilon)$. A closed knot $\beta : \mathbb{R}/\mathbb{Z} \to P\Omega N$ is in $\mathcal{K}(n, \varepsilon)$ if and only if the following condition holds:

(1) β is contractible.
(2) $d^0(\beta(\frac{k}{n}), \beta(\frac{k+1}{n})) < \varepsilon$ for $k = 0, 1, \ldots, n-1$.
(3) $\beta(\frac{k+t}{n}) = \gamma_{\beta(\frac{k}{n}), \beta(\frac{k+1}{n})}^0(t)$ for $t \in [0, 1]$ and $k = 0, 1, \ldots, n-1$.

In other words, the "distance" (d_h and d_v) between any two adjacent vertices p and q is at most ε and the edge between them is $\gamma_{p,q}^0$. $\mathcal{K}(n, \varepsilon)$ is an open subset of $(P\Omega N)^n$, and thus of dimension $3n$.

Let $\beta \in \mathcal{K}(n, \varepsilon)$ be a piecewise smooth **contractible** knot with vertices $\{x_k = \beta(\frac{k}{n})\}$ and edges $\{e_k = \gamma_{x_k, x_{k+1}}^0\}$. Its projection $\pi \circ \beta$ is said to have a *singularity* at the vertex $\pi(x_i)$ if $\pi(x_i)$ is on $\pi \circ e_j$ for some $j \notin \{i, i-1\}$.

Let $\mathcal{K}_k(n, \varepsilon)$ be the set of knots in $\mathcal{K}(n, \varepsilon)$ whose projections on N have at most k singularities. Also, let $\mathcal{K}'_k(n, \varepsilon) = \mathcal{K}_k(n, \varepsilon) - \mathcal{K}_{k-1}(n, \varepsilon)$, knots with exactly k singularities. Then $\mathcal{K}_0(n, \varepsilon)$ is a open submanifold of $\mathcal{K}(n, \varepsilon)$, and $\mathcal{K}'_1(n, \varepsilon)$ is a submanifold of $\mathcal{K}(n, \varepsilon)$ of co-dimension 1.

Notice that the $W_g(\beta)$ can be defined for $\beta \in \mathcal{K}_0(n, \varepsilon)$ as before without any modifications. Consider a continuous family of knots $\beta_t \in \mathcal{K}_0(n, \varepsilon)$. As t varies, crossings of β_t also moves continuously with their types unchanged. Therefore, W_g is constant on each component of $\mathcal{K}_0(n, \varepsilon)$.

Next, we extend W_g to $\mathcal{K}_1(n, \varepsilon)$. The old definition can not be adapted directly since there might singularities. Pick any $\beta_0, \beta_1 \in \mathcal{K}_0(n, \varepsilon)$ such that β_0, β_1 are in the same component of $\mathcal{K}_1(n, \varepsilon)$. We aim to show that $W_g(\beta_0) = W_g(\beta_1)$, and then we can extend W_g to $\mathcal{K}_1(n, \varepsilon)$ by making it constant on each component. Note that W_g will remain the same on $\mathcal{K}_0(n, \varepsilon)$.

Pick a smooth path $H : [0, 1] \to \mathcal{K}_1(n, \varepsilon)$ from β_0 to β_1. Perturbing H if necessary, we may assume that H intersects $\mathcal{K}'_1(n, \varepsilon)$ transversely a finite number of times. Let $x_0(t), x_1(t), \ldots, x_n(t) = x_0(t)$ be the n vertices of $H(t)$.

As long as $H(t)$ stays in $\mathcal{K}_0(n, \varepsilon)$, each crossing will just be moving without changing its type. When $H(t)$ passes through $\mathcal{K}'_1(n, \varepsilon)$, there are three possibilities corresponding to three types of singularities for knots in $\mathcal{K}'_1(n, \varepsilon)$ listed below. Suppose $H(c) \in \mathcal{K}'_1(n, \varepsilon)$ and $H(t) \notin \mathcal{K}'_1(n, \varepsilon)$ for $t \in (c - \delta, c) \bigcup (c, c + \delta)$. Then $\pi \circ H(c)$ has a singularity at $\pi(x_i(c))$ which is on $\pi \circ e_j$ where $x_i(c)$ is the i-th vertex of $H(c)$ and $e_j(c)$ is j-th edge of $H(c)$ (connecting $x_j(c)$ and $x_{j+1}(c)$).

(1) If $\pi \circ e_i(c)$ or $\pi \circ e_{i-1}(c)$ is tangent to $\pi \circ e_j(c)$, then the singularity is called a *cusp*. This happens when $i = j - 1$ or $i = j + 2$. In this case, $H(c+\delta)$ has one more or one less crossing than $H(c-\delta)$ has. We will show that the crossing involved is of the trivial type, (i.e., $g = 0$,) and hence

$W_g(H(c - \delta)) = W_g(H(c + \delta))$ for any non-trivial unoriented homotopy class g of closed curves immersed in $P\Omega N$.
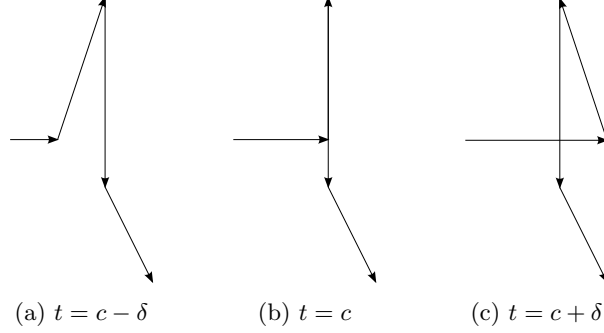

(a) $t = c - \delta$ (b) $t = c$ (c) $t = c + \delta$

FIGURE 9. 9c has one more crossing of the trivial type compared to 9a. This corresponds to a Reidemeister move of type I.

Without loss of of generality, assume that $i = j - 1$ and that $H(t)$ has one more crossing at $(l(t), l'(t))$ than $H(t')$ has when $c - \delta \le t' < c < t \le c + \delta$. (See Figure 9.) For any $t \in (c, c + \delta]$, swapping $l(t)$ and $l'(t)$ if necessary, we may assume that $l(t) \in (\frac{i-1}{n}, \frac{i}{n})$ and $l'(t) \in (\frac{i+1}{n}, \frac{i+2}{n})$. Lift $H : [0,1] \to \mathcal{K}(n, \varepsilon)$ to $\hat{H} : [0,1] \to (\mathbb{R}/\mathbb{Z} \to \Omega N)$. Since $H(t)$ is an embedding, $H(t)(l(t)) \ne H(t)(l'(t))$, and hence $\hat{H}(t)(l(t))$ and $\hat{H}(t)(l'(t))$ are not opposite vectors. It follows that there is a unique minimal geodesic $\hat{\alpha}(t) : [0,1] \to \pi^{-1}(x)$ connecting $\hat{H}(t)(l(t))$ and $\hat{H}(t)(l'(t))$. Let $\alpha(t) = P \circ \hat{\alpha}(t)$ and glue $\alpha(t)$ to $H(t)|_{[l(t), l'(t)]}$, obtaining a closed curve $C(t)$. Then the type of the crossing of $H(t)$ at $(l(t), l'(t))$ is the unoriented homotopy class of $C(t)$. It remains to show that $C(t)$ is contractible.

Let $l(c) = \lim_{t \to c+} l(t)$ and $l'(c) = \lim_{t \to c+} l'(t)$, then $C(c)$ can be defined as before, which is homotopic to $C(t)$ for $t \in (c, c + \delta)$. We shall show that $C(c)$ is contractible.

Reparametrize $C(c)$ as $\bar{\beta} : \mathbb{R}/\mathbb{Z} \to P\Omega N$ such that $\bar{\beta}(0) = H(c)(l(c))$, $\bar{\beta}(\frac{1}{3}) = x_{i+1}(c)$, $\bar{\beta}(\frac{2}{3}) = H(c)(l'(c))$ and $\pi(\bar{\beta}(\frac{1}{3}(1 + s))) = \pi(\bar{\beta}(\frac{1}{3}(1 - s)))$ for any $s \in [0, 1]$. To be precise, define $\bar{\beta}$ as

$$\bar{\beta}(t) = \begin{cases} H(c)(\frac{i+3t}{n}) & \text{if } t \in [0, \frac{1}{3}], \\ H(c)(\frac{i+1}{n} + (3t - 1)(l'(c) - \frac{i+1}{n})) & \text{if } t \in [\frac{1}{3}, \frac{2}{3}], \\ \alpha(3 - 3t) & \text{if } t \in [\frac{2}{3}, 1]. \end{cases}$$

Consider the homotopy $G : [0, 1] \to (\mathbb{R}/\mathbb{Z} \to P\Omega N)$ defined as

$$G(s)(t) = \begin{cases} \bar{\beta}(t) & \text{if } 0 \le t \le \frac{1}{3}(1 - s), \\ T(\bar{\beta}(t), \pi(\bar{\beta}(\frac{1}{3}(1 - s)))) & \text{if } \frac{1}{3}(1 - s) \le t \le \frac{1}{3}(1 + s), \\ \bar{\beta}(t) & \text{if } \frac{1}{3}(1 + s) \le t \le 1, \end{cases}$$

where $T(\bar{\beta}(t), \pi(\bar{\beta}(\frac{1}{3}(1 - s))))$ is a projectivized unit tangent vector at $\pi(\bar{\beta}(\frac{1}{3}(1 - s)))$ obtained by transporting $\bar{\beta}(t)$ parallelly along $\pi \circ e_j(c)$. Notice that $G(1)$ is a closed curve in $\Omega_{\pi(x_i)}N$, where $\Omega_{\pi(x_i)}N$ is a circle of length 2π, (using the Sasakian metric). We are going to show that $G(1)$

is contractible by showing that $\ell(G(1)) < 2\pi$. For any piecewise smooth curve $\gamma : [a, b] \to P\Omega N$, define its *vertical length* as

$$\ell_v(\gamma) := \int_a^b \left| \frac{D}{dt} \gamma(t) \right| dt,$$

where $\frac{D}{dt}$ is the covariant derivative. Loosely speaking, $\ell_v(\gamma)$ measure the angle that $\gamma(t)$ rotates by as t goes from a to b. By our construction, $\ell_v(G(s))$ is constant as s goes from 0 to 1. Notice that $\bar{\beta}$ has three edges. The edge from $\bar{\beta}(0)$ to $\bar{\beta}(\frac{1}{3})$ and the edge from $\bar{\beta}(\frac{1}{3})$ to $\bar{\beta}(\frac{2}{3})$ both have vertical lengths at most ε (by (3.2)), and the vertical length of the edge from $\bar{\beta}(\frac{2}{3})$ to $\bar{\beta}(0)$ (which is reparametrized α) is at most π. Since $\varepsilon < \frac{\pi}{2}$, $\ell_v(\bar{\beta}) < \pi + 2\varepsilon < 2\pi$, and hence $\ell(G(1)) = \ell_v(G(1)) = \ell_v(G(0)) = \ell_v(\bar{\beta}) < 2\pi$. It follows that $G(1)$ is contractible, and hence $C(t)$ is contractible for any $t \in [c, c + \delta]$.

(2) If $\pi \circ e_i(c)$ and $\pi \circ e_{i-1}(c)$ are not tangent to $\pi \circ e_j(c)$, and if $\pi \circ e_i(c)$ and $\pi \circ e_{i-1}(c)$ are on the same side of $\pi \circ e_j(c)$, then the singularity is called a *self-tangency*. In this case, $H(c + \delta)$ has two more or two less crossings than $H(c - \delta)$ has. We can show that the two crossings involved are of the same type g but opposite signs, and hence $W_g(H(c + \delta)) = W_g(H(c - \delta))$.

Without loss of of generality, assume that $H(t)$ has two more crossing at $(l_1(t), l_1'(t))$ and $(l_2(t), l_2'(t))$ than $H(t')$ has when $c - \delta \leq t' < c < t \leq c + \delta$. (See Figure 10.) Let $l_1(c) = \lim_{t \to c+} l_1(t)$ and define $l_1'(c)$, $l_2(c)$ and $l_2'(c)$

(a) $t = c - \delta$ (b) $t = c$ (c) $t = c + \delta$
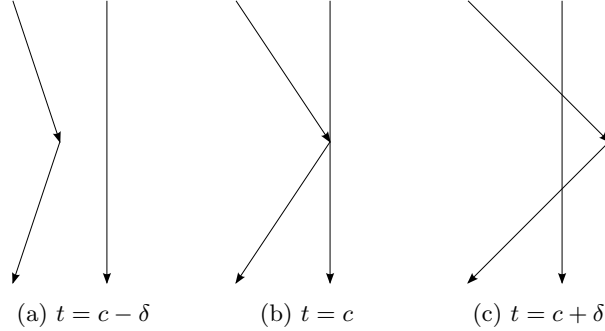
FIGURE 10. 10c has two more crossing of the same type but opposite signs compared to 10a. This corresponds to a Reidemeister move of type II.

similarly. Switching l_2 and l_2' if necessary, we may assume that $l_1(c) = l_2(c)$ and $l_1'(c) = l_2'(c)$. Also, either $H(l_1(c)) = x_i$ or $H(l_1'(c)) = x_i$. Without loss of generality, we assume that $H(l_1(c)) = x_i$. Lift $H : [0, 1] \to \mathcal{K}(n, \varepsilon)$ to $\hat{H} : [0, 1] \to (\mathbb{R}/\mathbb{Z} \to \Omega N)$, and denote the vertices of $\hat{H}(t)$ by $\hat{x}_k(t)$ and edges by $\hat{e}_k(t)$.

For any $t \in (c, c + \delta]$, we can separate $\hat{H}(t)$ into two arcs by cutting at $\hat{H}(l_1(t))$ and $\hat{H}(l_1'(t))$. Pick the arc which contains $\hat{x}_i(t)$ and glue it to $\hat{H}_{(l_1(t), l_1'(t))}$, obtaining a closed curve $C_1(t)$. We can also separate $\hat{H}(t)$ into two arcs by cutting at $\hat{H}(l_2(t))$ and $\hat{H}(l_2'(t))$. Pick the arc which does not contain $i\hat{x}_i(t)$ and glue it to $\hat{H}_{(l_2(t), l_2'(t))}$, obtaining a closed curve

$C_2(t)$. Next, define $C_1(c)$ and $C_2(c)$ by taking limits. It is then clear that $P \circ C_1(t)$ and $P \circ C_2(t)$ are in the same unoriented homotopy class since $C_1(c) = C_2(c)$. Hence the two crossings at $(l_1(t), l_1'(t))$ and $(l_2(t), l_2'(t))$ are of the same type.

Finally, it remains to show that the two crossings have opposite signs. Without loss of generality, assume that the crossing at $(l_1(t), l_1'(t))$ is positive. In other words, $(\hat{H}(l_1(t)), \hat{H}(l_1'(t)))$ and $((\pi \circ H)'(l_1(t)), (\pi \circ H)'(l_1'(t)))$ have the same orientation. It follows that $(\hat{H}(l_1(c)), \hat{H}(l_1'(c)))$ and $(\lim_{t \to c+}(\pi \circ H)'(l_1(t)), (\pi \circ H)'(l_1'(c)))$ have the same orientation. Since $\pi \circ e_i(c)$ and $\pi \circ e_{i-1}(c)$ are on the same side of $\pi \circ e_j(c)$, $(\lim_{t \to c+}(\pi \circ H)'(l_1(t)), (\pi \circ H)'(l_1'(c)))$ and $(\lim_{t \to c+}(\pi \circ H)'(l_2(t)), (\pi \circ H)'(l_2'(c)))$ have the opposite orientation. Since $(\hat{H}(l_1(c)), \hat{H}(l_1'(c)))$ and $(\hat{H}(l_2(c)), \hat{H}(l_2'(c)))$ are the same, $(\hat{H}(l_2(c)), \hat{H}(l_2'(c)))$ and $(\lim_{t \to c+}(\pi \circ H)'(l_2(t)), (\pi \circ H)'(l_2'(c)))$ have the opposite orientation, and thus the crossing at $(l_2(t), l_2'(t))$ is negative.

(3) If $\pi \circ e_i$ and $\pi \circ e_{i-1}$ are not tangent to $\pi \circ e_j$, and if $\pi \circ e_i$ and $\pi \circ e_{i-1}$ are on different sides of $\pi \circ e_j$, then the singularity is called a *transverse self-intersection*. In this case, all crossings moves continuously as t goes from $c - \delta$ to $c + \delta$, although one crossing will be also a singularity at $t = c$. The type and the sign of that crossing will be unchanged, which follows from an argument very similar to the one used for the previous case.

In any case, we have $W_g(H(c - \delta)) = W_g(H(c + \delta))$ for any non-trivial type g. It follows that $W_g(\beta_0) = W_g(H(0)) = W_g(H(1)) = W_g(\beta_1)$, and thus we may extend W_g to $\mathcal{K}_1(n, \varepsilon)$ by making it constant on each component.

Next, we will extent W_g to the whole $\mathcal{K}(n, \varepsilon)$. Pick any $\beta_0, \beta_1 \in \mathcal{K}_1(n, \varepsilon)$ such that β_0, β_1 are in the same component of $\mathcal{K}(n, \varepsilon)$. We aim to show that $W_g(\beta_0) = W_g(\beta_1)$, and then we can extend W_g to $\mathcal{K}(n, \varepsilon)$ by making it constant on each component.

The manifold $\mathcal{K}(n, \varepsilon)$ has a natural stratified structure as follows. For any $\gamma \in \mathcal{K}(n, \varepsilon)$, pick any neighborhood U of γ. If γ has k singularities, then let U_γ be the component of $U \bigcap \mathcal{K}_k'(n, \varepsilon)$ containing γ, which is a submanifold embedded in $\mathcal{K}(n, \varepsilon)$. Now, let $X_m = \{\gamma \in \mathcal{K}(n, \varepsilon) : \dim(U_\gamma) = m\}$. Then $\mathcal{K}(n, \varepsilon)$ is a stratified space whose m-dimensional stratum is X_m. We obviously have $X_{3n} = \mathcal{K}_0(n, \varepsilon)$ and $X_{3n-1} = \mathcal{K}_1'(n, \varepsilon)$. (Note that $X_m = \mathcal{K}_{3n-m}'(n, \varepsilon)$ is not true when m is big since the singularities are not necessarily independent. See Figure 11.) Pick a smooth path $H : [0, 1] \to \mathcal{K}_1(n, \varepsilon)$ from β_0 to β_1. Perturbing H if necessary, we may assume that H intersects each stratum X_m transversely. In other words, H does not intersect X_m at all if $m < 3n - 1$. Hence H is actually a path in $\mathcal{K}_1(n, \varepsilon)$, and thus $W_g(\beta_0) = W_g(\beta_1)$. Therefore, we can extend W_g to $\mathcal{K}(n, \varepsilon)$ by making it constant on each component.

We can extend W_g to a knot invariant for all contractible knots embedded in $P\Omega N$ using Lemma A.1, which will be proved in Appendix A. For any smooth contractible knot β, we can approximate β by a piecewise linear knot β' that is homotopic to β. Then we set $W_g(\beta) = W_g(\beta')$. W_g is well-defined according to Lemma A.1. □

Now, we are ready to prove Theorem 1.12.

Proof of Theorem 1.12. Let $\beta = P \circ \tilde{\gamma}$ be the projectivized unit tangent vector field of γ.
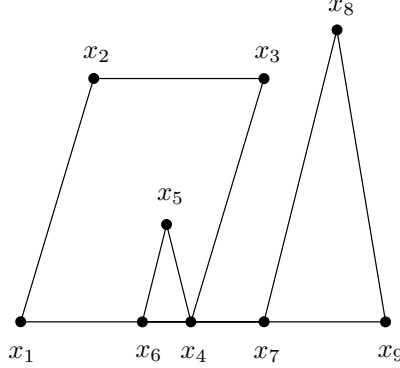
FIGURE 11. This is the projection of a knot $\beta \in X_{24}$ to N. Notice that x_4 is on the edge from x_9 to x_1 and also the edge from x_6 to x_7. Hence, there are two singularities involving x_4. There are also two singularities involving x_6 and x_7, and thus $\beta \in \mathcal{K}'_4(9, \varepsilon)$. This counterexample shows that $X_m = \mathcal{K}'_{3n-m}(n, \varepsilon)$ is not true in general when m is big.

If β is not contractible, then β is a non-trivial knot. Assume that β is contractible. We are going to show that $W_g(\beta) > 0$ for some g, while $W_g(\text{Unknot}) = 0$ for any g.

Assume that γ has no self-intersections. Then β is not contractible by Proposition 3.4. So γ has at least one self-intersection.

We will start at any point on γ and trace along γ until hitting the trace. To be precise, let $q = \max\{t : \gamma|_{[0,t]}$ has no self-intersection$\}$. Then there is $p \in [0, q)$ such that $\gamma(p) = \gamma(q)$ and β has a crossing at (p, q). By Lemma 3.7, the crossing of β at (p, q) is positive. Separate $\hat{\beta}$ into two arcs by cutting at $\hat{\beta}(p)$ and $\hat{\beta}(q)$. Then $\hat{\beta}|_{[p,q]}$ will be one of these two arcs. Glue $\hat{\beta}|_{[p,q]}$ to $\hat{\beta}_{(p,q)}$, obtaining a closed curve $\hat{\beta}'$. We can gradually widen the angle of $\gamma|_{[p,q]}$ at the corner until it becomes a simple smooth closed curve, and $\hat{\beta}'$ will converge to the unit tangent vector field along that simple smooth closed curved. By Proposition 3.4, $\hat{\beta}'$ is not contractible in ΩN.

Denote by g the non-orientable homotopy type of $\hat{\beta}'$, then $W_g(\beta) \geq 1$ by Lemma 3.7. Since $W_g(\text{Unknot}) = 0$, β is isotopically non-trivial. $\square$

Actually, a stronger (but more technical) result can be proved with exactly the same proof.

Theorem 3.14. *Suppose that $\gamma_1 : [0, 1] \to N$ is smoothly immersed curve without self-tangencies and that $\beta_2 : [0, 1] \to P\Omega N$ is a smoothly embedded curve connecting the end points of $\beta_1 := P \circ \tilde{\gamma}$. Glue β_2 to β_1, obtaining a closed curve β in $P\Omega N$. If γ_1 has at least one self-intersection, γ_1 and $\pi \circ \beta_2$ have no intersections and $\pi \circ \beta_2$ has no self-intersections, then β is isotopically non-trivial.*

Proof. Just let

$$\beta = \begin{cases} \beta_1(2t) & \text{if } t \in [0, \tfrac{1}{2}], \\ \beta_2(2 - 2t) & \text{if } t \in [\tfrac{1}{2}, 1]. \end{cases}$$

and $\gamma = \pi \circ \beta$. Let $q = \max\{t : \gamma|_{[0,t]}$ has no self-intersection$\}$. Then there is $p \in [0, q)$ such that $\gamma(p) = \gamma(q)$ and β has a crossing at (p, q).

Note that $q < \frac{1}{2}$ because γ_1 has at least one self-intersection. The rest of the proof is very similar to the proof of Theorem 1.12 since it only involves $\beta_{[p,q]}$.

$\square$

4. Closed geodesics tangent to the boundary

In this section, M and N will be two Riemannian surfaces with the same scattering data rel $h : \partial M \to \partial N$ where h is an isometry. Also, M is assumed to be simple. $\varphi : \partial \Omega M \to \partial \Omega N$ is the induced bundle map defined in (1.1). In this section, we shall prove Theorem 1.10 by studying closed geodesics tangent to the boundary.

Recall that $L = \tau_N(\varphi(X)) - \tau_M(X) \geq 0$ is a constant. We need to show that $L = 0$.

For any $Y \in \partial_0 \Omega N$, Recall that γ_Y is the limit of geodesic segments γ_X as $X \to Y$ where $X \in \partial_+ \Omega N$. γ_Y is a closed geodesic of length L.

Proposition 4.1. *If $L > 0$, then $P \circ \tilde{\gamma}_Y$ is contractible in $P\Omega N$ for any $Y \in \partial_0 \Omega N$.*

Proof. Pick $p \in \partial \Omega N$. Let $Y \in \partial_0 \Omega_p N$ be one of the two unit vector at p which are tangent to ∂N. Put

$$Y_s = \cos(\pi s)Y + \sin(\pi s)\nu(x),$$

for each $s \in [0, 1]$.

Now define a continuous family of loops $H : [0, 1] \times \mathbb{R}/\mathbb{Z} \to \Omega N$ as

$$H_s(t) = \begin{cases} \gamma'_{Y_s}(3\tau(Y_s)t) & \text{if } 0 \leq t \leq \frac{1}{3}, \\ \alpha_N(Y_{(2-3t)s}) & \text{if } \frac{1}{3} \leq t \leq \frac{2}{3}, \\ Y_{(3t-2)s} & \text{if } \frac{2}{3} \leq t \leq 1. \end{cases}$$

We shall show that $H_1|_{[\frac{1}{3}, 1]}$ is contractible. Since N is a disk, there is a diffeomorphism $\psi : N \to \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$. For any $X \in \Omega N$ and $x \in N$, let $\xi(x, X)$ be the unit vector based at x such that $\psi_*(\xi(x, X))$ and $\psi_*(X)$ have the same directions as two vectors in $\mathbb{R}^2$.

Since N is a disk, ∂N is homotopic the constant curve at p. So there is a homotopy $q : [0, 1] \times [0, 1] \to N$ such that $q(1, \cdot) = p$ and that $q(0, t) = \pi(\alpha_N(Y_t))$. Next, define a continuous family of loops $G : [0, 1] \times \mathbb{R}/\mathbb{Z} \to \Omega N$ as

$$G_s(t) = \begin{cases} \xi(q(s, 2t), \alpha_M(Y_{2t})) & \text{if } 0 \leq t \leq \frac{1}{2}, \\ Y_{2-2t} & \text{if } \frac{1}{2} \leq t \leq 1. \end{cases}$$

Let A_t be the angle that $r_1(\bar{t}) := \xi(p, \alpha_M(Y_{\bar{t}}))$ rotates by as $\bar{t}$ goes from 0 to t. We shall show that $A_1 = \pi$. Let B_t be the angle that $r_2(\bar{t}) := \xi(p, \frac{d}{d\bar{t}}q(0, \bar{t}))$ rotates by as $\bar{t}$ goes from 0 to t. Notice that $\psi(q(0, \bar{t}))$ goes around the unit circle in $\mathbb{R}^2$ for a full circle as t goes from 0 to 1. Hence $B_1 = 2\pi$. Let C_t be the signed angle between r_1 and r_2. Since $r_1(0)$ and $r_2(0)$ have the same direction, $C_t = B_t - A_t$. For any $t \in (0, 1)$, $\alpha_M(Y_{\bar{t}})$ is not tangent to ∂N, and hence $C_t \in [0, \pi]$ for $t \in [0, 1]$. Since $r_1(1)$ and $r_2(1)$ have opposite directions, $C_1 = \pi$, which implies that $A_1 = \pi$. Therefore, $\xi(p, \alpha_M(Y_t))$ rotates counterclockwise by π as t goes from 0 to 1. On

the other hand, the Y_{2-2t} rotate by π clockwise as t goes from $\frac{1}{2}$ to 1. Hence G_1 is contractible. It follows that G_0 is contractible, and thus $H_1|_{[\frac{1}{3},1]}$ is contractible.

Since $H_0|_{[\frac{1}{3},1]}$ is constant and $H_0|_{[0,\frac{1}{3}]}$ coincides with $\tilde{\gamma}_Y$, H_0 is homotopic to $\tilde{\gamma}_Y$. Since $H_1|_{[\frac{1}{3},1]}$ is contractible and $H_1|_{[0,\frac{1}{3}]}$ coincides with $\tilde{\gamma}_{-Y}$, H_1 is homotopic to $\tilde{\gamma}_{-Y}$. Therefore, $\tilde{\gamma}_Y$ is homotopic to $\tilde{\gamma}_{-Y}$.

If we rotate each vector counterclockwise by π, then $\tilde{\gamma}_{-Y}$ becomes $R(\tilde{\gamma}_Y)$. Hence $\tilde{\gamma}_Y$ is homotopic to $R(\tilde{\gamma}_Y)$. It follows that $[\tilde{\gamma}_Y] = [R(\tilde{\gamma}_Y)] = -[\tilde{\gamma}_Y]$, where $[\alpha]$ means the homology class of α. Since N is a disk, ΩN is a solid torus, and hence $H_1(\Omega N, \mathbb{Z}) = \mathbb{Z}$. Thus $[\tilde{\gamma}_Y] = 0$, that is, $\tilde{\gamma}_Y$ is contractible. Hence $P \circ \tilde{\gamma}_Y$ is contractible in $P\Omega N$. $\qquad\square$

Fix an orientation of ∂N and let $X_0(x)$ be the unit vector tangent to ∂N at $x \in \partial N$ such that $X_0(x)$ and ∂N have the same orientation. Let $h_1 : \mathbb{R}/\mathbb{Z} \to \partial N$ be an orientation preserving diffeomorphism. Pick $\varepsilon > 0$ small and let $T : \partial N \to \partial N$ be a diffeomorphism defined as

$$T(x) = h_1(h_1^{-1}(x) + \varepsilon),$$

and let $X_1(x) \in \partial_+\Omega_x N$ be the vector which is tangent to the geodesic from x to $T(x)$. Finally, put $X_2(x) = \alpha(X_1(T^{-1}(x)))$. When $T^{-1}(x)$, x and $T(x)$ are close, both $X_1(x)$ and $X_2(x)$ are close to $X_0(x)$, so we may assume that the angle between $X_1(x)$ and $X_2(x)$ is smaller than π by picking ε small enough. See Figure 12.
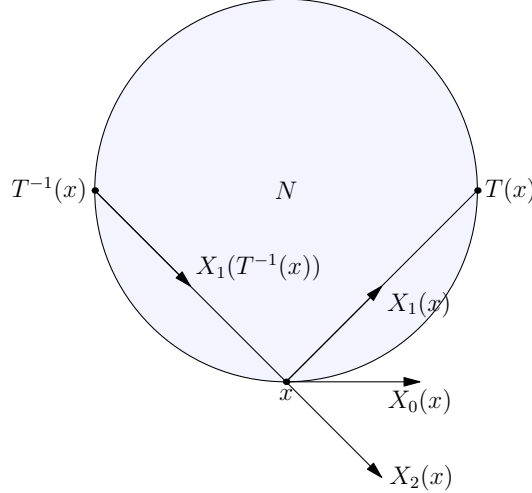

FIGURE 12

Now $X_1(x)$ and $X_2(x)$ separate the circle $\Omega_x N$ into two segments. Let $A(x)$ be the segment containing $X_0(x)$ (which is the shorter segment). Then $A = \bigcup_{x \in \partial N} A(x)$ is an annulus with boundaries $X_1(\partial N)$ and $X_2(\partial N)$. We have a natural diffeomorphism $u : \mathbb{R}/\mathbb{Z} \times [0,1] \to A$ where $u(s,t)$ is the unique vector in $\partial\Omega_{h_1(s)}N$ such that

$$t = \frac{\text{The angle between } u(x,t) \text{ and } X_1(h_1(s))}{\text{The angle between } X_1(h_1(s)) \text{ and } X_2(h_1(s))}.$$

In particular, we have $u(s,0) = X_1(h_1(s))$ and $u(s,1) = X_2(h_1(s))$.

Thus we can can break A down to a family of disjoint curves $\eta_x : [0,1] \to A$ from $X_1(x)$ to $\alpha(X_1(x))$ defined as

$$\eta_{h_1(s)}(t) = u(s + \varepsilon t, t).$$

See Figure 13.
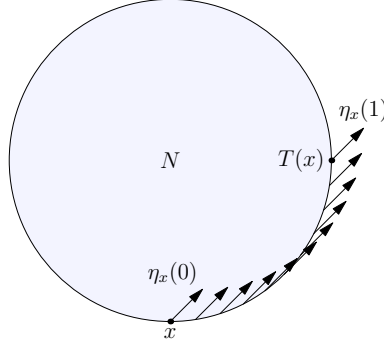

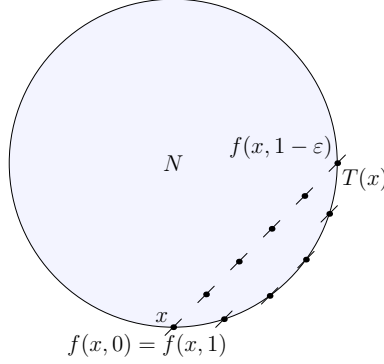FIGURE 13. Values of η_x from 0 to 1

FIGURE 14. Values of $f(x, \cdot)$ from 0 to 1

Define

$$f : \partial N \times \mathbb{R}/\mathbb{Z} \to P\Omega N$$

as

$$f(x,t) = \begin{cases} P\left(\tilde{\gamma}_{X_1(x)}(\frac{t}{1-\varepsilon})\right) & \text{if } 0 \leq t \leq 1 - \varepsilon, \\ P\left(\eta_x(\frac{1-t}{\varepsilon})\right) & \text{if } 1 - \varepsilon \leq t \leq 1. \end{cases}$$

See Figure 14.

Proposition 4.2. $f : \partial N \times \mathbb{R}/\mathbb{Z} \to P\Omega N$ *is an embedding.*

Proof. Suppose that $f(x,t) = f(x',t')$. If $0 < t < 1 - \varepsilon$, then $\pi(f(x,t))$ is not on the boundary, so $\pi(f(x',t'))$ is also not on the boundary, which implies that $0 < t' < 1 - \varepsilon$. However, $\pi \circ f(x,\cdot)|_{(0,1-\varepsilon)}$ and $\pi \circ f(x',\cdot)|_{(0,1-\varepsilon)}$ are geodesics in N, so they always intersect transversely, and thus $f(x,t)$ and $f(x',t')$ are equal if and

only if $(x,t) = (x',t')$. If $1 - \varepsilon \le t \le 1$, then $1 - \varepsilon \le t' \le 1$. Now $f(x,t) = P(\eta_x(t))$ and $f(x',t') = P(\eta_{x'}(t'))$, where $P : A \to P\Omega N$ is injective because the angle between $X_1(x)$ and $X_2(x)$ is smaller than π. Hence $f(x,t) = f(x',t')$ if and only if $\eta_x(t) = \eta_{x'}(t)$, which, by our definition of η, is equivalent to $(x,t) = (x',t')$. Therefore, f is an embedding of the torus $\mathbb{R}/\mathbb{Z} \times \partial N$ into the solid torus $P\Omega N$. $\quad\square$

Proposition 4.3. $f(x, \cdot)$ *is contractible in $P\Omega N$.*

Proof. We shall show that $f(x, \cdot)$ is homotopic to $P \circ \tilde{\gamma}_{X_0(x)}$. As $\varepsilon \to 0$, T converges to the identity map, X_1 and X_2 converge to X_0, and $f(x, \frac{t}{1-\varepsilon})$ converges to $P(\tilde{\gamma}_{X_0}(t))$ for $t \in [0, 1 - \varepsilon]$. Thus $f(x, \cdot)$ is homotopic to $P \circ \tilde{\gamma}_{X_0(x)}$. $\quad\square$

Proposition 4.4. $f(x, \cdot)$ *is isotopically trivial in $P\Omega N$.*

Proof. Define $\beta_1 : \mathbb{R}/\mathbb{Z} \to \partial N \times \mathbb{R}/\mathbb{Z}$ as

$$\beta_1(t) = (h_1(t), 0),$$

and define $\beta_2 : \mathbb{R}/\mathbb{Z} \to \partial N \times \mathbb{R}/\mathbb{Z}$ as

$$\beta_2(t) = (h_1(0), t).$$

Let $p = \beta_1(0)$, then

$$\pi_1(\partial N \times \mathbb{R}/\mathbb{Z}, p) = \{[\beta_1]_p^k [\beta_2]_p^l : k, l \in \mathbb{Z}\} \simeq \mathbb{Z}^2.$$

Let

$$f_* : \pi_1(\partial N \times \mathbb{R}/\mathbb{Z}, p) \to \pi_1(P\Omega N, f(p))$$

be the induced homomorphism between fundamental groups. Since $\pi_1(P\Omega N, f(p)) = \mathbb{Z}$, f_* is not injective. Since [12, Corollary 3.3] a two-sided surface f is incompressible if and only if f_* is injective, the torus $f(\partial N \times \mathbb{R}/\mathbb{Z})$ has a compressing disk embedded in $P\Omega N$.

As $\varepsilon \to 0$, $f \circ \beta_1$ converges to the projectivized unit tangent vector field along ∂N, so $f_*([\beta_1]_p) \ne 0$ by Proposition 3.4. Since $f(h_1(0), \cdot)$ is homotopic to $\gamma_{X_0(h_1(0))}$, both of them are contractible in $P\Omega N$ by Proposition 4.1. Since $\pi_1(P\Omega N, f(p)) = \mathbb{Z}$, there is no difference between free homotopy and based homotopy, hence $f_*([\beta_2]_p) = 0$.

Now, let B be a compressing disk of the torus $f(\partial N \times \mathbb{R}/\mathbb{Z})$, then ∂B is a circle embedded $f(\partial N \times \mathbb{R}/\mathbb{Z})$ which is contractible in $P\Omega N$. It follows that ∂B is homotopic to $f \circ \beta_2$ on $f(\mathbb{R}/\mathbb{Z} \times S)$. Any two simple closed essential curves on a surface are isotopic to each other if and only if they are freely homotopic to each other [9]. Therefore, ∂B is isotopic to $f \circ \beta_2 = f(h_1(0), \cdot)$ on $f(\mathbb{R}/\mathbb{Z} \times S) \subset P\Omega N$. Since ∂B bounds a disk B in $P\Omega N$, ∂B is isotopically trivial, and so is $f(h_1(0), \cdot)$ in $P\Omega N$. This completes the proof of Proposition 4.4 $\quad\square$

Notice (see below) that Proposition 4.4 contradicts Theorem 3.14 when $L > 0$, which proves Theorem 1.10.

Proof of Theorem 1.10. Suppose that $L > 0$. Pick any $x \in \partial M$. Let $\gamma_1 = \gamma_{X_1(x)}$ and $\beta_2 = P \circ \eta_x$.

If γ_1 has no self-intersections, then $\gamma_{X_0(x)}$, the limit of γ_1 as $\varepsilon \to 0$, also has no self-intersections. By Proposition 3.4, $P \circ \tilde{\gamma}_{X_0(x)}$ is not contractible in $P\Omega N_1$, which contradicts Proposition 4.1. Therefore, γ_1 has at least one self-intersection.

Let $\beta_1 = P \circ \tilde{\gamma}_1$ Notice that γ_1 and $P \circ \beta_2$ have no intersections except at end points and that $f(x, \cdot)$ is the closed curve obtained by gluing β_1 and β_2. By Theorem 3.14, $f(x, \cdot)$ is isotopically non-trivial in $P\Omega N_1$, which contradicts Proposition 4.4. Thus $L = 0$, which finishes the proof of Theorem 1.10. $\qquad\square$

Appendix A. Approximating smooth knots by piecewise linear knots

The goal of this section is to prove the following lemma, which allows us to approximate knot isotopies using piecewise-linear knot isotopies.

Lemma A.1. *Suppose that there is a continuous knot isotopy $G : [0,1] \times \mathbb{R}/\mathbb{Z} \to P\Omega N$. Then there is continuous family of knot isotopies $H : [0,1] \times [0,1] \times \mathbb{R}/\mathbb{Z} \to P\Omega N$ such that $H(0, \cdot, \cdot) = G$, that $H(l, s, \cdot)$ is a knot embedded in $P\Omega N$ for each $(l, s) \in [0,1] \times [0,1]$, and that $H(1, s, \cdot)$ is a piecewise linear knot for each $s \in [0,1]$.*

The following proposition and its corollaries will be our main tool used in this section.

Proposition A.2. *Assume that K is a compact smooth manifold and M is a Riemannian manifold. Suppose that $G : K \times [a, b] \to M$ is smooth and that each $G(s, \cdot)$ is a smooth curve whose speed is never 0. For any $\varepsilon > 0$, there is $\delta > 0$ such that the angles between $G(s, \cdot)|_{[t_1, t_2]}$ and the minimal geodesic connecting its end points are smaller than ε whenever $|t_1 - t_2| < \delta$.*

Proof. Pick any $\varepsilon > 0$. There is $\varepsilon_1 > 0$ such that $|\theta| < \varepsilon$ if $|\cos(\theta) - 1| < \varepsilon_1$ and if $|\theta| \leq \pi$.

Define $L : K \times [a, b] \times [a, b] \to \mathbb{R}$ as

$$L(s, t_1, t_2) = \begin{cases} \ell(G(s, \cdot)|_{[t_1, t_2]}) & \text{if } t_2 \geq t_1, \\ -L(s, t_1, t_2) & \text{if } t_2 < t_1, \end{cases}$$

where $\ell(G(s, \cdot)|_{[t_1, t_2]})$ is the length of $G(s, \cdot)|_{[t_1, t_2]}$. Similarly, define $D : K \times [a, b] \times [a, b] \to \mathbb{R}$ as

$$D(s, t_1, t_2) = \begin{cases} d(G(s, t_1), G(s, t_2)) & \text{if } t_2 \geq t_1, \\ -D(s, t_1, t_2) & \text{if } t_2 < t_1. \end{cases}$$

For any fixed s, we have

(A.1) $$L(s, t_1, t_2) - D(s, t_1, t_2) = o((t_2 - t_1)^2)$$

as $t_2 \to t_1$. Put

$$Q(s, t_1, t_2) = \frac{\frac{\partial}{\partial t_2} D(s, t_1, t_2)}{\frac{\partial}{\partial t_2} L(s, t_1, t_2)}.$$

Then we have $Q(s, t_1, t_1) = 1$ by (A.1).

We shall show the Q is continuous near $K \times \Delta[a, b]$ where $\Delta[a, b]$ is the diagonal of $[a, b] \times [a, b]$. Since G is continuous and K is compact, there is $\delta_1 > 0$ such that $d(G(s, t_1), G(s, t_2)) < inj(M)$ if $|t_1 - t_2| < \delta_1$. Since the squared distance function is smooth within the injectivity radius, D^2 is smooth on $K \times V_{\delta_1}$ where $V_{\delta_1} = \{(t_1, t_2) \in [a, b] \times [a, b] : |t_1 - t_2| < \delta_1\}$ and hence $\frac{\partial}{\partial t_2} D$ is continuous. Also, $\frac{\partial}{\partial t_2} L$ is obviously continuous on $K \times [a, b] \times [a, b]$ (since $L(s, t_1, \cdot)$ is just the signed arc length). Therefore, Q is continuous on $K \times V_{\delta_1}$. Since Q is continuous and K is

compact, there is $\delta \in (0, \delta_1)$ such that $|Q(s, t_1, t_2) - Q(s, t_1, t_1)| < \varepsilon_1$ if $|t_1 - t_2| < \delta$, that is, $|Q - 1| < \varepsilon_1$ on $K \times V_\delta$.

For any $(s, t_1, t_2) \in K \times V_\delta \setminus \Delta[a, b]$, let $\theta(s, t_1, t_2)$ be the angle between $G(s, \cdot)|_{[t_1, t_2]}$ and the minimal geodesic connecting its end points at the endpoint $G(s, t_2)$. By the first variation of arc length,

$$\cos \theta(s, t_1, t_2) = \frac{\frac{\partial}{\partial t_2} D(s, t_1, t_2)}{\frac{\partial}{\partial t_2} L(s, t_1, t_2)} = Q(s, t_1, t_2).$$

Hence $\theta < \varepsilon$ on $K \times V_\delta$. Therefore the angles between $G(s, \cdot)|_{[t_1, t_2]}$ and the minimal geodesic connecting its end points are smaller than ε whenever $|t_1 - t_2| < \delta$. $\qquad \square$

For any compact Riemannian surface N, we can apply the above proposition to unit-speed linear curves of length ≤ 1 in $P\Omega N$ (which has the Sasakian metric on it), which is a compact family of curves in $P\Omega N$. For any $p, q \in P\Omega N$ such that $d(p, q) < inj(P\Omega N)$ and that $d_h(p, q) < inj(N)$, let $\gamma_{p,q}$ be the minimal geodesic connecting p and q. Recall that $\gamma_{p,q}^0$ is the minimal linear curve connecting p and q.

Corollary A.3. *Assume that N is a compact Riemannian manifold. For any $\varepsilon > 0$, there is $\delta > 0$ such that the angles between $\gamma_{p,q}$ and $\gamma_{p,q}^0$ are smaller than ε for any $p, q \in P\Omega N$ such that $0 < d(p, q) < \delta$.*

Suppose that $p, q, r \in P\Omega N$ are close enough that there are minimal linear curves $\gamma_{p,q}^0$, $\gamma_{q,r}^0$ and $\gamma_{p,r}^0$. Denote by $A(p, q, r)$ the sum of the three angles between $\gamma_{p,q}^0$, $\gamma_{q,r}^0$ and $\gamma_{p,r}^0$. Since the sum of the inner angles of small geodesic triangles are close to π, Corollary A.3 implies that $A(p, q, r)$ is also close to π when p, q and r are close enough.

Corollary A.4. *Assume that N is a compact Riemannian manifold. For any $\varepsilon > 0$, there is $\delta > 0$ such that $|A(p, q, r) - \pi| < \varepsilon$ for any $p, q, r \in P\Omega N$ such that $d(p, q), d(p, r), d(q, r) \in (0, \delta)$.*

The following result follows from Proposition A.2 and Corollary A.3.

Corollary A.5. *Assume that K is a compact smooth manifold and N is a compact Riemannian manifold. Suppose that $G : K \times [a, b] \to P\Omega N$ is smooth and that each $G(s, \cdot)$ is a smooth curve whose speed is never 0. For any $\varepsilon > 0$, there is $\delta > 0$ such that the angles between $G(s, \cdot)|_{[t_1, t_2]}$ and $\gamma_{G(s,t_1),G(s,t_2)}^0$ are smaller than ε whenever $|t_1 - t_2| < \delta$.*

Proof of Lemma A.1. We shall assume that G is smooth since it is standard to approximate continuous isotopies by smooth isotopies.

Put $\varepsilon = 0.1$. By Corollary A.5, there is $\delta_1 > 0$ such that the angles between $G(s, \cdot)|_{[t_1, t_2]}$ and $\gamma_{G(s,t_1),G(s,t_2)}^0$ are smaller than ε whenever $|t_1 - t_2| < \delta_1$. By Corollary A.4 there is $\varepsilon_0 > 0$ such that

$$(\text{A.2}) \qquad\qquad |A(p, q, r) - \pi| < \varepsilon$$

for any $p, q, r \in P\Omega N$ such that $d(p, q), d(p, r), d(q, r) \in (0, \varepsilon_0)$. Also, by corollary A.3, there is $\varepsilon_1 \in (0, \varepsilon_0)$ such that the angles between $\gamma_{p,q}^0$ and $\gamma_{p,q}$ are smaller than ε for any $p, q \in P\Omega N$ such that $0 < d(p, q) < \varepsilon_1$. Since G is continuous, there is $\delta_2 \in (0, \delta_1)$ such that $d(G(s, t_1), G(s, t_2)) < \varepsilon_1$ if $d(t_1, t_2) < \delta_2$.

For each $(s,t) \in [0,1] \times \mathbb{R}/\mathbb{Z}$, let $I(t) = \mathbb{R}/\mathbb{Z} \setminus (t - \delta_2, t + \delta_2)$, then $D(s,t) :=$ $d(G(s,t), G(s, I(t))) > 0$. Let $\varepsilon_2 = \min(\varepsilon_1, \inf_{(s,t) \in [0,1] \times \mathbb{R}/\mathbb{Z}} D(s,t))$, then $\varepsilon_2 > 0$. Since G is continuous, there is $\delta \in (0, \delta_2)$ such that $d(G(s,t_1), G(s,t_2)) < \frac{1}{2}\varepsilon_2$ if $d(t_1, t_2) < \delta$.

Pick $n \in \mathbb{N}$ such that $n\delta > 1$. Define $H : [0,1] \times [0,1] \times \mathbb{R}/\mathbb{Z} \to P\Omega N$ as

$$H(l, s, \frac{k+t}{n}) = \begin{cases} \gamma^0_{G(s,\frac{k}{n}),G(s,\frac{k+l}{n})}(\frac{t}{l}) & \text{if } 0 \le t < l \\ G(s, \frac{k+t}{n}) & \text{if } l \le t \le 1 \end{cases}$$

where $k \in \mathbb{Z}/n\mathbb{Z}$, and $t \in [0,1]$. It is obvious that $H(0, \cdot, \cdot) = G$ and $H(1, \cdot, \cdot)$ is an isotopy of piecewise linear knots. We shall show that each $H(l, s, \cdot)$ is a knot embedded in $P\Omega N$.

Suppose that $H(l, s, \cdot)$ has a self-intersection, that is, $H(l, s, \frac{k_1+t_1}{n}) = H(l, s, \frac{k_2+t_2}{n})$ for some $k_i \in \mathbb{Z}/n\mathbb{Z}$ and $t_i \in [0,1)$ such that $k_1 \ne k_2$ or $t_1 \ne t_2$.

Since $G(s, \cdot)$ is an embedding, we have either $t_1 < l$ or $t_2 < l$. Without loss of generality, assume that $t_1 < l$. Write $t_3 = \frac{k_1+t_1}{n}$ and $t_4 = \frac{k_2+t_2}{n}$. We shall show that $\gamma^0_{G(s,\frac{k_1}{n}),G(s,\frac{k_1+l}{n})}$ is in $N_{\frac{\varepsilon_2}{2}}(G(s,\frac{k_1}{n}))$, the ball of radius $\frac{\varepsilon_2}{2}$ centered at $G(s,\frac{k_1}{n})$. Put $p = G(s,\frac{k_1}{n})$ and $q = G(s,\frac{k_1+l}{n})$. Since $d(\frac{k_1}{n},\frac{k_1+l}{n}) < \delta < \frac{1}{3}\delta_2$, $d(p,q) < \varepsilon_1$. Define $Q : [0,1] \to \mathbb{R}$ as

$$Q(t) = \frac{\frac{\partial}{\partial t}d(p, \gamma^0_{p,q}(t))}{\frac{\partial}{\partial t}\ell(\gamma^0_{p,q}|_{[0,t]})}$$

By the first variation of arc length, $Q(t) = \cos\theta(t)$ where $\theta(t)$ is the angle between $\gamma^0_{p,q}|_{[0,t]}$ and the minimal geodesic connecting their end points at the end point $\gamma^0_{p,q}(t)$. Since $d(p,q) < \frac{1}{2}\varepsilon_2 < \varepsilon_1$, $\theta(t) < \varepsilon = 0.01$, and hence $Q(1) > 0$. If $Q(t) = 0$ for some $t \in [0,1)$, then let $t_0 = \sup\{t \in [0,1) : Q(t) = 0\}$. Then we have $Q(t_0) = 0$ and $Q(t) > 0$ for $t \in (t_0, 1]$. Since

$$d(p, \gamma^0_{p,q}(t_0)) = d(p,q) - \int_{t_0}^1 \frac{\partial}{\partial t}d(p, \gamma^0_{p,q}(t))dt$$

$$= d(p,q) - \int_{t_0}^1 \frac{\frac{\partial}{\partial t}\ell(\gamma_{p,q}|_{[0,t]})}{Q(t)}dt$$

$$< d(p,q)$$

$$< \varepsilon_1,$$

$\theta(t_0) < \varepsilon$, and hence $Q(t_0) > 0$, which contradict our assumption that $Q(t_0) = 0$. So $Q(t) > 0$ for any $t \in [0,1]$, which implies that $d(p, \gamma^0_{p,q}(t))$ is strictly increasing. Hence $d(p, \gamma^0_{p,q}(t)) \le d(p,q) < \frac{\varepsilon_2}{2}$ for all $t \in [0,1)$, that is, $\gamma^0_{G(s,\frac{k_1+l}{n}),G(s,\frac{k_1}{n})}$ is in $N_{\frac{\varepsilon_2}{2}}(p)$. It also follows that $\gamma^0_{G(s,\frac{k_1+l}{n}),G(s,\frac{k_1}{n})}$ has no self-intersections, and thus $k_1 \ne k_2$.

Assume that $d(\frac{k_1}{n},\frac{k_2}{n}) \ge \delta_2$. Then $G(s,\frac{k_2}{n}) \notin N_{\varepsilon_2}(G(s,\frac{k_1}{n}))$. Hence we have $N_{\frac{\varepsilon_2}{2}}(G(s,\frac{k_1}{n})) \bigcap N_{\frac{\varepsilon_2}{2}}(G(s,\frac{k_2}{n})) = \emptyset$. When $t_2 \le l$, $H(l,s,t_4)$ is on $\gamma^0_{G(s,\frac{k_2}{n}),G(s,\frac{k_2+l}{n})}$, which is contained in $N_{\frac{\varepsilon_2}{2}}(G(s,\frac{k_2}{n}))$, and hence $H(l,s,t_4) \in N_{\frac{\varepsilon_2}{2}}(G(s,\frac{k_2}{n}))$. However, for the same reason, $H(l,s,t_3) \in N_{\frac{\varepsilon_2}{2}}(G(s,\frac{k_1}{n}))$, and hence $H(l,s,t_3) \ne H(l,s,t_4)$, which contradicts our assumption. When $t_2 \ge l$, $H(l,s,t_4) = G(s,t_4)$. Since $d(\frac{k_2}{n},t_4) = \frac{t}{n} < \frac{1}{n} < \delta$, $d(G(s,\frac{k_2}{n}),G(s,t_4)) < \frac{\varepsilon_2}{2}$, and hence $H(l,s,t_4) \in$

$N_{\frac{\varepsilon_2}{2}}(G(s, \frac{k_2}{n}))$. Again, $H(l, s, t_3) \neq H(l, s, t_4)$, which contradicts our assumption. Therefore, $d(\frac{k_1}{n}, \frac{k_2}{n}) < \delta_2$. We shall assume that $\frac{k_2}{n} \in (\frac{k_1}{n}, \frac{k_1}{n} + \delta_2)$, and the other case $(\frac{k_1}{n} \in (\frac{k_2}{n}, \frac{k_2}{n} + \delta_2))$ can be addressed similarly. Then we have $d(H(l, s, \frac{k_1+l}{n}), H(l, s, \frac{k_2}{n})) < \varepsilon_2$.

We shall show that $d(G(s, \frac{k_2}{n}), H(l, s, t_4)) < \varepsilon_2$. If $t_2 \geq l$, then $H(l, s, t_4) = G(s, t_4)$. We have $d(G(s, \frac{k_2}{n}), G(s, t_4)) < \varepsilon_2$ since $d(\frac{k_2}{n}, t_4) < \delta$. If $t_2 < l$, then $d(G(s, \frac{k_2}{n}), H(l, s, t_4)) < d(G(s, \frac{k_2}{n}), H(l, s, \frac{k_2+l}{n})) < \varepsilon_2$. Using the same argument, we have $d(G(s, \frac{k_1+l}{n}), H(l, s, t_3)) < \varepsilon_2$.

Write $p_1 = H(l, s, \frac{k_1+l}{n})$, $p_2 = H(l, s, \frac{k_2}{n})$ and $p_3 = H(l, s, t_3) = H(l, s, t_4)$. Then the angle between $\gamma^0_{p_i, p_j}$ and $G(s, \cdot)$ is at most ε for any $i \neq j$. Hence the angle between $\gamma^0_{p_1, p_2}$ and $\gamma^0_{p_1, p_3}$ and the angle between $\gamma^0_{p_1, p_2}$ and $\gamma^0_{p_2, p_3}$ are at least $\pi - 2\varepsilon$. Hence $A(p_1, p_2, p_3) > 2\pi - 4\varepsilon$, which contradicts (A.2).

This completes the proof of Lemma A.1. $\qquad\square$

References

[1] Stephanie B. Alexander, I. David Berg, and Richard L. Bishop, *The Riemannian obstacle problem*, Illinois Journal of Mathematics **31** (1987), no. 1, 167–184.

[2] G. Besson, G. Courtois, and S. Gallot, *Entropies et rigidités des espaces localement symétriques de courbure strictement négative*, Geom. Funct. Anal. **5** (1995), no. 5, 731–799. MR 1354289 (96i:58136)

[3] Dmitri Burago and Sergei Ivanov, *Boundary rigidity and filling volume minimality of metrics close to a flat one*, Ann. of Math. (2) **171** (2010), no. 2, 1183–1211. MR 2630062 (2011d:53079)

[4] ———, *Area minimizers and boundary rigidity of almost hyperbolic metrics*, Duke Mathematical Journal **162** (2013), no. 7, 1205–1248.

[5] S. Chmutov, V. Goryunov, and H. Murakami, *Regular Legendrian knots and the HOMFLY polynomial of immersed plane curves*, Math. Ann. **317** (2000), no. 3, 389–413. MR 1776110 (2001h:57009)

[6] Christopher B. Croke, *Scattering rigidity with trapped geodesics*, 2011.

[7] Christopher B. Croke and Bruce Kleiner, *A rigidity theorem for simply connected manifolds without conjugate points*, Ergodic Theory Dynam. Systems **18** (1998), no. 4, 807–812. MR 1645381 (99j:53035)

[8] J.E. Eaton, *On spherically symmetric lenses*, IRE Trans. Antennas Propa. **AP-4** (1952), 66–71.

[9] D BA Epstein, *Curves on 2-manifolds and isotopies*, Acta Mathematica **115** (1966), no. 1, 83–107.

[10] Mikhael Gromov, *Filling Riemannian manifolds*, J. Differential Geom **18** (1983), no. 1, 1–147.

[11] J.H. Hannay and T.M. Haeusser, *Retroreflection by refraction*, J. Mod. Opt **40** (1993), no. 8, 1437–1442.

[12] A. Hatcher, *Notes on basic 3-manifold topology*, 2000.

[13] M. Kerker, *The scattering of light and other electromagnetic radiation*, Academic Press, 1969.

[14] René Michel, *Sur la rigidité imposée par la longueur des géodésiques*, Invent. Math. **65** (1981/82), no. 1, 71–83.

[15] Leonid Pestov and Gunther Uhlmann, *Two dimensional compact simple Riemannian manifolds are boundary distance rigid*, Ann. of Math. (2) **161** (2005), no. 2, 1093–1110.

Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104-6395 USA

E-mail address: `weh@math.upenn.edu`