Homotopy and cohomology of spaces of locally convex curves in the sphere

Nicolau C. Saldanha

July 23, 2004

Abstract

We discuss the homotopy type and the cohomology of spaces of locally convex parametrized curves $\gamma : [0,1] \to \mathbb{S}^2$, i.e., curves with positive geodesic curvature. The space of all such curves with $\gamma(0) = \gamma(1) = e_1$ and $\gamma'(0) = \gamma'(1) = e_2$ is known to have three connected components $X_{-1,c}$, X_1, X_{-1}. We show several results concerning the homotopy type and cohomology of these spaces. In particular, $X_{-1,c}$ is contractible, X_1 and X_{-1} are simply connected, $\pi_2(X_{-1})$ contains a copy of $\mathbb{Z}$ and $\pi_2(X_1)$ contains a copy of $\mathbb{Z}^2$. Also, $H^n(X_1, \mathbb{R})$ and $H^n(X_{-1}, \mathbb{R})$ are nontrivial for all even n. More, $\dim H^{4n-2}(X_1, \mathbb{R}) \geq 2$ and $\dim H^{4n}(X_{-1}, \mathbb{R}) \geq 2$ for all positive n.

1 Introduction

A curve $\gamma : [0,1] \to \mathbb{S}^2$ is called *locally convex* if its geodesic curvature is always positive, i.e., if $\det(\gamma(t), \gamma'(t), \gamma''(t)) > 0$ for all t. Let X be the space of all locally convex curves with prescribed initial point and initial direction: $\gamma(0) = e_1$ and $\gamma'(0) = c_0 e_2$ for some $c_0 > 0$. We define $X_Q \subset X$ to be the spaces of curves with prescribed final point and final direction. More precisely, for $Q \in SO(3)$, X_Q is the set of $\gamma \in X$ for which $\gamma(1) = Qe_1$ and $\gamma'(1) = c_1 Q e_2$ for some $c_1 > 0$. In particular, X_I is the set of closed parametrized curves of positive geodesic curvature with a prescribed base point and base direction. The topology in these spaces of curves can be taken to be C^∞, C^k for some $k \geq 2$ or H^k (as in Sobolev spaces) for some $k \geq 2$: it actually makes very little difference since it is easy to smoothen out a curve while keeping its geodesic curvature positive. In this paper we discuss the homotopy type and cohomology of the spaces X_Q. The topology of these spaces has been discussed, among others, by Little ([7]), B. Shapiro, M. Shapiro and Khesin ([9], [10], [8]) but, as far as we could ascertain, the main results presented here are new.

Let Y_Q be the set of parametrized curves $\gamma : [0,1] \to \mathbb{S}^2$ with $\gamma'(t) \neq 0$ for all t, $\gamma(0) = e_1$, $\gamma'(0) = c_0 e_2$, $\gamma(1) = Qe_1$, $\gamma'(1) = c_1 Q e_2$ ($c_0, c_1 > 0$). Clearly $X_Q \subset Y_Q$. For each $\gamma \in Y_Q$, define $\Gamma : [0,1] \to SO(3)$ by

$$\begin{pmatrix} \gamma(t) & \gamma'(t) & \gamma''(t) \end{pmatrix} = \Gamma(t) R(t),$$

$R(t)$ being an upper triangular matrix with positive diagonal (the left hand side is the 3×3 matrix with columns $\gamma(t)$, $\gamma'(t)$ and $\gamma''(t)$). In other words,

$$\Gamma(t) = \left(\gamma(t) \quad \widehat{\gamma'(t)} \quad \gamma(t) \times \widehat{\gamma'(t)} \right).$$

Recall that the universal (double) cover of $SO(3)$ is $\mathbb{S}^3 \subset \mathbb{H}$, the group of quaternions of absolute value 1: define $\tilde{\Gamma} : [0,1] \to \mathbb{S}^3$ by $\tilde{\Gamma}(0) = 1$, $\Pi \circ \tilde{\Gamma} = \Gamma$ where $\Pi : \mathbb{S}^3 \to SO(3)$ is the canonical projection. This defines an injective map from Y_Q to $Z_z \cup Z_{-z}$, where z is a quaternion with $|z| = 1$, $\Pi(z) = Q$ and Z_z is the set of continuous maps $\alpha : [0,1] \to \mathbb{S}^3$ with $\alpha(0) = 1$ and $\alpha(1) = z$. Clearly each Z_z is homotopically equivalent to $Z_1 = \Omega\mathbb{S}^3$, the space of continuous maps $\alpha : [0,1] \to \mathbb{S}^3$ with $\alpha(0) = \alpha(1) = 1$. From the Hirsch-Smale theorem ([11], [6]), the map $Y_Q \hookrightarrow Z_z \cup Z_{-z}$ is a homotopy equivalence. In particular, Y_Q has two connected components, each one mapped to one of Z_z and Z_{-z}: call them Y_z and Y_{-z}.

Theorem 1 *The inclusion $i_Q : X_Q \to Y_Q$ is homotopically surjective. More precisely, for any compact space K and any function $f : K \to Y_Q$ there exists $g : K \to X_Q$ and a homotopy $H : [0,1] \times K \to Y_Q$ with $H(0,\cdot) = f$ and $H(1,\cdot) = g$.*

Theorem 1 implies that X_Q has at least two connected components, one in each of $Y_{\pm z}$. Actually, we shall describe a set $A \subset \mathbb{S}^3$ with the following proporties. If $z \notin A$, then $X \cap Y_z$ is connected: we call this set X_z. If $z \in A$, then $X \cap Y_z$ has two connected components X_z and $X_{z,c}$: the inclusion $X_z \subset Y_z$ is homotopically surjective and $X_{z,c}$ is contractible. These facts will be proved in theorems 6 and 7. Since A and $-A = \{-z, z \in A\}$ will turn out to be disjoint, X_Q may have two or three components, depending on Q. In [7], Little proved that X_I has three connected components. B. Shapiro, M. Shapiro and Khesin studied the connected components of other X_Q and studied a similar problem in higher dimensions ([8], [9], [10]).

We know that $H^*(Y_z, \mathbb{R}) = H^*(\Omega\mathbb{S}^3, \mathbb{R}) = \mathbb{R}[\mathbf{x}]$ where $\mathbf{x} \in H^2(\Omega\mathbb{S}^3, \mathbb{R})$. Theorem 1 implies that the map $H^*(i_z) : H^*(Y_z, \mathbb{R}) \to H^*(X_z, \mathbb{R})$ is injective and therefore $\dim H^n(X_z, \mathbb{R}) > 0$ for all even n. In particular, X_z is not homotopically equivalent to a finite CW-complex.

The two spaces X_1 and X_{-1} are the objects of central interest in this paper. We just saw that their homotopy and cohomology is at least as large as that of

$\Omega\mathbb{S}^3$, and we shall see that the inclusion $X_z \subset Y_z$ is not a homotopy equivalence in either case. On the other hand, the maps $\pi_n(i_z) : \pi_n(X_z) \to \pi_n(Y_z)$ are isomorphisms for $n \leq 1$.

Theorem 2 *The spaces X_1 and X_{-1} are simply connected.*

The maps $\pi_n(i_z) : \pi_n(X_z) \to \pi_n(Y_z)$ and, more generally, $[K, i_z] : [K, X_z] \to [K, Y_z]$ are surjective. The choice of the map g in theorem 1 is uniform up to homotopy: intuitively, g is obtained from f adding many small positively oriented loops along each $\gamma = f(k)$ so that the geodesic curvature of γ becomes positive (see figure 6). This defines maps $w_{z,K} : [K, Y_z] \to [K, X_z]$ such that $[K, i_z] \circ w_{z,K}$ is the identity. In particular we have injective maps $w_{z,\mathbb{S}^n} : \pi_n(Y_z) \to \pi_n(X_z)$ which allow us to identify $\pi_n(Y_z)$ with a subgroup of $\pi_n(X_z)$. Let $G_{n,z}$ be the kernel of $\pi_n(i_z)$: we have a natural isomorphism between $\pi_n(X_z)$ and $\pi_n(Y_z) \oplus G_{n,z}$.

For $\gamma_1 \in Y_I$ and $\gamma_2 \in Y_Q$, set $(\gamma_1 * \gamma_2)(t) = \gamma_1(2t)$ for $t \leq 1/2$ and $(\gamma_1 * \gamma_2)(t) = \gamma_2(2t-1)$ for $t \geq 1/2$, thus defining $* : Y_I \times Y_Q \to Y_Q$. For $k > 0$, let $\nu^k \in X_I$ be given by

$$\nu^k(t) = \left(\frac{1 + \cos(2\pi k t)}{2}, \frac{\sqrt{2}\sin(2\pi k t)}{2}, \frac{1 - \cos(2\pi k t)}{2} \right).$$

Define $p_Q^k : X_Q \to X_Q$ by $p_Q^k(\gamma) = \nu^k * \gamma$. It turns out that $p_Q = p_Q^1$ takes X_z to X_{-z} and vice versa.

Theorem 3 *The maps p_Q and p_Q^3 are homotopic. Furthermore, given $f : \mathbb{S}^n \to X_Q$ and $H : \mathbb{B}^{n+1} \to Y_Q$ with $f(s) = H(s)$ for all $s \in \mathbb{S}^n$ there exists $\tilde{H} : \mathbb{B}^{n+1} \to X_Q$ with $\tilde{H}(s) = p_Q(f(s))$ for all $s \in \mathbb{S}^n$.*

Let p_z^k be the restriction of p_Q^k to $X_z \subset X_Q$. It follows from theorem 3 that the map $\pi_n(p_z^2) : \pi_n(X_z) \to \pi_n(X_z)$ is a projection and that the kernel of $\pi_n(p_z^2)$ is $G_{n,z}$. We do not know what the groups $G_{n,z}$ are: we know, however, that they are nontrivial.

Theorem 4 *If $-z \in A$ then there exist functions $f_z : \mathbb{S}^2 \to X_z$ and $g_z : X_z \to \mathbb{S}^2$ such that $g_z \circ f_z$ is homotopic to the identity. Also, $g_z \circ p_z^2$ is constant. In particular, $G_{n,z} = \pi_n(\mathbb{S}^2) \oplus \ker(\pi_n(g_z))$.*

We can also say something about the cohomology of X_z.

Theorem 5 *If $(-1)^n z \in A$ then $\dim H^{2n}(X_z, \mathbb{R}) \geq 2$.*

This work was motivated by the study of the differential equation of order 3:

$$u'''(t) = h_1(t)u'(t) + h_0(t)u(t), \quad t \in [0, 2\pi]. \tag{$\dagger$}$$

The set of pairs of potentials (h_0, h_1) for which equation ($\dagger$) admits 3 linearly independent periodic solutions is homotopically equivalent to X_I. This appears to be the same motivation as that of B. Shapiro and M. Shapiro for studying these same spaces. This topological study of differential equations is continuation of the work done together with Dan Burghelea and Carlos Tomei in [4] and [5].

The author would like to thank Dan Burghelea for helpful conversations. The author acknowledges the hospitality of The Mathematics Department of The Ohio State University during the winter quarter of 2004 and the support of CNPq, Capes and Faperj (Brazil).

2 The homotopy type of Y_Q

The results in this section are not new are are presented to fix notation and for the convenience of the reader; see [1] for more information concerning the geometry of curves in the sphere.

Recall that the unit tangent bundle of $\mathbb{S}^2$ is $SO(3)$: indeed, the base point, the unit tangent vector and the cross product of the two are the columns of an orthogonal matrix. Also, $\pi_1(SO(3)) = \mathbb{Z}/(2)$ and the universal (double) cover of $SO(3)$ is $\mathbb{S}^3 \subset \mathbb{H}$, the group of quaternions of absolute value equal to 1. We fix notation by taking the projection $\Pi : \mathbb{S}^3 \to SO(3)$ to be given by

$$\Pi(a+bi+cj+dk) = \begin{pmatrix} a^2+b^2-c^2-d^2 & -2ad+2bc & 2ac+2bd \\ 2ad+2bc & a^2-b^2+c^2-d^2 & -2ab+2cd \\ -2ac+2bd & 2ab+2cd & a^2-b^2-c^2+d^2 \end{pmatrix}.$$

For an immersion $\gamma : [0,1] \to \mathbb{S}^2$, define $\Gamma : [0,1] \to SO(3)$ by

$$\begin{pmatrix} \gamma(t) & \gamma'(t) & \gamma''(t) \end{pmatrix} = \Gamma(t)R(t),$$

$R(t)$ being an upper triangular matrix with positive diagonal (the left hand side is the 3×3 matrix with columns $\gamma(t)$, $\gamma'(t)$ and $\gamma''(t)$). In other words,

$$\Gamma(t) = \begin{pmatrix} \gamma(t) & \widehat{\gamma'(t)} & \gamma(t) \times \widehat{\gamma'(t)} \end{pmatrix}.$$

If $\gamma(0) = e_1$ and $\gamma'(0) = ce_2$, $c > 0$, define $\tilde{\Gamma} : [0,1] \to \mathbb{S}^3$ by $\tilde{\Gamma}(0) = 1$, $\Pi \circ \tilde{\Gamma} = \Gamma$.

For instance, if $\theta \in (0, \pi)$, set

$$\gamma(t) = \nu_\theta(t) = \left(\cos^2\theta + \sin^2\theta\cos(2\pi t), \sin\theta\sin(2\pi t), \cos\theta\sin\theta(1-\cos(2\pi t))\right).$$

The curve γ is a circle in a plane passing through e_1, parallel to e_2 and making an angle θ with e_3. Notice that ν, as defined in the introduction, is $\nu_{\pi/4}$, that $\nu_\theta \in X_I$ for $\theta < \pi/2$ and that $\nu_{\pi/2}$ is a geodesic. A simple computation yields

$$\Gamma(t) = \begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(2\pi t) & -\sin(2\pi t) \\ 0 & \sin(2\pi t) & \cos(2\pi t) \end{pmatrix} \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix}$$

and $\tilde\Gamma(t) = \exp(\pi \ell t)$ where $\ell = \cos(2\theta)i + \sin(2\theta)k$. In particular, $\tilde\Gamma(1) = -1$ for all θ.

The map ϕ taking γ to Γ defines a map from Y_Q to Z_Q, the set of maps $f : [0,1] \to SO(3)$ with $f(0) = I$ and $f(1) = Q$. Notice that Z_Q is naturally identified with $Z_z \cup Z_{-z}$ where z and $-z$ are the two preimages of Q under Π and Z_z is the set of maps $f : [0,1] \to \mathbb{S}^3$ with $f(0) = 1$ and $f(1) = z$. Clearly, Z_Q is homotopically equivalent to $\Omega SO(3) = Z_I$ and each Z_z is homotopically equivalent to $\Omega\mathbb{S}^3 = Z_1$.

Recall that $\pi_n(\Omega\mathbb{S}^3) = \pi_{n+1}(\mathbb{S}^3)$: this implies that each Z_z is connected and simply connected with $\pi_2(Z_z) = \mathbb{Z}$. Also, $H^*(Z_z, \mathbb{R}) = H^*(\Omega\mathbb{S}^3, \mathbb{R}) = \mathbb{R}[\mathbf{x}]$ where $\mathbf{x} \in H^2(\Omega\mathbb{S}^3, \mathbb{R})$ satisfies $\mathbf{x}^n \neq 0$ for all n (see, for instance, [2]).

The Hirsch-Smale theorem proves that $\phi : Y_Q \to Z_Q$ is a homotopy equivalence: this fact admits a direct, simple proof in our special case but we do not discuss it. As a consequence, each Y_z is connected and simply connected, $\pi_2(Y_z) = \mathbb{Z}$ and $H^*(Y_z, \mathbb{R}) = \mathbb{R}[\mathbf{x}]$.

As in the introduction, define $p_Q^k : Y_Q \to Y_Q$ by $p_Q^k(\gamma) = \nu^k * \gamma$. Notice that p_Q^k as defined here is trivially homotopic to the composition of k copies of $p_Q = p_Q^1$, justifying the notation. The fact that $\nu \in Y_{-1}$ implies that $p_Q : Y_Q \to Y_Q$ takes Y_z to Y_{-z} and vice-versa.

Lemma 2.1 *The function $p_Q^2 : Y_z \to Y_z$ is homotopic to the identity.*

Proof: First notice that p_Q^2 is homotopic to $\gamma \mapsto (\nu_\epsilon * \nu_{\pi-\epsilon}) * \gamma$ for any $\epsilon > 0$.

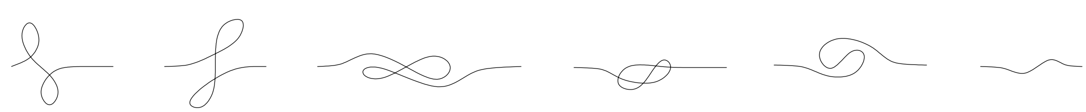

Figure 1: How to eliminate two small loops of opposite curvature.

Figure 1 shows how to move from that to a reparametrization of γ where a time slightly longer than $1/2$ is spent in a short initial segment. The figure shows only the beginning of γ: we do not think a formula is necessary or helpful. ■

3 Convex curves and the connected components of X_Q

Given a parametrized curve of positive geodesic curvature $\gamma : [0,1] \to \mathbb{S}^2$, let $V_\gamma \subseteq \mathbb{R}^3$ be the closure of the set of all positive linear combinations $a_1\gamma(t_1) + \cdots + a_n\gamma(t_n)$, where $a_1, \ldots, a_n$ are positive real numbers and $t_1, \ldots, t_n \in [0,1]$. The set V_γ is either a closed convex cone or $\mathbb{R}^3$. We say that γ is *convex* if γ is simple and the image of γ is contained in the boundary of V_γ.

Let $X_{Q,c} \subset X_Q$ be the set of convex curves $\gamma : [0,1] \to \mathbb{S}^2$ with $\gamma(0) = e_1$, $\gamma'(0) = c_0 e_2$, $\gamma(1) = Qe_1$ and $\gamma'(1) = c_1 Q e_2$ ($c_0, c_1 > 0$). For z with $\Pi(z) = Q$, define $X_z = (X_Q - X_{Q,c}) \cap Y_z$: we shall prove in theorem 7 that each X_z is connected. For $Q = I$, this is proved in [7]. In particular, $\nu \in X_{I,c}$, $\nu^{2n} \in X_1$ and $\nu^{2n+1} \in X_{-1}$ for n a positive integer.

We show how to decide, given Q, whether $X_{Q,c}$ is empty or not. The criterion is harder to state than to prove, so instead of proclaiming a proposition we explain the criterion together with its justification. We split our discussion into three cases: $Qe_1 = e_1$, $Qe_1 = -e_1$ and Qe_1 and e_1 linearly independent.

If $Qe_1 = e_1$, let $\alpha \in (-\pi, \pi]$ be the angle from Qe_2 to e_2. If $\alpha < 0$ then $X_{Q,c} = \emptyset$: the points $\gamma(t)$ fot t near 0 or 1 cannot possibly be in the boundary of V_γ if $\gamma \in X_Q$ (see figure 2, (a): the region within the dashed line must belong to V_γ). Similarly, if $\alpha = \pi$ (see figure 2, (b)). On the other hand, if $0 \le \alpha < \pi$, it is easy to construct $\gamma \in X_{Q,c}$ (see figure 2. (c)).

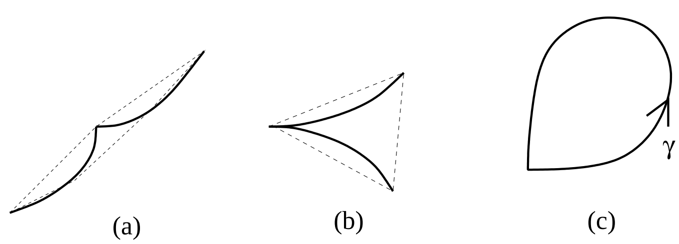

Figure 2: How to decide if $X_{Q,c} = \emptyset$ if $Qe_1 = e_1$.

If $Qe_1 = -e_1$, we always have $X_{Q,c} = \emptyset$. Indeed, if a convex cone contains both e_1 and $-e_1$, it must be bounded by two half-planes and there is no curve of positive geodesic curvature contained the intersection of these halfplanes with the unit sphere.

Finally, consider the case when e_1 and Qe_1 are linearly independent. Draw the shortest geodesic δ from e_1 to Qe_1: notice that δ is contained in V_γ for any $\gamma \in X_Q$. Let v_0 and v_1 be the tangent vectors to δ at e_1 and Qe_1 (see figure 3). Let α_0 (resp. α_1) be the angle from e_2 to v_0 (resp. from v_1 to Qe_2), $\alpha_i \in (-\pi, \pi]$.

If $\alpha_0 \leq 0$ or $\alpha_1 \leq 0$ then $X_{Q,c} = \emptyset$ (see figure 3, (b): the region within the dashed line must belong to V_γ). On the other hand, if $\alpha_0 > 0$ and $\alpha_1 > 0$ then it is easy to construct $\gamma \in X_{Q,c}$: just keep close to δ (see figure 3, (c)).

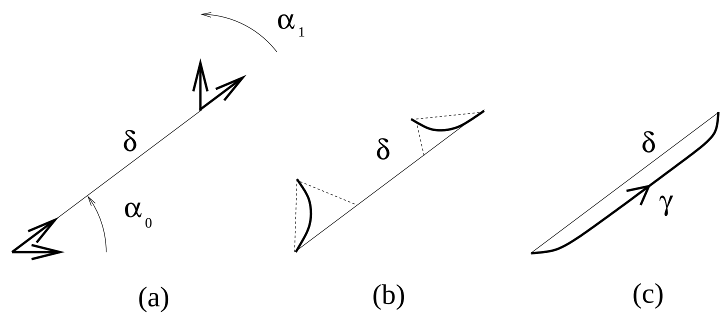

Figure 3: How to decide if $X_{Q,c} = \emptyset$ if e_1 and Qe_1 are linearly independent.

Let $A \subset \mathbb{S}^3$ be the set of quaternions z for which there exists a convex curve in Y_z. The paragraphs above give a description of A. The set A is neither closed nor open and its interior is given by

$$\{a + bi + cj + dk \mid b, d > 0, bd > |ac|\}.$$

The sets A and $-A$ are disjoint, but their closures are not.

Theorem 6 *Given $Q \in SO(3)$, the set $X_{Q,c} \subset X_Q$ is either empty or a contractible connected component of X_Q.*

Proof: Assume that Q is such that $X_{Q,c} \neq \emptyset$. It is not hard to see that both $X_{Q,c}$ and its complement (in X_Q) are open: we must prove that $X_{Q,c}$ is contractible (and in particular that it is connected).

From the discussion above, in all cases there exists $v \in \mathbb{S}^2$, $v \perp e_1$, $v \perp Qe_1$ such that $\langle v, \gamma(t)\rangle > 0$ for all $\gamma \in X_{Q,c}$, $t \in (0,1)$. Indeed, if $Qe_1 = e_1$ we may take $v = e_3$ and if $Qe_1 \neq e_1$ we take v to be one of the vectors perpendicular to the geodesic δ.

Let p be the plane $\{u | \langle v, u\rangle = 1\}$. For each $\gamma \in X_{Q,c}$, use radial projection to define a curve $\hat{\gamma} : (0,1) \to p$ such that $\hat{\gamma}(t)$ is a positive multiple of $\gamma(t)$. This defines a bijection from $X_{Q,c}$ to the set of curves of positive curvature in the plane with prescribed asymptotic behavior when t tends to 0 or 1 ($\gamma(0)$ and $\gamma(1)$ indicate the asymptotic direction and $\gamma'(0)$ and $\gamma'(1)$ indicate the asymptotic line or lack thereof). By putting axes in an appropriate position, the image of $\hat{\gamma}$ is the graph of a convex function from $\mathbb{R}$ to $\mathbb{R}$ (with prescribed asymptotic behavior). The parametrization of the curve does not affect the homotopy type of the space of curves, and we therefore have a homotopy equivalence between $X_{Q,c}$ and a convex space: this proves its contractibility. ∎

4 Construction of $f_z^{[n]} : (\mathbb{S}^2)^n \to X_z$

We now give an explicit generator for $\pi_2(Y_1)$. This function actually has the from $f_1 : \mathbb{S}^2 \to X_I \subset Y_I$ and will play an important part throughout the paper. Let

$$\begin{aligned}
\alpha_0(s,t) &= (\sin s \cos t, \sin s \sin t, \cos s),\\
\alpha_1(s,t) &= (-\sin t, \cos t, 0),\\
\alpha_2(s,t) &= (-\cos s \cos t, -\cos s \sin t, \sin s),\\
g_s(t) &= \frac{\sqrt{2}}{2}\left(\alpha_0(s,t) + \cos 3t\ \alpha_1(s,t) + \sin 3t\ \alpha_2(s,t)\right).
\end{aligned}$$

The curve g_0 is a circle drawn 4 times and the curve g_π is a circle drawn 2 times. A computation verifies that $\det(g_s(t), g_s'(t), g_s''(t)) > 0$ for all s and t. Let $\Gamma_s(t)$ be defined as above. It easy to verify that

$$\Gamma_s(t + (2\pi/3)) = \begin{pmatrix} -1/2 & -\sqrt{3}/2 & 0 \\ \sqrt{3}/2 & -1/2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \Gamma_s(t)$$

for all s and t. Finally, let $f_1 : [0, 2\pi] \times [0, \pi] \to X_I$ be defined by

$$f_1(s_1, s_2)(t) = (\Gamma_{s_2}(s_1/3))^{-1} \Gamma_{s_2}(t + (s_1/3)) e_1, \quad s_2 \in [0, \pi]$$

If $s_2 = 0$ or π, the value of s_1 is irrelevant for the value of f_1: actually, $f_1(s_1, 0) = \nu^4$ and $f_1(s_1, \pi) = \nu^2$. Also, the remark above shows that $f_1(0, s_2) = f_1(2\pi, s_2)$ for all s_2. Performing these identifications, the domain of f_1 becomes the sphere $\mathbb{S}^2$.

If we follow the identification of $\pi_2(Y_I)$ with $\pi_3(SO(3))$ described in the previous section, we see that in order to verify that f_1 is indeed a generator of $\pi_2(Y_{I,+})$ we have to compute the topological degree of the function $\tilde{f}_1 : \mathbb{S}^2 \times \mathbb{S}^1 \to \mathbb{S}^3$ which is the double cover of $\hat{f}_1 : \mathbb{S}^2 \times \mathbb{S}^1 \to SO(3)$ given by

$$\hat{f}_1(s,t) = \begin{pmatrix} f_1(s)(t) & (\widehat{f_1(s))'}(t) & f_1(s)(t) \times (\widehat{f_1(s))'}(t) \end{pmatrix} :$$

the absolute value of the degree of $\tilde{f}_1$ is 1, confirming that it is a generator. In order to see that, it suffices to verify that $j \in \mathbb{S}^3$ is a regular value with a single preimage under $\tilde{f}_1$; we skip the details.

Figure 4 should give a rough idea of the image of a line $(s_1, \cdot)$ under f_1. The fat dot in the figure is e_1; between the second and third images most of the curve went around the sphere. Figure 5 shows the image of a line $(\cdot, s_2)$: the first and last curves are intentionally equal: this is a closed curve in X_I.

More generally, we construct, for each $z \in -A$, a function $f_z : \mathbb{S}^2 \to X_z$ which is a generator of $\pi_2(Y_z)$. Here, instead of a formula, we indicate the construction

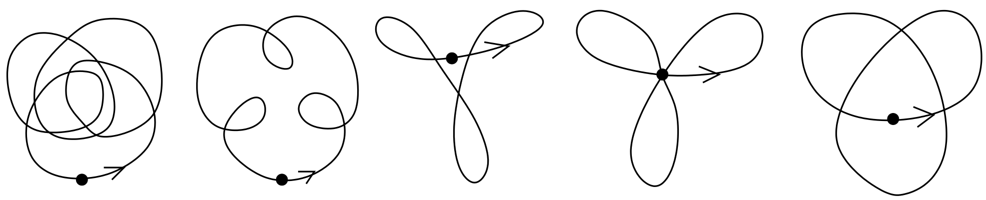

Figure 4: The image of a line under a generator of $\pi_2(Y_{I,+})$.

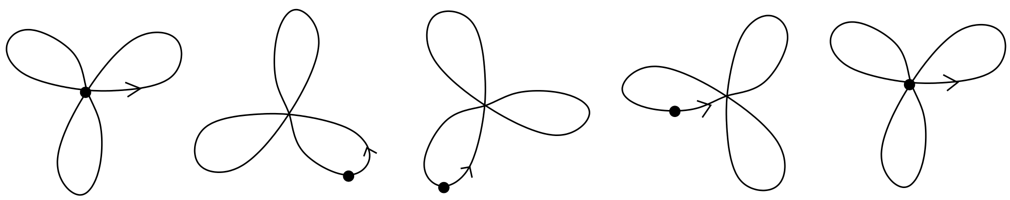

Figure 5: The image of a circle under a generator of $\pi_2(Y_{I,+})$.

in figure 19. The dashed line is an arbitrary curve $\delta : [0,1] \to \mathbb{S}^2$ with $\delta(0) = Qe_1$ ($\Pi z = Q$) and $\delta(1) = e_1$ such that $\gamma * \delta$ is convex for $\gamma \in X_{-z,c}$. The top and bottom lines are adjacent, forming a cylinder. The curves shown are all contained in a relatively small portion of the sphere and in the transition from the third to the fourth column, most of the curve passed around the back of the sphere. The way to define f_z in each of the 24 squares in this grid should be visually obvious (perhaps with a little effort). The octagons on the right and left can likewise be filled in a natural way, having in mind that the center of each octagon is approximately a circle drawn two or four times. The function f_1 constructed above is a special case of f_z.

Finally, for a positive integer n and $z = (-1)^n z'$, $z' \in A$, let $\gamma_0 \in X_{z',c}$ be fixed but arbitrary, $\tilde{\Gamma}_0 : [0,1] \to \mathbb{S}^3 \subset \mathbb{H}$ be the associated function. For $i = 1, 2, \ldots, n$ set $z'_i = (\tilde{\Gamma}_0((i-1)/n))^{-1}\tilde{\Gamma}_0(i/n)$ and $z_i = -z'_i$ so that $z' = z'_1 z'_2 \cdots z'_n$ and $z = z_1 z_2 \cdots z_n$. Notice that $z'_i \in A$. Set $Q_i = \Pi(z_i) = \Pi(z'_i)$ and define $f_z^{[n]} : (\mathbb{S}^2)^n \to X_z$ by

$$f_z^{[n]}(s_1, s_2, \ldots, s_n) = f_{z_1}(s_1) * (Q_1 f_{z_2}(s_2)) * \cdots * (Q_1 Q_2 \cdots Q_{n-1} f_{z_n}(s_n)).$$

Here we are stretching a bit our original definition of $*$: if $\gamma_1, \gamma_2 : [0,1] \to \mathbb{S}^2$ satisfy $\gamma_1(1) = \gamma_2(0)$ then

$$\gamma_1 * \gamma_2(t) = \begin{cases} \gamma_1(2t), & t \le 1/2, \\ \gamma_2(2t-1), & t \ge 1/2. \end{cases}$$

Also, for $\gamma \in X$ and $Q \in SO(3)$, $Q\gamma$ is the function from $[0,1]$ to $\mathbb{S}^2$ defined by $(Q\gamma)(t) = Q(\gamma(t))$.

The following lemma is a simple consequence of the existence of f_1.

Lemma 4.1 *Let $n_1, n_2 > 1$ and $\theta_1, \theta_2 \in (0, \pi/2)$. Then $\nu_{\theta_1}^{n_1}$ and $\nu_{\theta_2}^{n_2}$ are in the same connected component of X_I if and only if n_1 and n_2 have the same parity.*

Proof: First notice that if n_1 and n_2 have different parities then $\nu_{\theta_1}^{n_1}$ and $\nu_{\theta_2}^{n_2}$ are in different connected components of Y_I and therefore with stronger reason in different connected components of X_I.

Our function f_1 shows that ν^2 and ν^4 are in the same connected component; it follows from that that, for any $n > 0$, $\nu^{2+n} \sim \nu^n * \nu^2$ and $\nu^{4+n} \sim \nu^n * \nu^4$ are also in the same connected component. The value of θ can be changed continuously and is therefore not a problem. The result follows. ∎

Recall that it follows from the results of the previous section that $\nu \in X_{-1,c}$ is not in the same connected component of X_I as $\nu^3 \in X_{-1}$.

5 Proof of theorem 1

For $\gamma \in Y_Q$ and corresponding $\Gamma : [0, 1] \to SO(3)$, define $F_{n,\theta}(\gamma)(t) = \Gamma(t)\nu_\theta^{2n}(t)$. Intuitively, for small values of θ, $F_{n,\theta}(\gamma)$ is obtained from γ by attaching $2n$ positively oriented small loops along γ (see figure 6).

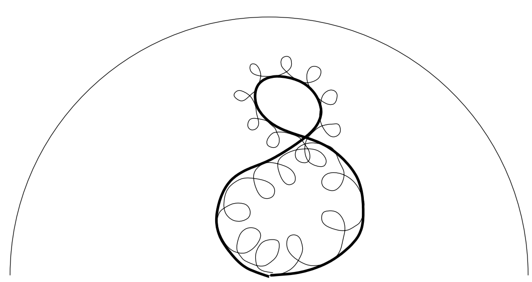

Figure 6: A curve γ (thicker) and $F_{9,\theta}(\gamma)$.

Lemma 5.1 *Let $\theta \in (0, \pi/2)$, let K be a compact set and let $f : K \to Y_Q$ a continuous function. Then, for sufficiently large n, $F_{n,\theta} \circ f$ is a function from K to X_Q.*

Proof: Let $C > 1$ be a constant such that $|\Gamma'(t)| < C$ and $|\Gamma''(t)| < C$ for any $\gamma = f(k)$, $k \in K$. Let $\epsilon > 0$ be such that if $|v_1 - \nu_\theta'(0)| < \epsilon$ and $|v_2 - \nu_\theta''(0)| < \epsilon$ then $\det(\nu_\theta(0), v_1, v_2) > 0$. Notice that this implies that if $|v_1 - \nu_\theta'(t)| < \epsilon$ and $|v_2 - \nu_\theta''(t)| < \epsilon$ then $\det(\nu_\theta(t), v_1, v_2) > 0$. Take $n > 20C/\epsilon$.

For $\gamma = f(k)$, write

$$\tilde{\gamma}(t) = (F_{n,\theta}\gamma)(t) = \Gamma(t)\nu_\theta^{2n}(t) = \Gamma(t)\nu_\theta(2nt)$$

so that

$$\begin{aligned}\tilde{\gamma}'(t) &= \Gamma'(t)\nu_\theta(2nt) + 2n\Gamma(t)\nu_\theta'(2nt)\\ \tilde{\gamma}''(t) &= \Gamma''(t)\nu_\theta(2nt) + 4n\Gamma'(t)\nu_\theta'(2nt) + 4n^2\Gamma(t)\nu_\theta''(2nt)\end{aligned}$$

and therefore, after a few manipulations,

$$\left|\frac{\tilde{\gamma}'(t)}{2n} - \Gamma(t)\nu_\theta'(2nt)\right| < \epsilon, \quad \left|\frac{\tilde{\gamma}''(t)}{4n^2} - \Gamma(t)\nu_\theta''(2nt)\right| < \epsilon$$

or, equivalently,

$$\left|\frac{(\Gamma(t))^{-1}\tilde{\gamma}'(t)}{2n} - \nu_\theta'(2nt)\right| < \epsilon, \quad \left|\frac{(\Gamma(t))^{-1}\tilde{\gamma}''(t)}{4n^2} - \nu_\theta''(2nt)\right| < \epsilon.$$

It follows that

$$\det\left(\nu_\theta(2nt), \frac{(\Gamma(t))^{-1}\tilde{\gamma}'(t)}{2n}, \frac{(\Gamma(t))^{-1}\tilde{\gamma}''(t)}{4n^2}\right) > 0$$

and therefore that $\det(\tilde{\gamma}(t), \tilde{\gamma}'(t), \tilde{\gamma}''(t)) > 0$, which is what we needed. ■

Theorem 1 now follows directly from the next lemma.

Lemma 5.2 *Let $\theta \in (0, \pi/2)$, let K be a compact set, $f : K \to Y_Q$. Then, for sufficiently large n, the image of $F_{n,\theta} \circ f$ is contained in X_Q and there exists $H : [0,1] \times K \to Y_Q$ such that $H(0, \cdot) = f$ and $H(1, \cdot) = F_{n,\theta} \circ f$.*

Proof: We know from lemma 2.1 that f if homotopic to f_1, $f_1(k) = \nu_\theta^{2n} * f(k)$ so all we have to do is construct a homotopy between $F_{n,\theta} \circ f$ and f_1. Intuitively, this is done by pushing the loops towards $t = 0$. More precisely, if $\gamma = f(k)$, $k \in K$, let

$$H_1(s,k)(t) = \begin{cases} \nu_s^{2n}(t), & t \le s/2,\\ \Gamma((2t-s)/(2-s))\nu_s^{2n}(t), & t \ge s/2\end{cases}$$

and

$$H_2(s,k)(t) = \begin{cases} \nu_s^{2n}((2t)/(2-s)), & t \le 1/2,\\ \Gamma(2t-1)\nu_s^{2n}((2t)/(2-s)), & 1/2 \le t \le 1-s/2,\\ \gamma(2t-1), & t \ge 1-s/2.\end{cases}$$

Estimates similar to those of the proof of lemma 5.1 guarantee that $H_1(s,k) \in X_Q$ and $H_2(s,k) \in Y_Q$ for sufficiently large n. ■

6 Proof of theorem 3 and construction of $w_{Q,K}$

Lemma 6.1 *Let $\theta \in (0,\pi)$. Let K be a compact space and $f : K \to Y_Q$ a continuous map. Then, for sufficiently large n, if $n_1, n_2 \geq n$ then $F_{n_1+n_2,\theta} \circ f$ and $\nu_\theta^{2n_1} * (F_{n_2,\theta} \circ f)$ are homotopic in the space of functions from K to X_Q.*

Proof: This is a pushing-the-loops argument similar to what was done in the proof of lemma 5.2. More precisely, for each $\gamma = f(k)$ first go from $F_{n_1+n_2,\theta}(\gamma) = \Gamma \cdot \nu_\theta^{2(n_1+n_2)}$ to $\Gamma \cdot (\nu_\theta^{2n_1} * \nu_\theta^{2n_2})$: this involves a reparametrization and the estimates in the proof of lemma 5.1 guarantee that we remain inside X_Q for sufficiently large n_1 and n_2. Next do

$$H(s,k)(t) = \begin{cases} (\nu_\theta^{2n_1} * \nu_\theta^{2n_2})(t), & t \leq s/2, \\ \Gamma((2t-s)/(2-s))(\nu_\theta^{2n_1} * \nu_\theta^{2n_2})(t), & t \geq s/2. \end{cases}$$

Again, the estimates in lemma 5.1 show that we remain inside X_Q throughout the process. ■

The function $w_{Q,K} : [K, Y_Q] \to [K, X_Q]$ is defined to take $f : K \to Y_Q$ to $F_{n,\theta} \circ f : K \to X_Q$, where θ is arbitrary and n is taken to be sufficiently large. The following lemma shows that $w_{Q,K}$ is well defined.

Lemma 6.2 *Let $\theta_1, \theta_2 \in (0, \pi/2)$. Let K be a compact set and $f : K \to Y_Q$ a continuous function. Then, there exists N such that, if $n_1, n_2 > N$ then the functions $F_{n_1,\theta_1} \circ f$ and $F_{n_2,\theta_2} \circ f$ have images contained in X_Q and are homotopic in the class of functions from K to X_Q.*

Proof: Take $n = \lfloor \min(n_1, n_2)/2 \rfloor$. Use lemma 6.1 to obtain homotopies from $F_{n_i,\theta_i} \circ f$ to $\nu_{\theta_i}^{2(n_i-n)} * (F_{n,\theta_i} \circ f)$. Lemma 4.1 gives us a homotopy from $\nu_{\theta_1}^{2(n_1-n)}$ to $\nu_{\theta_2}^{2(n_2-n)}$. A homotopy from $F_{n,\theta_1} \circ f$ to $F_{n,\theta_2} \circ f$ is obtained just by changing the value of θ, finishing the proof. ■

We next show that if $f : K \to X_Q$ then $w_{Q,K} f = p_Q^2 f = \nu^2 * f$ (the equalities here being in $[K, X_Q]$: homotopies and not equalities as funcions).

Lemma 6.3 *Let $\theta \in (0,\pi)$. Let K be a compact space and $f : K \to X_Q$ a continuous map. Then, for sufficiently large n, $F_{n,\theta} \circ f$ and $\nu^2 * f$ are homotopic in the space of functions from K to X_Q.*

Proof: Write $n = n_1 + n_2$. From lemma 6.1, $F_{n,\theta} \circ f$ is homotopic to $\nu_\theta^{2n_1} * (F_{n_2,\theta} \circ f)$. From lemma 4.1, that is homotopic to $\nu^2 * (F_{n_2,\theta} \circ f)$. We now push the loops from the second interval to the first. More precisely, for any $k \in K$, set $\gamma = \nu^2 * f(k)$ and corresponding Γ so that

$$(\nu^2 * (F_{n_2,\theta} \circ f))(t) = \begin{cases} \Gamma(t) e_1, & t \leq 1/2, \\ \Gamma(t) \nu_\theta^{2n_2}(2t-1), & t \geq 1/2. \end{cases}$$

Set

$$H(s,k)(t) = \begin{cases} \Gamma(t)e_1, & t \leq (1-s)/2, \\ \Gamma(t)\nu_\theta^{2n_2}(2t-1), & (1-s)/2 \leq t \leq (2-s)/2, \\ \Gamma(t)e_1, & t \geq (2-s)/2. \end{cases}$$

This is a homotopy from $\nu^2 * (F_{n_2,\theta} \circ f)$ to $(F_{n_2,\theta} \circ \nu^2) * f$: at any point the curvature is positive either because n_2 is large, using estimates similar to those of 5.1, or because our curve is simply $\nu^2 * f(k)$, which is known to be in X_Q. Finally, changing the value of θ yields a homotopy from $F_{n_2,\theta} \circ \nu^2$ to $F_{n_2,\pi/4} \circ \nu^2 = \nu^{2(1+n_2)}$ and lemma 4.1 gives us a homotopy from that to ν^2. ■

Proof of theorem 3: The fact that p_Q^2 and p_Q^4 are homotopic follows from lemma 4.1. Given $H : \mathbb{B}^{n+1} \to Y_Q$, define $\tilde{H}(s) = F_{n,\theta}(H(2s))$ for $|s| \leq 1/2$ where, as usual, $\theta \in (0, \pi/2)$ and n is sufficiently large. Use now lemma 6.3 for $K = \mathbb{S}^n$ to define $\tilde{H}$ for $1/2 < |s| < 1$. ■

7 Reidemeister moves and the connectivity of X_z

A *double point* of a curve $\gamma \in Y_I$ is a pair (t_0, t_1), $0 \leq t_0 < t_1 < 1$, with $\gamma(t_0) = \gamma(t_1)$. Similarly, a *triple point* is a triple (t_0, t_1, t_2), $0 \leq t_0 < t_1 < t_2 < 1$, such that $\gamma(t_0) = \gamma(t_1) = \gamma(t_2)$. A double point (t_0, t_1) is a *self-tangency* if $\gamma'(t_0)$ and $\gamma'(t_1)$ are parallel; otherwise the double point is *transversal*. We call a curve *generic* if it has neither triple points nor self-tangencies and define $Y_I^{(0)} \subset Y_I$ to be the set of generic curves. For $z = \pm 1$, set also $Y_z^{(0)} = Y_I^{(0)} \cap Y_z$, $X_I^{(0)} = X_I \cap Y_I^{(0)}$ and $X_z^{(0)} = X_I^{(0)} \cap X_z$. It is clear that $Y_z^{(0)}$ and $X_z^{(0)}$ are open and dense in Y_z and X_z, respectively. Also, $Y_z^{(0)}$ and $X_z^{(0)}$ are disconnected since the number of double points does not change in a connected component of these sets. In other words, the complement of $Y^{(0)}$ has codimension 1.

A triple point (t_0, t_1, t_2) is *generic* if the vectors $\gamma'(t_0)$, $\gamma'(t_1)$ and $\gamma'(t_2)$ are two by two linearly independent. A self-tangency (t_0, t_1) is *generic* if the curvatures of γ at t_0 and t_1 are distinct. A curve $\gamma \in Y_I - Y_I^{(0)}$ belongs to $Y_I^{(1)}$ if it has *either* a unique generic triple point *or* a unique generic self-tangency (but not both). The complement of $Y_I^{(0)} \cup Y_I^{(1)} \subset Y_I$ has codimension 2. We define $X_I^{(1)}$, $Y_z^{(1)}$ and $X_z^{(1)}$ in the obvious way.

The passage from one connected component of $Y^{(0)}$ to another through an element of $Y^{(1)}$ is a Reidemeister move ([3]). Reidemeister moves of type I are not allowed in Y_I; Reidemeister moves of types II and III correspond to generic self-tangencies and generic triple points, respectively. If the curve is in $X_I \subset Y_I$, one of the possibilities for orientations near a generic self-tangency is ruled out. Figure 7 shows the possible Reidemeister moves, or, equivalently, shows the neighborhood of a generic self-tangency or generic triple point.

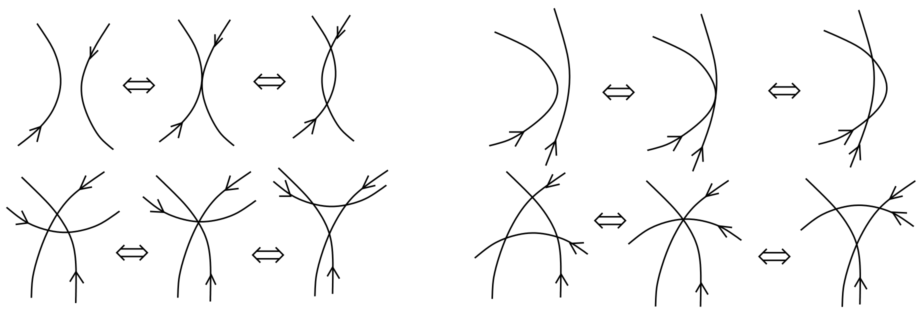

Figure 7: Reidemeister moves: type II on first line, type III on second line.

An *arc* of a curve $\gamma \in X_I$ is a pair (t_-, t_+) such that either (t_-, t_+) or (t_+, t_-) is a transversal double point. Intuitively, we think of the arc as the restriction of γ to $[t_-, t_+]$ or $[0, t_+] \cup [t_-, 1]$. An arc is *positive* (resp. *negative*) if $\det(\gamma(t_-), \gamma'(t_+), \gamma'(t_-)) > 0$ (resp. < 0). An arc is *simple* if one of the following two conditions holds:

1. $0 \leq t_- < t_+ < 1$ and the restriction of γ to $[t_-, t_+)$ is injective;
2. $0 \leq t_+ < t_- < 1$ and the restriction of γ to $[0, t_+) \cup [t_-, 1)$ is injective.

Figure 8 shows examples of simple arcs.

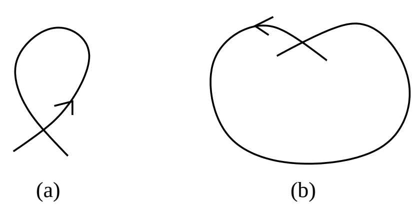

Figure 8: A simple positive arc and a simple negative arc.

The function H constructed in the following lemma is the fundamental building block in the proof that the spaces $X_{\pm 1}$ are connected and simply connected.

Lemma 7.1 *Let (t_-, t_+) be a simple positive arc of $\gamma_0 \in X_I$. Then there exists an open neighborhood V of γ_0 and continuous functions $t_-, t_+ : V \to \mathbb{S}^1$ such that $(t_-(\gamma), t_+(\gamma))$ is a simple positive arc of γ for all $\gamma \in V$. Furthermore, there exists $H : [0, 1] \times V \to X_I$ with $H(0, \gamma) = \gamma$ and $H(1, \gamma) = \nu^2 * \gamma$ for all γ.*

Proof: First assume $0 < t_- < t_+ < 1$. In this case, $H(s,\gamma)$ coincides with γ outside $(t_- - \epsilon, t_+ + \epsilon)$ and $H(s,\gamma)(t)$ for $t_- - \epsilon < t < t_+ + \epsilon$ is indicated in figure 9. Let us follow the process: the arc is first shrunk (a), then pushed along a geodesic all the way, until it comes back (b). This creates two chunks of curve which are very nearly geodesics: these two chunks are then shrunk (c), obtaining two new positive simple arcs which can be deformed so that we have a copy of ν_θ^2 (for small θ) somewhere in the middle of the curve (d). Finally, that copy of ν_θ^2 can be pushed back to $t = 0$, proving the lemma in this case.

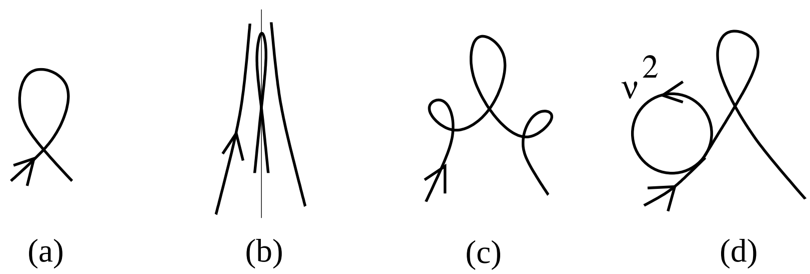

Figure 9: How to create a copy of ν_θ^2 in a curve which has a small loop.

If $t_- = 0$ or $t_+ < t_-$, we forget about the base point and perform the construction above to obtain $\tilde{H}(s,\gamma) : [0,1] \to \mathbb{S}^2$. Define $Q(s,\gamma)$ to be the matrix in $SO(3)$ with first two columns equal to $\tilde{H}(s,\gamma)(0)$ and $\widehat{\tilde{H}(s,\gamma)'(0)}$ and set $H(s,\gamma) = (Q(s,\gamma))^{-1}\tilde{H}(s,\gamma)$. ∎

Lemma 7.2 *Let (t_-, t_+) be a simple negative arc of $\gamma_0 \in X_I$. Then there exists an open neighborhood V of γ_0 and continuous functions $t_-, t_+ : V \to \mathbb{S}^1$ such that $(t_-(\gamma), t_+(\gamma))$ is a simple negative arc of γ for all $\gamma \in V$. Furthermore, there exists $H : [0,1] \times V \to X_I$ with $H(0,\gamma) = \gamma$ and $H(1,\gamma) = \nu^2 * \gamma$ for all γ.*

Proof: Shrink the negative simple arc as in figure 10. This creates a simple positive arc: apply lemma 7.1 and then unshrink. ∎

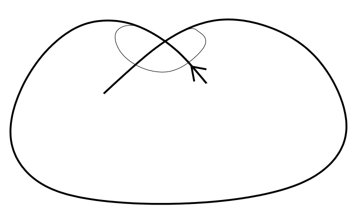

Figure 10: How to create a simple positive arc in a curve with a simple negative arc.

In the next result we become more general and consider again curves in X_z for $z \neq \pm 1$.

Lemma 7.3 *Let $\gamma \in X_Q - X_{Q,c}$. Then there exists $\gamma_1 \in X_Q$ and a path in X_Q from γ to $\nu^2 * \gamma_1$.*

Proof: Assume without loss of generality that γ is generic, i.e., that all self intersections (if any such exist) are transversal double (not triple) points in the interior (i.e., not in the endpoints). We may furthermore assume that $\theta(t) \neq \pm e_1$ for $t \in (0,1)$.

If γ is not injective, there exists a simple arc and we use the construction in lemma 7.1 or 7.2. If γ is simple but not convex, let $t_1 \in (0,1)$ be the largest number for which the restriction of γ to $[0,t_1]$ is convex. We want to prove that this can happen in the two ways illustrated if figure 11 (a) and (b). Let $\theta(t)$ be the argument of the vector obtained from the second and third coordinates of $\gamma(t)$: thus $\theta(t)$ is increasing for small t and its limit when t tends to 0 is 0. As long as $\theta'(0) \geq 0$ and $\theta(t) \leq \pi$, γ restricted to $[0,t]$ is convex: indeed, the image of this interval under γ is a graph of a function: one value of x (the e_1 coordinate) for each argument between 0 and $\theta(t)$. In this case $V_{\gamma|_{[0,t]}} \cap \mathbb{S}^2$ is the region "under" this graph (see figure 11 (c)): since the boundary is a locally convex simple closed curve this set is indeed convex.

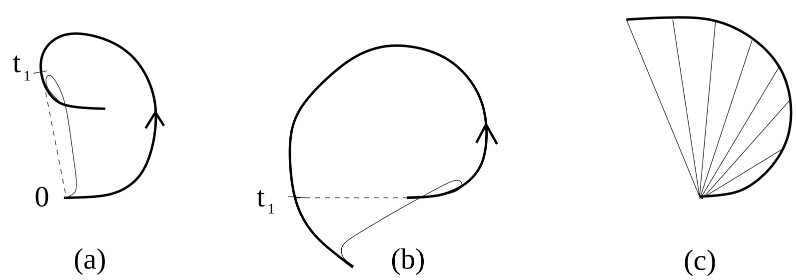

Figure 11: How to create a self-intersection in a non-convex curve.

Thus t_1 is the first number for which either $\theta'(t) = 0$ (case (a)) or $\theta(t) = \pi$ (case (b)). In either case it is easy to deform γ in the interval $[0, t_1+\epsilon]$ as indicated in figure 11. The new (thinner) curve is very near a geodesic from e_1 to $\gamma(t_1+\epsilon)$. In either case we create self-intersections, reducing the problem to the previous case. ■

Theorem 7 *For any $z \in \mathbb{S}^3$ the set X_z is connected.*

Proof: Let $\gamma_1, \gamma_{-1} \in X_z$. From lemma 7.3, each γ_i is in the same connected component as some $\nu^2 * \tilde{\gamma}_i$: we must prove that $\nu^2 * \tilde{\gamma}_1$ and $\nu^2 * \tilde{\gamma}_{-1}$ are in the same connected component of X_z. Since Y_z is connected, there exists $h : \mathbb{B}^1 = [-1,1] \to Y_{Q,+}$ with $h(-1) = \tilde{\gamma}_{-1}$, $h(1) = \tilde{\gamma}_1$. From theorem 3, there exists $\tilde{h} : [-1,1] \to X_Q$ with $\tilde{h}(-1) = \nu^2 * \tilde{\gamma}_{-1}$, $\tilde{h}(1) = \nu^2 * \tilde{\gamma}_1$. ■

The construction in the proof of lemma 7.3 is not at all uniform. This is not something which can be fixed with a more careful argument: a uniform construction would prove the inclusions $X_z \subset Y_z$ to be homotopy equivelences. In the following sections we show that this is not the case.

8 Proof of theorem 2

In this section, let $z = \pm 1$.

Lemma 8.1 *Let $z = \pm 1$. All generic curves $\gamma \in X_z$ have simple arcs. Furthermore, if $\gamma_0 \in X_z^{(1)}$ (a Reidemeister move) then there exists an open neighborhood $V \subset X_z$ of γ_0 and continuous functions $t_-, t_+ : V \to \mathbb{S}^1$ such that $(t_-(\gamma), t_+(\gamma))$ is a simple arc for all $\gamma \in V$.*

Proof: We define a *generalized arc* to be a pair (t_-, t_+) where $t_- \neq t_+$ and $\gamma(t_-) = \gamma(t_+)$. We identify the generalized arc (t_-, t_+) with the interval $[t_-, t_+]$ or with the set $[0, t_+] \cup [t_-, 1]$, depending on whether $t_- < t_+$ or $t_+ < t_-$ and we order generalized arcs by inclusion.

Any curve in $\gamma \in X_z$ has self intersections. If γ is generic or a Reidemeister move of type III then all generalized arcs are arcs and their number is finite; in particular, there exists a minimal arc with respect to inclusion. A minimal arc is clearly a simple arc.

Assume now that γ is a Reidemeister move of type II with self-tangency (t_0, t_1). Consider the two generalized arcs (t_0, t_1) and (t_1, t_0). If either is non minimal then it contains a minimal arc and we are done. If both are simple, they must intersect each other, otherwise we would have a Reidemeister move from a simple closed curve in $X_{-1,c}$ to a curve in X_{-1}, contradicting theorem 6. Let (t_2, t_3) be this intersection: it contains neither (t_0, t_1) nor (t_1, t_0). A minimal generalized arc contained in (t_2, t_3) is therefore a simple arc.

In any of these cases the neighborhood V and the functions $t_-, t_+ : V \to \mathbb{S}^1$ are constructed exactly as in lemma 7.1 or lemma 7.2. ■

Lemma 8.2 *Let $z = \pm 1$. Let $\gamma \in X_z$ be a generic curve with two simple arcs $(t_{a,-}, t_{a,+})$ and $(t_{b,-}, t_{b,+})$. Let $\delta_a, \delta_b : [0,1] \to X_z$ be the paths $\delta(s) = H(s, \gamma)$ constructed in lemma 7.1 or 7.2, so that $\delta_a(0) = \delta_b(0) = \gamma$ and $\delta_a(1) = \delta_b(1) = \nu^2 * \gamma$. Then the two paths δ_a and δ_b are homotopic with fixed endpoints in X_z.*

Proof: Notice that the statement includes the case $t_{a,\pm} = t_{b,\pm}$. This is not quite trivial because there is an ambiguity at the end of the construction of H in lemma 7.1: we did not specify which way the copy of ν^2 would roll back the base point. Thus, we have to prove that the map $\delta_\gamma : \mathbb{S}^1 \to X_z$ taking s_1 to γ

with a copy of ν_θ^2 attached at the point $\gamma(s_1)$ is homotopic to a point. If θ is small enough, the copy of ν_θ^2 can be attached to any point of $\delta(s)$, for any s, thus proving that δ_γ is homotopic to $\nu^2 * \delta_\gamma$. Since Y_z is simply connected, δ_γ is homotopic to a point in Y_z; from theorem 3, $\nu^2 * \delta_\gamma$ is homotopic to a point in X_z, proving the lemma in this special case.

We next consider the case when a third simple arc $(t_{c,-}, t_{c,+})$ exists which is disjoint from the first two. Since the arcs $(t_{a,-}, t_{a,+})$ (resp. $(t_{b,-}, t_{b,+})$) and $(t_{c,-}, t_{c,+})$ are disjoint, we may perform the construction in lemma 7.1 or 7.2 independently, thus defining $\delta_{ac} : [0,1]^2 \to X_z$ (resp. $\delta_{bc} : [0,1]^2 to X_z$) with $\delta_{ac}(s,0) = \delta_a(s)$ (resp. $\delta_{bc}(s,0) = \delta_b(s)$) and $\delta_{ac}(s,1) = \nu^2 * \delta_a(s)$ (resp. $\delta_{bc}(s,1) = \nu^2 * \delta_b(s)$). We also have $\delta_{ac}(0,s_1) = \delta_{bc}(0,s_1)$ and $\delta_{ac}(1,s_1) = \delta_{bc}(1,s_1)$: thus, δ_a and δ_b are homotopic with fixed endpoints if and only if $\nu^2 * \delta_a$ and $\nu^2 * \delta_b$ are, and this again follows from the simple connectivity of Y_z and theorem 3.

Next we consider the case when the two arcs $(t_{a,-}, t_{a,+})$ and $(t_{b,-}, t_{b,+})$ are non-disjoint positive arcs. In this case both δ_a and δ_b begin by performing a Reidemeister move of type III: this makes the two arcs disjoint and guarantees the existence of a third disjoint arc, thus reducing the problem to the previous case, as shown in figure 12.

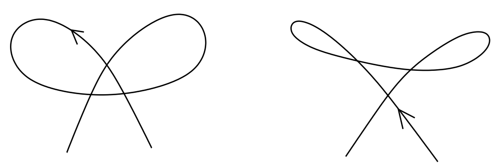

Figure 12: Two non-disjoint positive arcs become disjoint

If one of the two initial arcs is negative, the construction in lemma 7.2 creates a positive arc *and* a second spare positive arc, thus again reducing the problem to the previous cases.

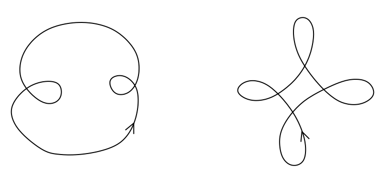

Figure 13: If there are exactly two arcs then the argument works

We are left with only one situation to consider: the two arcs $(t_{a,-}, t_{a,+})$ and $(t_{b,-}, t_{b,+})$ are positive, disjoint, and are the *only* arcs in γ. This guarantees that, up to a deformation, γ is the curve in figure 13. Again, both δ_a and δ_b begin by parforming a sequence of Reidemeister moves with the same result (up to deformation), shown in figure 13. This has four disjoint positive arcs, again reducing to previous cases. ■

Lemma 8.3 *Let $h : \mathbb{S}^1 \to X_z$ be a continuous function. There exists a continuous function $H : [0,1] \times \mathbb{S}^1 \to X_z$ with $H(0,s) = h(s)$ and $H(1,s) = \nu^2 * (h(s))$ for all s.*

Proof: We may assume without loss of generality that $h(s)$ is generic for all but a finite number of values of $s = s_1, \ldots, s_N$ and that these are Reidemeister moves. Cover $\mathbb{S}^1$ by small open sets whose image under h is contained in an open neighborhood V as in lemma 7.1 or 7.2: these two results allow us to contruct H except for thin neighborhoods of finitely many transition points from one arc to another. These transition points may be assumed not to be Reidemeister moves. Lemma 8.2 now guarantees that these holes can be plugged. ■

Proof of theorem 2: Take $h : \mathbb{S}^1 \to X_z$. Since Y_z is simply connected, h is homotopic to a point in Y_z. From theorem 3, $p_z^2 \circ h$ is homotopic to a point in X_z. From lemma 8.3, h is homotopic to $p_z^2 \circ h$ in X_z. Thus, h is homotopic to a point in X_z. ■

9 Stars, trefoils and the proof of theorem 4

A *star* is a curve γ in the same connected component of $X_1^{(0)}$ as one of the infinite family of curves given in figure 14. More precisely, a star has $2k+1$ double points; if $k > 0$, their images in the sphere are the vertices of a convex polygon and, for any pair of adjacent vertices, there are two arcs of γ joining them.

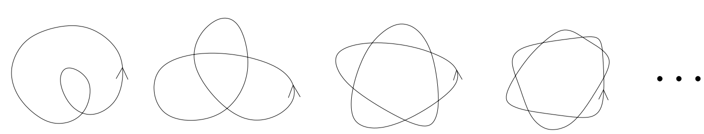

Figure 14: Stars ($k = 0, 1, 2, 3, \ldots$).

Let T^0 be the closure (in X_1) of the set of stars and let T^1 be its boundary.

A curve $\gamma \in X_1$ is called a *trefoil* if:

1. γ has a generic triple point (t_0, t_1, t_2);

2. γ has no self-tangencies and no double points besides (t_0, t_1), (t_0, t_2) and (t_1, t_2);

3. the restriction of γ to each of $[t_0, t_1]$, $[t_1, t_2]$ and $[t_2, 1 + t_0]$ is convex.

The restriction to $[t_2, 1 + t_0]$ is defined by $\gamma(t + 1) = \gamma(t)$. The fourth curve in figure 4 and all the curves in figure 5 are trefoils.

Lemma 9.1 *The set T^1 is the set of trefoils and is a manifold of codimension 1.*

Proof: We have to show that the only Reidemeister moves from a star to a generic γ which is not a star pass through a trefoil. In order to do this, we classify all possible Reidemeister moves starting at a star. Figure 15 shows how a Reidemeister move of type II takes a star to another star (changing the value of k) and how a Reidemeister move of type III takes a star ($k = 1$) to a generic curve which is not a star passing through a trefoil. We prove that these are the only possible moves.

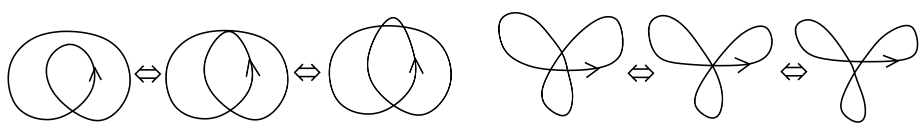

Figure 15: Reidemeister moves starting at a star.

The only possible star from which a Reidemeister move of type III is possible is the one shown in figure 15 ($k = 1$): indeed, a Reidemeister move of type III is quite impossible if the curve does not form a combinatorial triangle. In order to see that the only possible Reidemeister moves of type II are those indicated in figure 15, notice that if γ is a star, its image is trapped in the union of triangles shown in figure 16 (where straight lines indicate geodesics in the sphere). ■

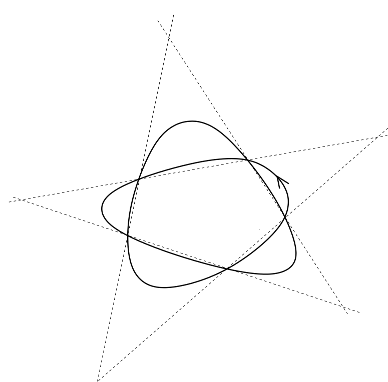

Figure 16: A star is trapped in a union of triangles.

Let $\tilde{T}^1 \subset X_1$ be an open tubular neighborhood of T^1 and $\Pi_{T^1} : \tilde{T}^1 \to T^1$ a projection onto T^1. We know from lemma 9.1 that $X_1 - \tilde{T}^1$ has two connected components: $T^0 - \tilde{T}^1$ and $X_1 - (T^0 \cup \tilde{T}^1)$. Let $g_{1,a} : X_1 \to [0,\pi]$ be a continuous function with $g_{1,a}(\gamma) = 0$ (resp. π) for $\gamma \in T^0 - \tilde{T}^1$ (resp. $X_1 - (T^0 \cup \tilde{T}^1)$). We may assume without loss of generality that $g_{1,a}(\gamma) \leq \pi/2$ if and only if $\gamma \in T^0$. Finally, let $g_1 : X_1 \to \mathbb{S}^2$ be defined by

$$g_1(\gamma) = (\sin(g_{1,a}(\gamma))\cos\theta, \sin(g_{1,a}(\gamma))\sin\theta, \cos(g_{1,a}(\gamma))$$

where, for $\gamma \in \tilde{T}^1$, we have $\theta = 2\pi t_0/(1 + t_0 - t_2)$, (t_0, t_1, t_2) being the triple point of the trefoil $\Pi_{T^1}(\gamma)$. For $\gamma \notin \tilde{T}^1$, θ is undefined but this does not affect the definition of g_1.

Lemma 9.2 *The function $g_1 \circ f_1 : \mathbb{S}^2 \to \mathbb{S}^2$ has topological degree ± 1.*

Proof: It is enough to look at the unique preimage of $(1,0,0)$, indicated by the first and last curves in figure 5: the function $g_1 \circ f_1$ is injective on a neighborhood of this point. ■

The sign of the degree depends on the choice of orientation for these two copies of $\mathbb{S}^2$: the domain of f_1 and the image of g_1; we may therefore assume that these orientations were chosen so that the degree is 1 and then these two copies of $\mathbb{S}^2$ were identified by an orientation preserving homeomorphism.

We wrote this section of $z = 1$ only, but everything applies to any $z \in -A$: just attach an arc δ at the end to make the curves closed. The definitions of star and trefoil are based on these closed curves $\gamma * \delta$ and we thus define $g_z : X_z \to \mathbb{S}^2$. The second column of figure 19 consists of trefoils and it is clear that g_z takes the region to the left of this circle to $(\cdot,\cdot,+)$, the region to the right to $(\cdot,\cdot,-)$ and that this circle is taken to the circle $(\cdot,\cdot,0)$ by a function of degree 1.

Proof of theorem 4: All we still have to do is prove that $g_z \circ p_z^2$ is constant: indeed, no curve of the form $p_z^2(\gamma) = \nu^2 * \gamma$ will be near a star or trefoil and therefore $g_z(p_z^2(\gamma)) = (0,0,-1)$ for all $\gamma \in X_z$. ■

Notice that this implies that f_z and $p_z^2 \circ f_z$ are not homotopic.

10 Flowers and the proof of theorem 5

For $z \in A$ and $Q = \Pi(z)$, let $\theta_M \in (0,\pi]$ be defined as follows: if $Qe_1 = e_1$, then θ_M is the argument of (x_2, x_3) where $(0, x_2, x_3) = -Qe_2$; if $Qe_1 \neq e_1$, then θ_M is the argument of (x_2, x_3) where $Qe_1 = (x_1, x_2, x_3)$. A curve $\gamma \in X_{(-1)^k z}$ is a *flower* of $2k+1$ petals if there exist $0 = t_0 < t_1 < t_2 < \cdots < t_{2k} < t_{2k+1} = 1$ and $0 = \theta_0 < \theta_1 < \theta_2 < \cdots < \theta_{2k} < \theta_{2k+1} = \theta_M$ such that:

1. $\gamma(t_1) = \gamma(t_2) = \cdots = \gamma(t_{2k}) = e_1$;

2. the only double points of γ are of the form (t_i, t_j), $i < j$;

3. the argument of $(x_{i,2}, x_{i,3})$ is θ_i, where $(0, x_{i,2}, x_{i,3}) = (-1)^i \gamma'(t_i)$;

4. the restriction of γ to an interval of the form $[t_i, t_{i+1}]$ is convex.

Thus a flower of 1 petal is a convex curve and a flower of 3 petals is a trefoil with the triple point at e_1. Figure 17 shows other examples of flowers. Notice that if γ is a flower than $\gamma(t) \neq -e_1$ for all t. For $k > 0$, let $F_{2k} \subset X_{(-1)^k z}$ be the set of flowers of $2k+1$ petals.

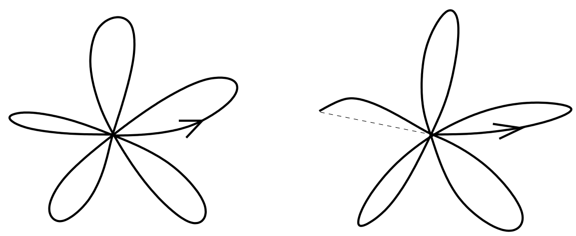

Figure 17: Examples of flowers.

Lemma 10.1 *The set F_{2k} is closed (as a subset of $X_{(-1)^k z}$ and a submanifold of codimension $2k$. Furthermore, the normal bundle to F_{2k} in $X_{(-1)^k z}$ is trivial.*

Proof: Any flower γ has as open neighborhood of curves $\tilde{\gamma}$ as shown in figure 18: an arc from $t = 0$ to $\tilde{t}_1$, with $\tilde{\gamma}(\tilde{t}_1) = (\cos \tilde{\eta}_1, \sin \tilde{\eta}_1, 0)$, another arc from there to $\tilde{t}_2$, with $\tilde{\gamma}(\tilde{t}_2) = (\cos \tilde{\eta}_2, \sin \tilde{\eta}_2, 0)$, and so on, and a final arc from $\tilde{t}_{2k}$ to $t_{2k+1} = 1$. This defines a submersion E from this neighborhood of γ to (a neighborhood of the origin in) $\mathbb{R}^{2k}$ taking $\tilde{\gamma}$ to $(\tilde{\eta}_1, \tilde{\eta}_2, \ldots, \tilde{\eta}_{2k})$. Notice that $\tilde{\gamma} \in F_{2k+1}$ if and only if $E(\tilde{\gamma}) = 0$.

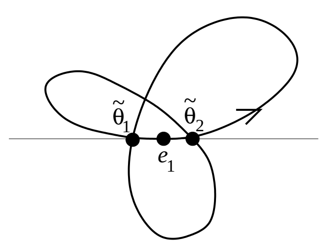

Figure 18: A curve near a flower.

This construction is uniform: E is a submersion from a tubular neighborhood of F_{2k} to $\mathbb{R}^{2k}$. This proves our claims. ■

The intersection number with F_{2k} is therefore well defined and can be interpreted as an element $\mathbf{f}_{2k} \in H^{2k}(X_{(-1)^k z}, \mathbb{Z})$. We may also consider $\mathbf{f}_{2k}$ to be the Poincaré dual of F_{2k}.

Proof of theorem 5: We claim that $\mathbf{f}_{2k} \cdot f^{[k]}_{(-1)^k z} = \pm 1$ where $f^{[k]}_z : (\mathbb{S}^2)^k \to X_z$ was constructed in section 4 (we will again not bother with checking orientations which we can define as we prefer anyway). Indeed, there is a single $(s_1, s_2, \ldots, s_k) \in (\mathbb{S}^2)^k$ such that $f^{[n]}(s_1, s_2, \ldots, s_k)$ is a flower: each $s_i \in \mathbb{S}^2$ has to be taken to be the only point such that (in the notation of section 4) $f_{z_i}(s_i)$ is a trefoil with the triple point at $(Q_1 Q_2 \cdots Q_{i-1})^{-1} e_1$, which is a point on the dashed line in figure 19. It is not hard to check that this intersection is transversal. This proves that $\mathbf{f}_{2k} \neq 0$.

On the other hand, there is obviously no flower in the image of $\nu^2 * f^{[k]}_{(-1)^k z}$ and therefore $\mathbf{f}_{2k} \cdot (\nu^2 * f_1) = 0$.

Let $\mathbf{x} \in H^2(Y_{(-1)^k z}) = H^2(\Omega \mathbb{S}^3)$ be as in section 2 and let $\tilde{\mathbf{x}} = (H^2(i))(\mathbf{x}) \in H^2(X_{(-1)^k z})$. We know from theorem 1 that $H^*(i) : H^*(Y_{(-1)^k z}) \to H^*(X_{(-1)^k z})$ is injective and therefore $0 \neq \tilde{\mathbf{x}}^k \in H^{2k}(X_{(-1)^k z})$. By lemma 2.1, $\tilde{\mathbf{x}}^k \cdot f^{[k]}_{(-1)^k z} = \tilde{\mathbf{x}}^k \cdot (\nu^2 * f^{[k]}_{(-1)^k z})$ and therefore $\mathbf{f}_{2k}$ and $\tilde{\mathbf{x}}^k$ are linearly independent. ■

We conclude by computing the product between the identified elements of $H^*(X_z, \mathbb{R})$.

Proposition 10.2 *For $z \in A$, let $\tilde{\mathbf{x}}, \mathbf{f}_4, \mathbf{f}_8, \ldots \in H^*(X_z, \mathbb{R})$ be defined as above. Then $\tilde{\mathbf{x}}\mathbf{f}_i = \mathbf{f}_i \mathbf{f}_j = 0$ for all i, j.*

Similarly, for $z \in -A$, let $\tilde{\mathbf{x}}, \mathbf{f}_2, \mathbf{f}_6, \ldots \in H^*(X_z, \mathbb{R})$ be defined as above. Then $\tilde{\mathbf{x}}\mathbf{f}_i = \mathbf{f}_i \mathbf{f}_j = 0$ for all i, j.

Proof: The product $\mathbf{f}_i \mathbf{f}_j$ can be interpreted in terms of the intersection between F_i and F_j. Since $F_i \cap F_j = \emptyset$ for $i \neq j$ it follows that $\mathbf{f}_i \mathbf{f}_j = 0$ in this case. Also, since the normal bundle to F_i is trivial, we can uniformly push F_i in some direction to obtain another manifold F_i' homologic to F_i and disjoint from it: thus $\mathbf{f}_i^2 = 0$.

In order to discuss the products $\tilde{\mathbf{x}}\mathbf{f}_i$, we recall the definition of $\tilde{\mathbf{x}}$. If K is a closed oriented surface and $f : K \to X_z$ is a continuous function with $f(s) = \gamma$ then define $\tilde{f} : K \times [0,1] \to \mathbb{S}^3$ by $\tilde{f}(s,t) = \tilde{\Gamma}(t)$. Let $\delta : [0,1] \to \mathbb{S}^3$ be an arbitrary curve with $\delta(0) = z$, $\delta(1) = 1$ and define $\hat{f} : K \times \mathbb{S}^1 \to \mathbb{S}^3$ (where $\mathbb{S}^1$ is $[0,1]$ with identified endpoints) by $\hat{f}(s,t) = \tilde{f}(s, 2t)$ for $t \in [0, 1/2]$ and $\hat{f}(s,t) = \delta(2t-1)$ for $t \in [1/2, 1]$. The product $\tilde{\mathbf{x}} \cdot f$ is the topological degree of $\hat{f}$: this can be computed using any point of $\mathbb{S}^3$ and counting its preimages under $\hat{f}$ with sign. If we take that point to be $j \in \mathbb{S}^3$ and thus obtain a 2-cocycle representing $\tilde{\mathbf{x}}$. Notice that having $\tilde{\Gamma}(t) = j$ implies $\gamma(t) = -e_1$. Since we can not have $\gamma(t) = e_1$ for a flower, the support of this cocycle is disjoint from F_i and therefore $\tilde{\mathbf{x}}\mathbf{f}_i = 0$ for all i. ■

References

[1] Arnol'd, V. I., *The geometry of spherical curves and the algebra of quaternions*, Russian Math. Surveys 50:1, 1-68, 1995.

[2] Bott, R. and Tu, L., *Differential forms in algebraic topology*, Spinger-Verlag, GTM 82, New York, 1982.

[3] Burde, G. and Zieschang, H., *Knots*, Walter de Gruyter, Berlin, 1985.

[4] Burghelea, D., Saldanha, N. and Tomei, C., *Results on infinite dimensional topology and applications to the structure of the critical set of non-linear Sturm-Liouville operators*, J. Differential Equations, 188, 569-590, 2003.

[5] Burghelea, D., Saldanha, N. and Tomei, C., *The topology of the monodromy map of the second order ODE and the structure of the critical set of nonlinear Sturm-Liouville operators*, preprint.

[6] Hirsch, M., *Immersions of manifolds*, Trans. Am. Math. Soc. 93, 242-276, 1959.

[7] Little, J. A., *Nondegenerate homotopies of curves on the unit 2-sphere*, J. Differential Geometry, 4, 339-348, 1970.

[8] Shapiro, B. and Khesin, B., *Homotopy classification of nondegenerate quasiperiodic curves on the* 2*-sphere*, Publ. Inst. Math. (Beograd) 66(80), 127-156, 1999.

[9] Shapiro, B. and Shapiro, M., *On the number of connected components of nondegenerate curves on* $\mathbb{S}^n$, Bull. of the AMS, 25, 75-79, 1991.

[10] Shapiro, M., *Topology of the space of nondegenerate curves*, Math. USSR, 57, 106-126, 1993.

[11] Smale, S., *The classification of immersions of spheres in euclidean spaces*, Ann. of Math. 69, 327-344, 1959.

Nicolau C. Saldanha, PUC-Rio
nicolau@mat.puc-rio.br; http://www.mat.puc-rio.br/∼nicolau/

Departamento de Matemática, PUC-Rio
R. Marquês de S. Vicente 225, Rio de Janeiro, RJ 22453-900, Brazil

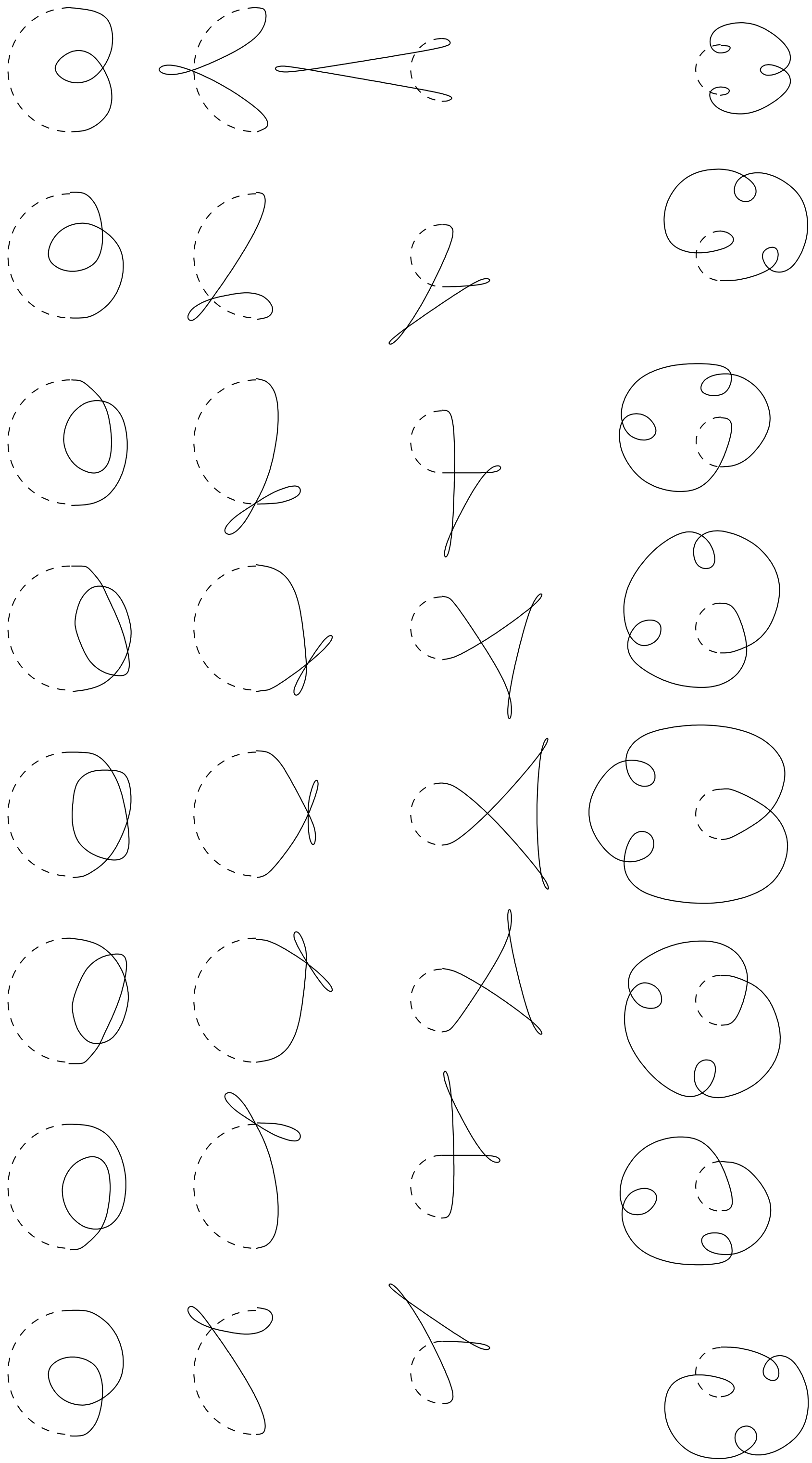

Figure 19: The function $f_z : \mathbb{S}^2 \to X_z$, $-z \in A$.