Combinatorial Floer Homology

Vin de Silva
Pomona College

Joel W. Robbin
University of Wisconsin*

Dietmar A. Salamon†
ETH-Zürich

6 February 2013

*Part of this paper was written while JWR visited the FIM at ETH Zürich
†Partially supported by the Swiss National Science Foundation Grant 200021-127136

Contents

1 Introduction

The Floer homology of a transverse pair of Lagrangian submanifolds in a symplectic manifold is, under favorable hypotheses, the homology of a chain complex generated by the intersection points. The boundary operator counts index one holomorphic strips with boundary on the Lagrangian submanifolds. This theory was introduced by Floer in [10, 11]; see also the three papers [21] of Oh. In this memoir we consider the following special case:

> **(H)** Σ *is a connected oriented 2-manifold without boundary and* $\alpha, \beta \subset \Sigma$ *are connected smooth one dimensional oriented submanifolds without boundary which are closed as subsets of* Σ *and intersect transversally.* We do not assume that Σ is compact, but when it is, α and β are embedded circles.

In this special case there is a purely combinatorial approach to Lagrangian Floer homology which was first developed by de Silva [6]. We give a full and detailed definition of this combinatorial Floer homology (see Theorem 9.1) under the hypothesis that α and β are noncontractible embedded circles and are not isotopic to each other. Under this hypothesis, *combinatorial Floer homology is invariant under isotopy, not just Hamiltonian isotopy, as in Floer's original work* (see Theorem 9.2). Combinatorial Floer homology *is* isomorphic to analytic Floer homology as defined by Floer (see Theorem 9.3).

Floer homology is the homology of a chain complex $\mathrm{CF}(\alpha, \beta)$ with basis consisting of the points of the intersection $\alpha \cap \beta$ (and coefficients in $\mathbb{Z}/2\mathbb{Z}$). The boundary operator $\partial : \mathrm{CF}(\alpha, \beta) \to \mathrm{CF}(\alpha, \beta)$ has the form

$$\partial x = \sum_y n(x, y) y.$$

In the case of analytic Floer homology as defined by Floer $n(x, y)$ denotes the number (mod two) of equivalence classes of holomorphic strips $v : \mathbb{S} \to \Sigma$ satisfying the boundary conditions

$$v(\mathbb{R}) \subset \alpha, \qquad v(\mathbb{R} + \mathbf{i}) \subset \beta, \qquad v(-\infty) = x, \qquad v(+\infty) = y$$

and having Maslov index one. The boundary operator in combinatorial Floer homology has the same form but now $n(x, y)$ denotes the number (mod two) of equivalence classes of smooth immersions $u : \mathbb{D} \to \Sigma$ satisfying

$$u(\mathbb{D} \cap \mathbb{R}) \subset \alpha, \qquad u(\mathbb{D} \cap S^1) \subset \beta, \qquad u(-1) = x, \qquad u(+1) = y.$$

We call such an immersion a *smooth lune*. Here

$$\mathbb{S} := \mathbb{R} + \mathbf{i}[0,1], \qquad \mathbb{D} := \{z \in \mathbb{C} \,|\, \mathrm{Im}\, z \geq 0,\ |z| \leq 1\}$$

denote the standard strip and the standard half disc respectively. We develop the combinatorial theory without appeal to the difficult analysis required for the analytic theory. The invariance under isotopy rather than just Hamiltonian isotopy (Theorem 9.3) is a benefit of this approach. A corollary is the formula

$$\dim \mathrm{HF}(\alpha, \beta) = \mathrm{geo}\,(\alpha, \beta)$$

for the dimension of the Floer Homology $\mathrm{HF}(\alpha, \beta)$ (see Corollary 9.5). Here $\mathrm{geo}\,(\alpha, \beta)$ denotes the *geometric intersection number* of the curves α and β. In Remark 9.11 we indicate how to define combinatorial Floer homology with integer coefficients, but we do not discuss integer coefficients in analytic Floer homology.

Let $\mathcal{D}$ denote the space of all smooth maps $u : \mathbb{D} \to \Sigma$ satisfying the boundary conditions $u(\mathbb{D} \cap \mathbb{R}) \subset \alpha$ and $u(\mathbb{D} \cap S^1) \subset \beta$. For $x, y \in \alpha \cap \beta$ let $\mathcal{D}(x, y)$ denote the subset of all $u \in \mathcal{D}$ satisfying the endpoint conditions $u(-1) = x$ and $u(1) = y$. Each $u \in \mathcal{D}$ determines a locally constant function

$$\mathrm{w} : \Sigma \setminus (\alpha \cup \beta) \to \mathbb{Z}$$

defined as the degree

$$\mathrm{w}(z) := \deg(u, z), \qquad z \in \Sigma \setminus (\alpha \cup \beta).$$

When z is a regular value of u this is the algebraic number of points in the preimage $u^{-1}(z)$. The function w depends only on the homotopy class of u. In Theorem 2.4 we prove that the homotopy class of $u \in \mathcal{D}$ is uniquely determined by its endpoints x, y and its degree function w. Theorem 3.4 says that the Viterbo–Maslov index of every smooth map $u \in \mathcal{D}(x, y)$ is determined by the values of w near the endpoints x and y of u, namely, it is given by the following *trace formula*

$$\mu(u) = \frac{m_x(\Lambda_u) + m_y(\Lambda_u)}{2}, \qquad \Lambda_u := (x, y, \mathrm{w}).$$

Here m_x denotes the sum of the four values of w encountered when walking along a small circle surrounding x, and similarly for y. Part I of this memoir is devoted to proving this formula.

Part II gives a combinatorial characterization of smooth lunes. Specifically, the equivalent conditions (ii) and (iii) of Theorem 6.7 are necessary for the existence of a smooth lune. This implies the fact (not obvious to us) that a lune cannot contain either of its endpoints in the interior of its image. In the simply connected case we prove in the same theorem that the necessary conditions are also sufficient. We conjecture that they characterize smooth lunes in general. Theorem 6.8 shows that any two smooth lunes with the same counting function w are isotopic and thus the equivalence class of a smooth lune is uniquely determined by its combinatorial data. The proofs of these theorems are carried out in Sections 7 and 8. Together they provide a solution to the Picard–Loewner problem in a special case; see for example [12] and the references cited therein, e.g. [38, 4, 28]. Our result is a special case because no critical points are allowed (lunes are immersions), the source is a disc and not a Riemann surface with positive genus, and the prescribed boundary circle decomposes into two embedded arcs.

Part III introduces combinatorial Floer homology. Here we restrict our discussion to the case where α and β are noncontractible embedded circles which are not isotopic to each other (with either orientation). The basic definitions are given in Section 9. That the square of the boundary operator is indeed zero in the combinatorial setting will be proved in Section 10 by analyzing *broken hearts*. Propositions 10.2 and 10.5 say that *there are two ways to break a heart* and this is why the square of the boundary operator is zero. In Section 11 we prove the isotopy invariance of combinatorial Floer homology by examining generic deformations of loops that change the number of intersection points. This is very much in the spirit of Floer's original proof of deformation invariance (under Hamiltonian isotopy of the Lagrangian manifolds) of analytic Floer homology. The main theorem in Section 12 asserts, in the general setting, that smooth lunes (up to isotopy) are in one-to-one correspondence with index one holomorphic strips (up to translation). The proof is self-contained and does not use any of the other results in this memoir. It is based on an equation (the *index formula* (69) in Theorem 12.2) which expresses the Viterbo–Maslov index of a holomorphic strip in terms of its critical points and its angles at infinity. A linear version of this equation (the *linear index formula* (76) in Lemma 12.3) also shows that every holomorphic strip is regular in the sense that the linearized operator is surjective. It follows from these observations that the combinatorial and analytic definitions of Floer homology agree as asserted in Theorem 9.3. In fact, our results show that the two chain complexes agree.

There are many directions in which the theory developed in the present memoir can be extended. Some of these are discussed in Section 13. For example, it has been understood for some time that the Donaldson triangle product and the Fukaya category have combinatorial analogues in dimension two, and that these analogues are isomorphic to the original analytic theories. The combinatorial approach to the Donaldson triangle product has been outlined in the PhD thesis of the first author [6], and the combinatorial approach to the derived Fukaya category has been used by Abouzaid [1] to compute it. Our formula for the Viterbo–Maslov index in Theorem 3.4 and our combinatorial characterization of smooth lunes in Theorem 6.7 are not needed for their applications. In our memoir these two results are limited to the elements of $\mathcal{D}$. (To our knowledge, they have not been extended to triangles or more general polygons in the existing literature.)

When $\Sigma = \mathbb{T}^2$, the Heegaard–Floer theory of Ozsvath–Szabo [26, 27] can be interpreted as a refinement of the combinatorial Floer theory, in that the winding number of a lune at a prescribed point in $\mathbb{T}^2 \setminus (\alpha \cup \beta)$ is taken into account in the definition of their boundary operator. However, for higher genus surfaces Heegaard–Floer theory does not include the combinatorial Floer theory discussed in the present memoir as a special case.

Appendix A contains a proof that, under suitable hypotheses, the space of paths connecting α to β is simply connected. Appendix B contains a proof that the group of orientation preserving diffeomorphisms of the half disc fixing the corners is connected. Appendix C contains an account of Floer's algebraic deformation argument. Appendix D summarizes the relevant results in [32] about the asymptotic behavior of holomorphic strips.

Acknowledgement. We would like to thank the referee for his/her careful work.

I. The Viterbo–Maslov Index

Throughout this memoir we assume (H). We often write "assume (H)" to remind the reader of our standing hypothesis.

2 Chains and Traces

Define a cell complex structure on Σ by taking the set of zero-cells to be the set $\alpha \cap \beta$, the set of one-cells to be the set of connected components of $(\alpha \setminus \beta) \cup (\beta \setminus \alpha)$ with compact closure, and the set of two-cells to be the set of connected components of $\Sigma \setminus (\alpha \cup \beta)$ with compact closure. (There is an abuse of language here as the "two-cells" need not be homeomorphs of the open unit disc if the genus of Σ is positive and the "one-cells" need not be arcs if $\alpha \cap \beta = \emptyset$.) Define a boundary operator ∂ as follows. For each two-cell F let

$$\partial F = \sum \pm E,$$

where the sum is over the one-cells E which abut F and the plus sign is chosen iff the orientation of E (determined from the given orientations of α and β) agrees with the boundary orientation of F as a connected open subset of the oriented manifold Σ. For each one-cell E let

$$\partial E = y - x$$

where x and y are the endpoints of the arc E and the orientation of E goes from x to y. (The one-cell E is either a subarc of α or a subarc of β and both α and β are oriented one-manifolds.) For $k = 0, 1, 2$ a **k-chain** is defined to be a formal linear combination (with integer coefficients) of k-cells, i.e. a two-chain is a locally constant map $\Sigma \setminus (\alpha \cup \beta) \to \mathbb{Z}$ (whose support has compact closure in Σ) and a one-chain is a locally constant map $(\alpha \setminus \beta) \cup (\beta \setminus \alpha) \to \mathbb{Z}$ (whose support has compact closure in $\alpha \cup \beta$). It follows directly from the definitions that $\partial^2 F = 0$ for each two-cell F.

Each $u \in \mathcal{D}$ determines a two-chain w via

$$\mathrm{w}(z) := \deg(u, z), \qquad z \in \Sigma \setminus (\alpha \cup \beta). \tag{1}$$

and a one-chain ν via

$$\nu(z) := \begin{cases} \deg(u|_{\partial\mathbb{D}\cap\mathbb{R}} : \partial\mathbb{D} \cap \mathbb{R} \to \alpha, z), & \text{for } z \in \alpha \setminus \beta, \\ -\deg(u|_{\partial\mathbb{D}\cap S^1} : \partial\mathbb{D} \cap S^1 \to \beta, z), & \text{for } z \in \beta \setminus \alpha. \end{cases} \tag{2}$$

Here we orient the one-manifolds $\mathbb{D}\cap\mathbb{R}$ and $\mathbb{D}\cap S^1$ from -1 to $+1$. For any one-chain $\nu:(\alpha\setminus\beta)\cup(\beta\setminus\alpha)\to\mathbb{Z}$ denote

$$\nu_\alpha := \nu|_{\alpha\setminus\beta}:\alpha\setminus\beta\to\mathbb{Z},\qquad \nu_\beta := \nu|_{\beta\setminus\alpha}:\beta\setminus\alpha\to\mathbb{Z}.$$

Conversely, given locally constant functions $\nu_\alpha:\alpha\setminus\beta\to\mathbb{Z}$ (whose support has compact closure in α) and $\nu_\beta:\beta\setminus\alpha\to\mathbb{Z}$ (whose support has compact closure in β), denote by $\nu=\nu_\alpha-\nu_\beta$ the one-chain that agrees with ν_α on $\alpha\setminus\beta$ and agrees with $-\nu_\beta$ on $\beta\setminus\alpha$.

Definition 2.1 (**Traces). Fix two (not necessarily distinct) intersection points $x,y\in\alpha\cap\beta$.

(i) Let $\mathrm{w}:\Sigma\setminus(\alpha\cup\beta)\to\mathbb{Z}$ be a two-chain. The triple $\Lambda=(x,y,\mathrm{w})$ is called an (α,β)**-trace** if there exists an element $u\in\mathcal{D}(x,y)$ such that w is given by (1). In this case $\Lambda=:\Lambda_u$ is also called the (α,β)**-trace of** u and we sometimes write $\mathrm{w}_u:=\mathrm{w}$.

(ii) Let $\Lambda=(x,y,\mathrm{w})$ be an (α,β)-trace. The triple $\partial\Lambda:=(x,y,\partial\mathrm{w})$ is called the **boundary of** Λ.

(iii) A one-chain $\nu:(\alpha\setminus\beta)\cup(\beta\setminus\alpha)\to\mathbb{Z}$ is called an (x,y)**-trace** if there exist smooth curves $\gamma_\alpha:[0,1]\to\alpha$ and $\gamma_\beta:[0,1]\to\beta$ such that $\gamma_\alpha(0)=\gamma_\beta(0)=x$, $\gamma_\alpha(1)=\gamma_\beta(1)=y$, γ_α and γ_β are homotopic in Σ with fixed endpoints, and

$$\nu(z)=\left\{\begin{array}{rl}\deg(\gamma_\alpha,z), & \text{for } z\in\alpha\setminus\beta,\\ -\deg(\gamma_\beta,z), & \text{for } z\in\beta\setminus\alpha.\end{array}\right. \tag{3}$$

Remark 2.2. Assume Σ is simply connected. Then the condition on γ_α and γ_β to be homotopic with fixed endpoints is redundant. Moreover, if $x=y$ then a one-chain ν is an (x,y)-trace if and only if the restrictions

$$\nu_\alpha:=\nu|_{\alpha\setminus\beta},\qquad \nu_\beta:=-\nu|_{\beta\setminus\alpha}$$

are constant. If $x\neq y$ and α,β are embedded circles and A,B denote the positively oriented arcs from x to y in α,β, then a one-chain ν is an (x,y)-trace if and only if

$$\nu_\alpha|_{\alpha\setminus(A\cup\beta)}=\nu_\alpha|_{A\setminus\beta}-1$$

and

$$\nu_\beta|_{\beta\setminus(B\cup\alpha)}=\nu_\beta|_{B\setminus\alpha}-1.$$

In particular, when walking along α or β, the function ν only changes its value at x and y.

Lemma 2.3. *Let $x, y \in \alpha \cap \beta$ and $u \in \mathcal{D}(x, y)$. Then the boundary of the (α, β)-trace Λ_u of u is the triple $\partial\Lambda_u = (x, y, \nu)$, where ν is given by (2). In other words, if* w *is given by (1) and ν is given by (2) then $\nu = \partial\mathrm{w}$.*

Proof. Choose an embedding $\gamma : [-1, 1] \to \Sigma$ such that u is transverse to γ, $\gamma(t) \in \Sigma \setminus (\alpha \cup \beta)$ for $t \neq 0$, $\gamma(-1)$, $\gamma(1)$ are regular values of u, $\gamma(0) \in \alpha \setminus \beta$ is a regular value of $u|_{\mathbb{D}\cap\mathbb{R}}$, and γ intersects α transversally at $t = 0$ such that orientations match in

$$T_{\gamma(0)}\Sigma = T_{\gamma(0)}\alpha \oplus \mathbb{R}\dot{\gamma}(0).$$

Denote $\Gamma := \gamma([-1, 1])$. Then $u^{-1}(\Gamma) \subset \mathbb{D}$ is a 1-dimensional submanifold with boundary

$$\partial u^{-1}(\Gamma) = u^{-1}(\gamma(-1)) \cup u^{-1}(\gamma(1)) \cup \big(u^{-1}(\gamma(0)) \cap \mathbb{R}\big).$$

If $z \in u^{-1}(\Gamma)$ then

$$\mathrm{im}\, du(z) + T_{u(z)}\Gamma = T_{u(z)}\Sigma, \qquad T_z u^{-1}(\Gamma) = du(z)^{-1} T_{u(z)}\Gamma.$$

We orient $u^{-1}(\Gamma)$ such that the orientations match in

$$T_{u(z)}\Sigma = T_{u(z)}\Gamma \oplus du(z)\mathbf{i}T_z u^{-1}(\Gamma).$$

In other words, if $z \in u^{-1}(\Gamma)$ and $u(z) = \gamma(t)$, then a nonzero tangent vector $\zeta \in T_z u^{-1}(\Gamma)$ is positive if and only if the pair $(\dot{\gamma}(t), du(z)\mathbf{i}\zeta)$ is a positive basis of $T_{\gamma(t)}\Sigma$. Then the boundary orientation of $u^{-1}(\Gamma)$ at the elements of $u^{-1}(\gamma(1))$ agrees with the algebraic count in the definition of $\mathrm{w}(\gamma(1))$, at the elements of $u^{-1}(\gamma(-1))$ is opposite to the algebraic count in the definition of $\mathrm{w}(\gamma(-1))$, and at the elements of $u^{-1}(\gamma(0)) \cap \mathbb{R}$ is opposite to the algebraic count in the definition of $\nu(\gamma(0))$. Hence

$$\mathrm{w}(\gamma(1)) = \mathrm{w}(\gamma(-1)) + \nu(\gamma(0)).$$

In other words the value of ν at a point in $\alpha \setminus \beta$ is equal to the value of w slightly to the left of α minus the value of w slightly to the right of α. Likewise, the value of ν at a point in $\beta \setminus \alpha$ is equal to the value of w slightly to the right of β minus the value of w slightly to the left of β. This proves Lemma 2.3. $\square$

Theorem 2.4. **(i)** *Two elements of $\mathcal{D}$ belong to the same connected component of $\mathcal{D}$ if and only if they have the same (α, β)-trace.*

(ii) *Assume Σ is diffeomorphic to the two-sphere. Let $x, y \in \alpha \cap \beta$ and let $\mathrm{w} : \Sigma \setminus (\alpha \cup \beta) \to \mathbb{Z}$ be a locally constant function. Then $\Lambda = (x, y, \mathrm{w})$ is an (α, β)-trace if and only if $\partial \mathrm{w}$ is an (x, y)-trace.*

(iii) *Assume Σ is not diffeomorphic to the two-sphere and let $x, y \in \alpha \cap \beta$. If ν is an (x, y)-trace, then there is a unique two-chain w such that $\Lambda := (x, y, \mathrm{w})$ is an (α, β)-trace and $\partial \mathrm{w} = \nu$.*

Proof. We prove (i). "Only if" follows from the standard arguments in degree theory as in Milnor [19]. To prove "if", fix two intersection points

$$x, y \in \alpha \cap \beta$$

and, for $X = \Sigma, \alpha, \beta$, denote by $\mathcal{P}(x, y; X)$ the space of all smooth curves $\gamma : [0, 1] \to X$ satisfying $\gamma(0) = x$ and $\gamma(1) = y$. Every $u \in \mathcal{D}(x, y)$ determines smooth paths $\gamma_{u,\alpha} \in \mathcal{P}(x, y; \alpha)$ and $\gamma_{u,\beta} \in \mathcal{P}(x, y; \beta)$ via

$$\gamma_{u,\alpha}(s) := u(-\cos(\pi s), 0), \qquad \gamma_{u,\beta}(s) = u(-\cos(\pi s), \sin(\pi s)). \tag{4}$$

These paths are homotopic in Σ with fixed endpoints. An explicit homotopy is the map

$$F_u := u \circ \varphi : [0, 1]^2 \to \Sigma$$

where $\varphi : [0, 1]^2 \to \mathbb{D}$ is the map

$$\varphi(s, t) := (-\cos(\pi s), t \sin(\pi s)).$$

By Lemma 2.3, the homotopy class of $\gamma_{u,\alpha}$ in $\mathcal{P}(x, y; \alpha)$ is uniquely determined by

$$\nu_\alpha := \partial \mathrm{w}_u|_{\alpha \setminus \beta} : \alpha \setminus \beta \to \mathbb{Z}$$

and that of $\gamma_{u,\beta}$ in $\mathcal{P}(x, y; \beta)$ is uniquely determined by

$$\nu_\beta := -\partial \mathrm{w}_u|_{\beta \setminus \alpha} : \beta \setminus \alpha \to \mathbb{Z}.$$

Hence they are both uniquely determined by the (α, β)-trace of u. If Σ is not diffeomorphic to the 2-sphere the assertion follows from the fact that each component of $\mathcal{P}(x, y; \Sigma)$ is contractible (because the universal cover of Σ is diffeomorphic to the complex plane). Now assume Σ is diffeomorphic to the 2-sphere. Then $\pi_1(\mathcal{P}(x,y;\Sigma)) = \mathbb{Z}$ acts on $\pi_0(\mathcal{D})$ because the correspondence $u \mapsto F_u$ identifies $\pi_0(\mathcal{D})$ with a space of homotopy classes of paths in $\mathcal{P}(x,y;\Sigma)$ connecting $\mathcal{P}(x,y;\alpha)$ to $\mathcal{P}(x,y;\beta)$. The induced action on the space of two-chains $\mathrm{w} : \Sigma \setminus (\alpha \cup \beta)$ is given by adding a global constant. Hence the map $u \mapsto \mathrm{w}$ induces an injective map

$$\pi_0(\mathcal{D}(x,y)) \to \{\text{2-chains}\}.$$

This proves (i).

We prove (ii) and (iii). Let w be a two-chain, suppose that $\nu := \partial\mathrm{w}$ is an (x,y)-trace, and denote $\Lambda := (x,y,\mathrm{w})$. Let $\gamma_\alpha : [0,1] \to \alpha$ and $\gamma_\beta : [0,1] \to \beta$ be as in Definition 2.1. Then there is a $u' \in \mathcal{D}(x,y)$ such that the map $s \mapsto u'(-\cos(\pi s), 0)$ is homotopic to γ_α and $s \mapsto u'(-\cos(\pi s), \sin(\pi s))$ is homotopic to γ_β. By definition the (α,β)-trace of u' is $\Lambda' = (x,y,\mathrm{w}')$ for some two-chain w'. By Lemma 2.3, we have

$$\partial\mathrm{w}' = \nu = \partial\mathrm{w}$$

and hence $\mathrm{w} - \mathrm{w}' =: d$ is constant. If Σ is not diffeomorphic to the two-sphere and Λ is the (α,β)-trace of some element $u \in \mathcal{D}$, then u is homotopic to u' (as $\mathcal{P}(x,y;\Sigma)$ is simply connected) and hence $d = 0$ and $\Lambda = \Lambda'$. If Σ is diffeomorphic to the 2-sphere choose a smooth map $v : S^2 \to \Sigma$ of degree d and replace u' by the connected sum $u := u' \# v$. Then Λ is the (α,β)-trace of u. This proves Theorem 2.4. $\square$

Remark 2.5. Let $\Lambda = (x,y,\mathrm{w})$ be an (α,β)-trace and define

$$\nu_\alpha := \partial\mathrm{w}|_{\alpha\setminus\beta}, \qquad \nu_\beta := -\partial\mathrm{w}|_{\beta\setminus\alpha}.$$

(i) The two-chain w is uniquely determined by the condition $\partial\mathrm{w} = \nu_\alpha - \nu_\beta$ and its value at one point. To see this, think of the embedded circles α and β as traintracks. Crossing α at a point $z \in \alpha \setminus \beta$ increases w by $\nu_\alpha(z)$ if the train comes from the left, and decreases it by $\nu_\alpha(z)$ if the train comes from the right. Crossing β at a point $z \in \beta \setminus \alpha$ decreases w by $\nu_\beta(z)$ if the train comes from the left and increases it by $\nu_\beta(z)$ if the train comes from the right. Moreover, ν_α extends continuously to $\alpha \setminus \{x,y\}$ and ν_β extends continuously to $\beta \setminus \{x,y\}$. At each intersection point $z \in (\alpha \cap \beta) \setminus \{x,y\}$ with intersection index $+1$ (respectively -1) the function w takes the values

$$k, \quad k + \nu_\alpha(z), \quad k + \nu_\alpha(z) - \nu_\beta(z), \quad k - \nu_\beta(z)$$

as we march counterclockwise (respectively clockwise) along a small circle surrounding the intersection point.

(ii) If Σ is not diffeomorphic to the 2-sphere then, by Theorem 2.4 (iii), the (α,β)-trace Λ is uniquely determined by its boundary $\partial\Lambda = (x, y, \nu_\alpha - \nu_\beta)$.

(iii) Assume Σ is not diffeomorphic to the 2-sphere and choose a universal covering $\pi : \mathbb{C} \to \Sigma$. Choose a point $\widetilde{x} \in \pi^{-1}(x)$ and lifts $\widetilde{\alpha}$ and $\widetilde{\beta}$ of α and β such that $\widetilde{x} \in \widetilde{\alpha} \cap \widetilde{\beta}$. Then Λ lifts to an $(\widetilde{\alpha}, \widetilde{\beta})$-trace

$$\widetilde{\Lambda} = (\widetilde{x}, \widetilde{y}, \widetilde{\mathrm{w}}).$$

More precisely, the one chain $\nu := \nu_\alpha - \nu_\beta = \partial\mathrm{w}$ is an (x, y)-trace, by Lemma 2.3. The paths $\gamma_\alpha : [0, 1] \to \alpha$ and $\gamma_\beta : [0, 1] \to \beta$ in Definition 2.1 lift to unique paths $\gamma_{\widetilde{\alpha}} : [0, 1] \to \widetilde{\alpha}$ and $\gamma_{\widetilde{\beta}} : [0, 1] \to \widetilde{\beta}$ connecting $\widetilde{x}$ to $\widetilde{y}$. For $\widetilde{z} \in \mathbb{C} \setminus (\widetilde{A} \cup \widetilde{B})$ the number $\widetilde{\mathrm{w}}(\widetilde{z})$ is the winding number of the loop $\gamma_{\widetilde{\alpha}} - \gamma_{\widetilde{\beta}}$ about $\widetilde{z}$ (by Rouché's theorem). The two-chain w is then given by

$$\mathrm{w}(z) = \sum_{\widetilde{z} \in \pi^{-1}(z)} \widetilde{\mathrm{w}}(\widetilde{z}), \qquad z \in \Sigma \setminus (\alpha \cup \beta).$$

To see this, lift an element $u \in \mathcal{D}(x, y)$ with (α, β)-trace Λ to the universal cover to obtain an element $\widetilde{u} \in \mathcal{D}(\widetilde{x}, \widetilde{y})$ with $\Lambda_{\widetilde{u}} = \widetilde{\Lambda}$ and consider the degree.

Definition 2.6 (Catenation). Let $x, y, z \in \alpha \cap \beta$. The **catenation of two (α, β)-traces $\Lambda = (x, y, \mathrm{w})$ and $\Lambda' = (y, z, \mathrm{w}')$** is defined by

$$\Lambda \# \Lambda' := (x, z, \mathrm{w} + \mathrm{w}').$$

Let $u \in \mathcal{D}(x, y)$ and $u' \in \mathcal{D}(y, z)$ and suppose that u and u' are constant near the ends $\pm 1 \in \mathbb{D}$. For $0 < \lambda < 1$ sufficiently close to one the **λ-catenation of u and u'** is the map $u\#_\lambda u' \in \mathcal{D}(x, z)$ defined by

$$(u\#_\lambda u')(\zeta) := \begin{cases} u\left(\frac{\zeta+\lambda}{1+\lambda\zeta}\right), & \text{for } \mathrm{Re}\,\zeta \le 0, \\ u'\left(\frac{\zeta-\lambda}{1-\lambda\zeta}\right), & \text{for } \mathrm{Re}\,\zeta \ge 0. \end{cases}$$

Lemma 2.7. *If $u \in \mathcal{D}(x, y)$ and $u' \in \mathcal{D}(y, z)$ are as in Definition 2.6 then*

$$\Lambda_{u\#_\lambda u'} = \Lambda_u \# \Lambda_{u'}.$$

Thus the catenation of two (α, β)-traces is again an (α, β)-trace.

Proof. This follows directly from the definitions. □

3 The Maslov Index

Definition 3.1. Let $x, y \in \alpha \cap \beta$ and $u \in \mathcal{D}(x, y)$. Choose an orientation preserving trivialization

$$\mathbb{D} \times \mathbb{R}^2 \to u^*T\Sigma : (z, \zeta) \mapsto \Phi(z)\zeta$$

and consider the Lagrangian paths

$$\lambda_0, \lambda_1 : [0, 1] \to \mathbb{R}\mathrm{P}^1$$

given by

$$\begin{aligned}
\lambda_0(s) &:= \Phi(-\cos(\pi s), 0)^{-1} T_{u(-\cos(\pi s),0)}\alpha, \\
\lambda_1(s) &:= \Phi(-\cos(\pi s), \sin(\pi s))^{-1} T_{u(-\cos(\pi s),\sin(\pi s))}\beta.
\end{aligned}$$

The **Viterbo–Maslov index of** u is defined as the relative Maslov index of the pair of Lagrangian paths (λ_0, λ_1) and will be denoted by

$$\mu(u) := \mu(\Lambda_u) := \mu(\lambda_0, \lambda_1).$$

By the naturality and homotopy axioms for the relative Maslov index (see for example [30]), the number $\mu(u)$ is independent of the choice of the trivialization and depends only on the homotopy class of u; hence it depends only on the (α, β)-trace of u, by Theorem 2.4. The relative Maslov index $\mu(\lambda_0, \lambda_1)$ is the degree of the loop in $\mathbb{R}\mathrm{P}^1$ obtained by traversing λ_0, followed by a counterclockwise turn from $\lambda_0(1)$ to $\lambda_1(1)$, followed by traversing λ_1 in reverse time, followed by a clockwise turn from $\lambda_1(0)$ to $\lambda_0(0)$. This index was first defined by Viterbo [39] (in all dimensions). Another exposition is contained in [30].

Remark 3.2. The Viterbo–Maslov index is additive under catenation, i.e. if

$$\Lambda = (x, y, \mathrm{w}), \qquad \Lambda' = (y, z, \mathrm{w}')$$

are (α, β)-traces then

$$\mu(\Lambda \# \Lambda') = \mu(\Lambda) + \mu(\Lambda').$$

For a proof of this formula see [39, 30].

Definition 3.3. Let $\Lambda = (x, y, \mathrm{w})$ be an (α, β)-trace and denote $\nu_\alpha := \partial \mathrm{w}|_{\alpha\setminus\beta}$ and $\nu_\beta := -\partial \mathrm{w}|_{\beta\setminus\alpha}$. Λ is said to satisfy the **arc condition** if

$$x \neq y, \qquad \min|\nu_\alpha| = \min|\nu_\beta| = 0. \tag{5}$$

When Λ satisfies the arc condition there are arcs $A \subset \alpha$ and $B \subset \beta$ from x to y such that

$$\nu_\alpha(z) = \begin{cases} \pm 1, & \text{if } z \in A, \\ 0, & \text{if } z \in \alpha \setminus \overline{A}, \end{cases} \qquad \nu_\beta(z) = \begin{cases} \pm 1, & \text{if } z \in B, \\ 0, & \text{if } z \in \beta \setminus \overline{B}. \end{cases} \tag{6}$$

Here the plus sign is chosen iff the orientation of A from x to y agrees with that of α, respectively the orientation of B from x to y agrees with that of β. In this situation the quadruple (x, y, A, B) and the triple $(x, y, \partial \mathrm{w})$ determine one another and we also write

$$\partial\Lambda = (x, y, A, B)$$

for the boundary of Λ. When $u \in \mathcal{D}$ and $\Lambda_u = (x, y, \mathrm{w})$ satisfies the arc condition and $\partial\Lambda_u = (x, y, A, B)$ then

$$s \mapsto u(-\cos(\pi s), 0)$$

is homotopic in α to a path traversing A and the path

$$s \mapsto u(-\cos(\pi s), \sin(\pi s))$$

is homotopic in β to a path traversing B.

Theorem 3.4. *Let $\Lambda = (x, y, \mathrm{w})$ be an (α, β)-trace. For $z \in \alpha \cap \beta$ denote by $m_z(\Lambda)$ the sum of the four values of w encountered when walking along a small circle surrounding z. Then the Viterbo–Maslov index of Λ is given by*

$$\mu(\Lambda) = \frac{m_x(\Lambda) + m_y(\Lambda)}{2}. \tag{7}$$

We call this the **trace formula**.

We first prove the trace formula for the 2-plane $\mathbb{C}$ and the 2-sphere S^2 (Section 4 on page 24). When Σ is not simply connected we reduce the result to the case of the 2-plane (Section 5 page 38). The key is the identity

$$m_{g\widetilde{x}}(\widetilde{\Lambda}) + m_{g^{-1}\widetilde{y}}(\widetilde{\Lambda}) = 0 \tag{8}$$

for every lift $\widetilde{\Lambda}$ to the universal cover and every deck transformation $g \neq \mathrm{id}$. We call this the **cancellation formula**.

4 The Simply Connected Case

A connected oriented 2-manifold Σ is called **planar** if it admits an (orientation preserving) embedding into the complex plane.

Proposition 4.1. *The trace formula* (7) *holds when* Σ *is planar.*

Proof. Assume first that $\Sigma = \mathbb{C}$ and $\Lambda = (x, y, \mathrm{w})$ satisfies the arc condition. Thus the boundary of Λ has the form

$$\partial\Lambda = (x, y, A, B),$$

where $A \subset \alpha$ and $B \subset \beta$ are arcs from x to y and $\mathrm{w}(z)$ is the winding number of the loop $A - B$ about the point $z \in \Sigma \setminus (A \cup B)$ (see Remark 2.5). Hence the trace formula (7) can be written in the form

$$\mu(\Lambda) = 2k_x + 2k_y + \frac{\varepsilon_x - \varepsilon_y}{2}. \tag{9}$$

Here $\varepsilon_z = \varepsilon_z(\Lambda) \in \{+1, -1\}$ denotes the intersection index of A and B at a point $z \in A \cap B$, $k_x = k_x(\Lambda)$ denotes the value of the winding number w at a point in $\alpha \setminus A$ close to x, and $k_y = k_y(\Lambda)$ denotes the value of w at a point in $\alpha \setminus A$ close to y. We now prove (9) under the hypothesis that Λ satisfies the arc condition. The proof is by induction on the number of intersection points of B and α and has seven steps.

Step 1. *We may assume without loss of generality that*

$$\Sigma = \mathbb{C}, \qquad \alpha = \mathbb{R}, \qquad A = [x, y], \qquad x < y, \tag{10}$$

and $B \subset \mathbb{C}$ *is an embedded arc from* x *to* y *that is transverse to* $\mathbb{R}$*.*

Choose a diffeomorphism from Σ to $\mathbb{C}$ that maps A to a bounded closed interval and maps x to the left endpoint of A. If α is not compact the diffeomorphism can be chosen such that it also maps α to $\mathbb{R}$. If α is an embedded circle the diffeomorphism can be chosen such that its restriction to B is transverse to $\mathbb{R}$; now replace the image of α by $\mathbb{R}$. This proves Step 1.

Step 2. *Assume* (10) *and let* $\bar{\Lambda} := (x, y, z \mapsto -\mathrm{w}(\bar{z}))$ *be the* $(\alpha, \bar{\beta})$*-trace obtained from* Λ *by complex conjugation. Then* Λ *satisfies* (9) *if and only if* $\bar{\Lambda}$ *satisfies* (9)*.*

Step 2 follows from the fact that the numbers $\mu, k_x, k_y, \varepsilon_x, \varepsilon_y$ change sign under complex conjugation.

Step 3. *Assume* (10)*. If* $B \cap \mathbb{R} = \{x, y\}$ *then* Λ *satisfies* (9)*.*

In this case B is contained in the upper or lower closed half plane and the loop $A \cup B$ bounds a disc contained in the same half plane. By Step 1 we may assume that B is contained in the upper half space. Then $\varepsilon_x = 1$, $\varepsilon_y = -1$, and $\mu(\Lambda) = 1$. Moreover, the winding number w is one in the disc encircled by A and B and is zero in the complement of its closure. Since the intervals $(-\infty, 0)$ and $(0, \infty)$ are contained in this complement, we have $k_x = k_y = 0$. This proves Step 3.

Step 4. *Assume* (10) *and* $\#(B \cap \mathbb{R}) > 2$*, follow the arc of* B*, starting at* x*, and let* x' *be the next intersection point with* $\mathbb{R}$*. Assume* $x' < x$*, denote by* B' *the arc in* B *from* x' *to* y*, and let* $A' := [x', y]$ *(see Figure 1). If the* (α, β)*-trace* Λ' *with boundary* $\partial\Lambda' = (x', y, A', B')$ *satisfies* (9) *so does* Λ*.*

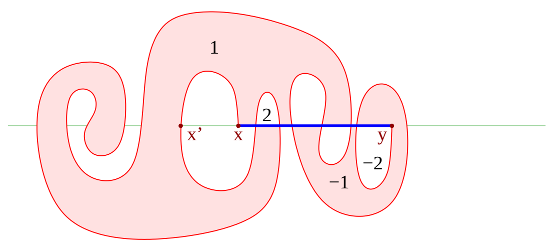

Figure 1: Maslov index and catenation: $x' < x < y$.

By Step 2 we may assume $\varepsilon_x(\Lambda) = 1$. Orient B from x to y. The Viterbo–Maslov index of Λ is minus the Maslov index of the path $B \to \mathbb{R}\mathrm{P}^1 : z \mapsto T_z B$, relative to the Lagrangian subspace $\mathbb{R} \subset \mathbb{C}$. Since the Maslov index of the arc in B from x to x' is $+1$ we have

$$\mu(\Lambda) = \mu(\Lambda') - 1. \tag{11}$$

Since the orientations of A' and B' agree with those of A and B we have

$$\varepsilon_{x'}(\Lambda') = \varepsilon_{x'}(\Lambda) = -1, \qquad \varepsilon_y(\Lambda') = \varepsilon_y(\Lambda). \tag{12}$$

Now let $x_1 < x_2 < \cdots < x_m < x$ be the intersection points of $\mathbb{R}$ and B in the interval $(-\infty, x)$ and let $\varepsilon_i \in \{-1, +1\}$ be the intersection index of $\mathbb{R}$ and B at x_i. Then there is an integer $\ell \in \{1, \ldots, m\}$ such that $x_\ell = x'$ and $\varepsilon_\ell = -1$. Moreover, the winding number w slightly to the left of x is

$$k_x(\Lambda) = \sum_{i=1}^{m} \varepsilon_i.$$

It agrees with the value of w slightly to the right of $x' = x_\ell$. Hence

$$k_x(\Lambda) = \sum_{i=1}^{\ell} \varepsilon_i = \sum_{i=1}^{\ell-1} \varepsilon_i - 1 = k_{x'}(\Lambda') - 1, \qquad k_y(\Lambda') = k_y(\Lambda). \tag{13}$$

It follows from equation (9) for Λ' and equations (11), (12), and (13) that

$$\begin{aligned}
\mu(\Lambda) &= \mu(\Lambda') - 1 \\
&= 2k_{x'}(\Lambda') + 2k_y(\Lambda') + \frac{\varepsilon_{x'}(\Lambda') - \varepsilon_y(\Lambda')}{2} - 1 \\
&= 2k_{x'}(\Lambda') + 2k_y(\Lambda') + \frac{-1 - \varepsilon_y(\Lambda)}{2} - 1 \\
&= 2k_{x'}(\Lambda') + 2k_y(\Lambda') + \frac{1 - \varepsilon_y(\Lambda)}{2} - 2 \\
&= 2k_x(\Lambda) + 2k_y(\Lambda) + \frac{\varepsilon_x(\Lambda) - \varepsilon_y(\Lambda)}{2}.
\end{aligned}$$

This proves Step 4.

Step 5. *Assume (10) and $\#(B \cap \mathbb{R}) > 2$, follow the arc of B, starting at x, and let x' be the next intersection point with $\mathbb{R}$. Assume $x < x' < y$, denote by B' the arc in B from x' to y, and let $A' := [x', y]$ (see Figure 2). If the (α, β)-trace Λ' with boundary $\partial\Lambda' = (x', y, A', B')$ satisfies (9) so does Λ.*

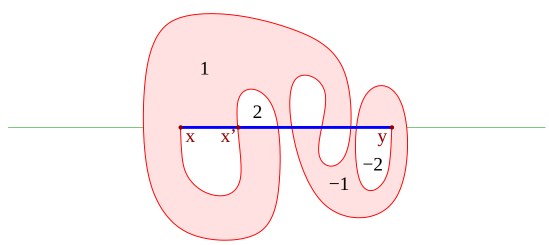

Figure 2: Maslov index and catenation: $x < x' < y$.

By Step 2 we may assume $\varepsilon_x(\Lambda) = 1$. Since the Maslov index of the arc in B from x to x' is -1, we have

$$\mu(\Lambda) = \mu(\Lambda') + 1. \tag{14}$$

Since the orientations of A' and B' agree with those of A and B we have

$$\varepsilon_{x'}(\Lambda') = \varepsilon_{x'}(\Lambda) = -1, \qquad \varepsilon_y(\Lambda') = \varepsilon_y(\Lambda). \tag{15}$$

Now let $x < x_1 < x_2 < \cdots < x_m < x'$ be the intersection points of $\mathbb{R}$ and B in the interval (x, x') and let $\varepsilon_i \in \{-1, +1\}$ be the intersection index of $\mathbb{R}$ and B at x_i. Since the interval $[x, x']$ in A and the arc in B from x to x' bound an open half disc, every subarc of B in this half disc must enter and exit through the open interval (x, x'). Hence the intersections indices of $\mathbb{R}$ and B at the points $x_1, \ldots, x_m$ cancel in pairs and thus

$$\sum_{i=1}^{m} \varepsilon_i = 0.$$

Since $k_{x'}(\Lambda')$ is the sum of the intersection indices of $\mathbb{R}$ and B' at all points to the left of x' we obtain

$$k_{x'}(\Lambda') = k_x(\Lambda) + \sum_{i=1}^{m} \varepsilon_i = k_x(\Lambda), \qquad k_y(\Lambda') = k_y(\Lambda). \tag{16}$$

It follows from equation (9) for Λ' and equations (14), (15), and (16) that

$$\begin{aligned} \mu(\Lambda) &= \mu(\Lambda') + 1 \\ &= 2k_{x'}(\Lambda') + 2k_y(\Lambda') + \frac{\varepsilon_{x'}(\Lambda') - \varepsilon_y(\Lambda')}{2} + 1 \\ &= 2k_x(\Lambda) + 2k_y(\Lambda) + \frac{-1 - \varepsilon_y(\Lambda)}{2} + 1 \\ &= 2k_x(\Lambda) + 2k_y(\Lambda) + \frac{\varepsilon_x(\Lambda) - \varepsilon_y(\Lambda)}{2}. \end{aligned}$$

This proves Step 5.

Step 6. *Assume (10) and $\#(B \cap \mathbb{R}) > 2$, follow the arc of B, starting at x, and let y' be the next intersection point with $\mathbb{R}$. Assume $y' > y$. Denote by B' the arc in B from y to y', and let $A' := [y, y']$ (see Figure 3). If the (α, β)-trace Λ' with boundary $\partial\Lambda' = (y, y', A', B')$ satisfies (9) so does Λ.*

By Step 2 we may assume $\varepsilon_x(\Lambda) = 1$. Since the orientation of B' from y to y' is opposite to the orientation of B and the Maslov index of the arc in B from x to y' is -1, we have

$$\mu(\Lambda) = 1 - \mu(\Lambda'). \tag{17}$$

Using again the fact that the orientation of B' is opposite to the orientation of B we have

$$\varepsilon_y(\Lambda') = -\varepsilon_y(\Lambda), \qquad \varepsilon_{y'}(\Lambda') = -\varepsilon_{y'}(\Lambda) = 1. \tag{18}$$

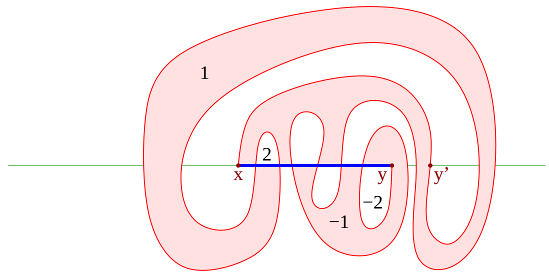

Figure 3: Maslov index and catenation: $x < y < y'$.

Now let $x_1 < x_2 < \cdots < x_m$ be all intersection points of $\mathbb{R}$ and B and let $\varepsilon_i \in \{-1, +1\}$ be the intersection index of $\mathbb{R}$ and B at x_i. Choose

$$j < k < \ell$$

such that

$$x_j = x, \qquad x_k = y, \qquad x_\ell = y'.$$

Then

$$\varepsilon_j = \varepsilon_x(\Lambda) = 1, \qquad \varepsilon_k = \varepsilon_y(\Lambda), \qquad \varepsilon_\ell = \varepsilon_{y'}(\Lambda) = -1,$$

and

$$k_x(\Lambda) = \sum_{i<j} \varepsilon_i, \qquad k_y(\Lambda) = -\sum_{i>k} \varepsilon_i.$$

For $i \neq j$ the intersection index of $\mathbb{R}$ and B' at x_i is $-\varepsilon_i$. Moreover, $k_y(\Lambda')$ is the sum of the intersection indices of $\mathbb{R}$ and B' at all points to the left of y and $k_{y'}(\Lambda')$ is minus the sum of the intersection indices of $\mathbb{R}$ and B' at all points to the right of y'. Hence

$$k_y(\Lambda') = -\sum_{i<j} \varepsilon_i - \sum_{j<i<k} \varepsilon_i, \qquad k_{y'}(\Lambda') = \sum_{i>\ell} \varepsilon_i.$$

We claim that

$$k_{y'}(\Lambda') + k_x(\Lambda) = 0, \qquad k_y(\Lambda') + k_y(\Lambda) = \frac{1 + \varepsilon_y(\Lambda)}{2}. \tag{19}$$

To see this, note that the value of the winding number w slightly to the left of x agrees with the value of w slightly to the right of y', and hence

$$0 = \sum_{i<j} \varepsilon_i + \sum_{i>\ell} \varepsilon_i = k_x(\Lambda) + k_{y'}(\Lambda').$$

This proves the first equation in (19). To prove the second equation in (19) we observe that

$$\sum_{i=1}^{m} \varepsilon_i = \frac{\varepsilon_x(\Lambda) + \varepsilon_y(\Lambda)}{2}$$

and hence

$$\begin{aligned}
k_y(\Lambda') + k_y(\Lambda) &= -\sum_{i<j} \varepsilon_i - \sum_{j<i<k} \varepsilon_i - \sum_{i>k} \varepsilon_i \\
&= \varepsilon_j + \varepsilon_k - \sum_{i=1}^{m} \varepsilon_i \\
&= \varepsilon_x(\Lambda) + \varepsilon_y(\Lambda) - \sum_{i=1}^{m} \varepsilon_i \\
&= \frac{\varepsilon_x(\Lambda) + \varepsilon_y(\Lambda)}{2} \\
&= \frac{1 + \varepsilon_y(\Lambda)}{2}.
\end{aligned}$$

This proves the second equation in (19).

It follows from equation (9) for Λ' and equations (17), (18), and (19) that

$$\begin{aligned}
\mu(\Lambda) &= 1 - \mu(\Lambda') \\
&= 1 - 2k_y(\Lambda') - 2k_{y'}(\Lambda') - \frac{\varepsilon_y(\Lambda') - \varepsilon_{y'}(\Lambda')}{2} \\
&= 1 - 2k_y(\Lambda') - 2k_{y'}(\Lambda') - \frac{-\varepsilon_y(\Lambda) - 1}{2} \\
&= 2k_y(\Lambda) - \varepsilon_y(\Lambda) + 2k_x(\Lambda) + \frac{1 + \varepsilon_y(\Lambda)}{2} \\
&= 2k_x(\Lambda) + 2k_y(\Lambda) + \frac{1 - \varepsilon_y(\Lambda)}{2}.
\end{aligned}$$

Here the first equality follows from (17), the second equality follows from (9) for Λ', the third equality follows from (18), and the fourth equality follows from (19). This proves Step 6.

Step 7. *The trace formula (7) holds when $\Sigma = \mathbb{C}$ and Λ satisfies the arc condition.*

It follows from Steps 3-6 by induction that equation (9) holds for every (α, β)-trace $\Lambda = (x, y, \mathrm{w})$ whose boundary $\partial\Lambda = (x, y, A, B)$ satisfies (10). Hence Step 7 follows from Step 1.

Next we drop the hypothesis that Λ satisfies the arc condition and extend the result to planar surfaces. This requires a further three steps.

Step 8. *The trace formula (7) holds when $\Sigma = \mathbb{C}$ and $x = y$.*

Under these hypotheses $\nu_\alpha := \partial \mathrm{w}|_{\alpha\setminus\beta}$ and $\nu_\beta := -\partial \mathrm{w}|_{\beta\setminus\alpha}$ are constant. There are four cases.

Case 1. *α is an embedded circle and β is not an embedded circle.* In this case we have $\nu_\beta \equiv 0$ and $B = \{x\}$. Moroeover, α is the boundary of a unique disc Δ_α and we assume that α is oriented as the boundary of Δ_α. Then the path $\gamma_\alpha : [0,1] \to \Sigma$ in Definition 2.1 satisfies $\gamma_\alpha(0) = \gamma_\alpha(1) = x$ and is homotopic to $\nu_\alpha\alpha$. Hence

$$m_x(\Lambda) = m_y(\Lambda) = 2\nu_\alpha = \mu(\Lambda).$$

Here the last equation follows from the fact that Λ can be obtained as the catenation of ν_α copies of the disc Δ_α.

Case 2. *α is not an embedded circle and β is an embedded circle.* This follows from Case 1 by interchanging α and β.

Case 3. *α and β are embedded circles.* In this case there is a unique pair of embedded discs Δ_α and Δ_β with boundaries α and β, respectively. Orient α and β as the boundaries of these discs. Then, for every $z \in \Sigma \setminus \alpha \cup \beta$, we have

$$\mathrm{w}(z) = \begin{cases} \nu_\alpha - \nu_\beta, & \text{for } z \in \Delta_\alpha \cap \Delta_\beta, \\ \nu_\alpha, & \text{for } z \in \Delta_\alpha \setminus \overline{\Delta}_\beta, \\ -\nu_\beta, & \text{for } z \in \Delta_\beta \setminus \overline{\Delta}_\alpha, \\ 0, & \text{for } z \in \Sigma \setminus \overline{\Delta}_\alpha \cup \overline{\Delta}_\beta. \end{cases}$$

Hence

$$m_x(\Lambda) = m_y(\Lambda) = 2\nu_\alpha - 2\nu_\beta = \mu(\Lambda).$$

Here the last equation follows from the fact Λ can be obtained as the catenation of ν_α copies of the disc Δ_α (with the orientation inherited from Σ) and ν_β copies of $-\Delta_\beta$ (with the opposite orientation).

Case 4. *Neither α nor β is an embedded circle.* Under this hypothesis we have $\nu_\alpha = \nu_\beta = 0$. Hence it follows from Theorem 2.4 that $\mathrm{w} = 0$ and $\Lambda = \Lambda_u$ for the constant map $u \equiv x \in \mathcal{D}(x,x)$. Thus

$$m_x(\Lambda) = m_y(\Lambda) = \mu(\Lambda) = 0.$$

This proves Step 8.

Step 9. *The trace formula* (7) *holds when* $\Sigma = \mathbb{C}$.

By Step 8, it suffices to assume $x \neq y$. It follows from Theorem 2.4 that every $u \in \mathcal{D}(x,y)$ is homotopic to a catentation $u = u_0 \# v$, where $u_0 \in \mathcal{D}(x,y)$ satisfies the arc condition and $v \in \mathcal{D}(y,y)$. Hence it follows from Steps 7 and 8 that

$$\begin{aligned}\mu(\Lambda_u) &= \mu(\Lambda_{u_0}) + \mu(\Lambda_v) \\ &= \frac{m_x(\Lambda_{u_0}) + m_y(\Lambda_{u_0})}{2} + m_y(\Lambda_v) \\ &= \frac{m_x(\Lambda_u) + m_y(\Lambda_u)}{2}.\end{aligned}$$

Here the last equation follows from the fact that $\mathrm{w}_u = \mathrm{w}_{u_0} + \mathrm{w}_v$ and hence $m_z(\Lambda_u) = m_z(\Lambda_{u_0}) + m_z(\Lambda_v)$ for every $z \in \alpha \cap \beta$. This proves Step 9.

Step 10. *The trace formula* (7) *holds when* Σ *is planar.*

Choose an element $u \in \mathcal{D}(x,y)$ such that $\Lambda_u = \Lambda$. Modifying α and β on the complement of $u(\mathbb{D})$, if necessary, we may assume without loss of generality that α and β are embedded circles. Let $\iota : \Sigma \to \mathbb{C}$ be an orientation preserving embedding. Then $\iota_*\Lambda := \Lambda_{\iota \circ u}$ is an $(\iota(\alpha), \iota(\beta))$-trace in $\mathbb{C}$ and hence satisfies the trace formula (7) by Step 9. Since $m_{\iota(x)}(\iota_*\Lambda) = m_x(\Lambda)$, $m_{\iota(y)}(\iota_*\Lambda) = m_y(\Lambda)$, and $\mu(\iota_*\Lambda) = \mu(\Lambda)$ it follows that Λ also satisfies the trace formula. This proves Step 10 and Proposition 4.1 □

Remark 4.2. Let $\Lambda = (x, y, A, B)$ be an (α, β)-trace in $\mathbb{C}$ as in Step 1 in the proof of Theorem 3.4. Thus $x < y$ are real numbers, A is the interval $[x, y]$, and B is an embedded arc with endpoints x, y which is oriented from x to y and is transverse to $\mathbb{R}$. Thus $Z := B \cap \mathbb{R}$ is a finite set. Define a map

$$f : Z \setminus \{y\} \to Z \setminus \{x\}$$

as follows. Given $z \in Z \setminus \{y\}$ walk along B towards y and let $f(z)$ be the next intersection point with $\mathbb{R}$. This map is bijective. Now let I be any of the three open intervals $(-\infty, x)$, (x, y), (y, ∞). Any arc in B from z to $f(z)$ with both endpoints in the same interval I can be removed by an isotopy of B which does not pass through x, y. Call Λ a **reduced** (α, β)**-trace** if $z \in I$ implies $f(z) \notin I$ for each of the three intervals. Then every (α, β)-trace is isotopic to a reduced (α, β')-trace and the isotopy does not affect the numbers $\mu, k_x, k_y, \varepsilon_x, \varepsilon_y$.

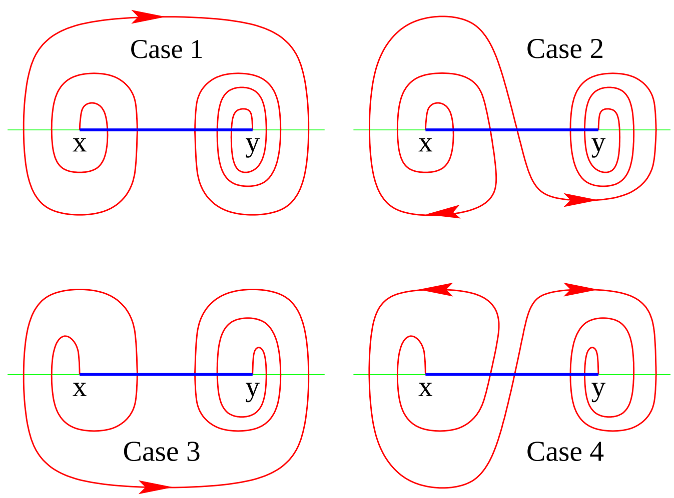

Figure 4: Reduced (α, β)-traces in $\mathbb{C}$.

Let Z^+ (respectively Z^-) denote the set of all points $z \in Z = B \cap \mathbb{R}$ where the positive tangent vectors in $T_z B$ point up (respectively down). One can prove that every reduced (α, β)-trace satisfies one of the following conditions.

Case 1: If $z \in Z^+ \setminus \{y\}$ then $f(z) > z$. **Case 2:** $Z^- \subset [x, y]$.
Case 3: If $z \in Z^- \setminus \{y\}$ then $f(z) > z$. **Case 4:** $Z^+ \subset [x, y]$.

(Examples with $\varepsilon_x = 1$ and $\varepsilon_y = -1$ are depicted in Figure 4.) One can then show directly that the reduced (α, β)-traces satisfy equation (9). This gives rise to an alternative proof of Proposition 4.1 via case distinction.

Proof of Theorem 3.4 in the Simply Connected Case. If Σ is diffeomorphic to the 2-plane the result has been established in Proposition 4.1. Hence assume

$$\Sigma = S^2.$$

Let $u \in \mathcal{D}(x, y)$. If u is not surjective the assertion follows from the case of the complex plane (Proposition 4.1) via stereographic projection. Hence assume u is surjective and choose a regular value $z \in S^2 \setminus (\alpha \cup \beta)$ of u. Denote

$$u^{-1}(z) = \{z_1, \ldots, z_k\}.$$

For $i = 1, \ldots, k$ let $\varepsilon_i = \pm 1$ according to whether or not the differential $du(z_i) : \mathbb{C} \to T_z\Sigma$ is orientation preserving. Choose an open disc $\Delta \subset S^2$ centered at z such that

$$\bar{\Delta} \cap (\alpha \cup \beta) = \emptyset$$

and $u^{-1}(\Delta)$ is a union of open neighborhoods $U_i \subset \mathbb{D}$ of z_i with disjoint closures such that

$$u|_{U_i} : U_i \to \Delta$$

is a diffeomorphism for each i which extends to a neighborhood of $\bar{U}_i$. Now choose a continuous map $u' : \mathbb{D} \to S^2$ which agrees with u on $\mathbb{D} \setminus \bigcup_i U_i$ and restricts to a diffeomorphism from $\bar{U}_i$ to $S^2 \setminus \Delta$ for each i. Then z does not belong to the image of u' and hence the trace formula (7) holds for u' (after smoothing along the boundaries ∂U_i). Moreover, the diffeomorphism

$$u'|_{\bar{U}_i} : \bar{U}_i \to S^2 \setminus \Delta$$

is orientation preserving if and only if $\varepsilon_i = -1$. Hence

$$\mu(\Lambda_u) = \mu(\Lambda_{u'}) + 4\sum_{i=1}^{k} \varepsilon_i,$$

$$m_x(\Lambda_u) = m_x(\Lambda_{u'}) + 4\sum_{i=1}^{k} \varepsilon_i,$$

$$m_y(\Lambda_u) = m_y(\Lambda_{u'}) + 4\sum_{i=1}^{k} \varepsilon_i.$$

By Proposition 4.1 the trace formula (7) holds for $\Lambda_{u'}$ and hence it also holds for Λ_u. This proves Theorem 3.4 when Σ is simply connected. □

5 The Non Simply Connected Case

The key step for extending Proposition 4.1 to non-simply connected two-manifolds is the next result about lifts to the universal cover.

Proposition 5.1. *Suppose Σ is not diffeomorphic to the 2-sphere. Let $\Lambda = (x, y, \mathrm{w})$ be an (α, β)-trace and $\pi : \mathbb{C} \to \Sigma$ be a universal covering. Denote by $\Gamma \subset \mathrm{Diff}(\mathbb{C})$ the group of deck transformations. Choose an element $\widetilde{x} \in \pi^{-1}(x)$ and let $\widetilde{\alpha}$ and $\widetilde{\beta}$ be the lifts of α and β through $\widetilde{x}$. Let $\widetilde{\Lambda} = (\widetilde{x}, \widetilde{y}, \widetilde{\mathrm{w}})$ be the lift of Λ with left endpoint $\widetilde{x}$. Then $\widetilde{\Lambda}$ satisfies the* **cancellation formula**

$$m_{g\widetilde{x}}(\widetilde{\Lambda}) + m_{g^{-1}\widetilde{y}}(\widetilde{\Lambda}) = 0 \qquad (20)$$

for every $g \in \Gamma \setminus \{\mathrm{id}\}$. (Proof on page 32.)

Lemma 5.2 (**Annulus Reduction**). *Suppose Σ is not diffeomorphic to the 2-sphere. Let Λ, π, Γ, $\widetilde{\Lambda}$ be as in Proposition 5.1. If*

$$m_{g\widetilde{x}}(\widetilde{\Lambda}) + m_{g^{-1}\widetilde{y}}(\widetilde{\Lambda}) = m_{g^{-1}\widetilde{x}}(\widetilde{\Lambda}) + m_{g\widetilde{y}}(\widetilde{\Lambda}) \tag{21}$$

for all $g \in \Gamma \setminus \{\mathrm{id}\}$ then the cancellation formula (20) holds for all $g \in \Gamma \setminus \{\mathrm{id}\}$.

Proof. If (20) does not hold then there is a deck transformation $h \in \Gamma \setminus \{\mathrm{id}\}$ such that $m_{h\widetilde{x}}(\widetilde{\Lambda}) + m_{h^{-1}\widetilde{y}}(\widetilde{\Lambda}) \neq 0$. Since there can only be finitely many such $h \in \Gamma \setminus \{\mathrm{id}\}$, there is an integer $k \geq 1$ such that $m_{h^k\widetilde{x}}(\widetilde{\Lambda}) + m_{h^{-k}\widetilde{y}}(\widetilde{\Lambda}) \neq 0$ and $m_{h^\ell\widetilde{x}}(\widetilde{\Lambda}) + m_{h^{-\ell}\widetilde{y}}(\widetilde{\Lambda}) = 0$ for every integer $\ell > k$. Define $g := h^k$. Then

$$m_{g\widetilde{x}}(\widetilde{\Lambda}) + m_{g^{-1}\widetilde{y}}(\widetilde{\Lambda}) \neq 0 \tag{22}$$

and $m_{g^k\widetilde{x}}(\widetilde{\Lambda}) + m_{g^{-k}\widetilde{y}}(\widetilde{\Lambda}) = 0$ for every integer $k \in \mathbb{Z} \setminus \{-1, 0, 1\}$. Define

$$\Sigma_0 := \mathbb{C}/\Gamma_0, \qquad \Gamma_0 := \left\{ g^k \,|\, k \in \mathbb{Z} \right\}.$$

Then Σ_0 is diffeomorphic to the annulus. Let $\pi_0 : \mathbb{C} \to \Sigma_0$ be the obvious projection, define $\alpha_0 := \pi_0(\widetilde{\alpha})$, $\beta_0 := \pi_0(\widetilde{\beta})$, and let $\Lambda_0 := (x_0, y_0, \mathrm{w}_0)$ be the (α_0, β_0)-trace in Σ_0 with $x_0 := \pi_0(\widetilde{x})$, $y_0 := \pi_0(\widetilde{y})$, and

$$\mathrm{w}_0(z_0) := \sum_{\widetilde{z} \in \pi_0^{-1}(z_0)} \widetilde{\mathrm{w}}(\widetilde{z}), \qquad z_0 \in \Sigma_0 \setminus (\alpha_0 \cup \beta_0).$$

Then

$$\begin{aligned} m_{x_0}(\Lambda_0) &= m_{\widetilde{x}}(\widetilde{\Lambda}) + \sum_{k \in \mathbb{Z} \setminus \{0\}} m_{g^k\widetilde{x}}(\widetilde{\Lambda}), \\ m_{y_0}(\Lambda_0) &= m_{\widetilde{y}}(\widetilde{\Lambda}) + \sum_{k \in \mathbb{Z} \setminus \{0\}} m_{g^{-k}\widetilde{y}}(\widetilde{\Lambda}). \end{aligned}$$

By Proposition 4.1 both $\widetilde{\Lambda}$ and Λ_0 satisfy the trace formula (7) and they have the same Viterbo–Maslov index. Hence

$$\begin{aligned} 0 &= \mu(\Lambda_0) - \mu(\widetilde{\Lambda}) \\ &= \frac{m_{x_0}(\Lambda_0) + m_{y_0}(\Lambda_0)}{2} - \frac{m_{\widetilde{x}}(\widetilde{\Lambda}) + m_{\widetilde{y}}(\widetilde{\Lambda})}{2} \\ &= \frac{1}{2} \sum_{k \neq 0} \Big(m_{g^k\widetilde{x}}(\widetilde{\Lambda}) + m_{g^{-k}\widetilde{y}}(\widetilde{\Lambda}) \Big) \\ &= m_{g\widetilde{x}}(\widetilde{\Lambda}) + m_{g^{-1}\widetilde{y}}(\widetilde{\Lambda}). \end{aligned}$$

Here the last equation follows from (21). This contradicts (22) and proves Lemma 5.2. □

Lemma 5.3. *Suppose Σ is not diffeomorphic to the 2-sphere. Let Λ, π, Γ, $\widetilde{\Lambda}$ be as in Proposition 5.1 and denote $\nu_{\widetilde{\alpha}} := \partial\widetilde{\mathrm{w}}|_{\widetilde{\alpha}\setminus\widetilde{\beta}}$ and $\nu_{\widetilde{\beta}} := -\partial\widetilde{\mathrm{w}}|_{\widetilde{\beta}\setminus\widetilde{\alpha}}$. Choose smooth paths*

$$\gamma_{\widetilde{\alpha}} : [0,1] \to \widetilde{\alpha}, \qquad \gamma_{\widetilde{\beta}} : [0,1] \to \widetilde{\beta}$$

from $\gamma_{\widetilde{\alpha}}(0) = \gamma_{\widetilde{\beta}}(0) = \widetilde{x}$ to $\gamma_{\widetilde{\alpha}}(1) = \gamma_{\widetilde{\beta}}(1) = \widetilde{y}$ such that $\gamma_{\widetilde{\alpha}}$ is an immersion when $\nu_{\widetilde{\alpha}} \not\equiv 0$ and constant when $\nu_{\widetilde{\alpha}} \equiv 0$, the same holds for $\gamma_{\widetilde{\beta}}$, and

$$\begin{aligned}
\nu_{\widetilde{\alpha}}(\widetilde{z}) &= \deg(\gamma_{\widetilde{\alpha}}, \widetilde{z}) \quad \textit{for} \quad \widetilde{z} \in \widetilde{\alpha} \setminus \{\widetilde{x}, \widetilde{y}\}, \\
\nu_{\widetilde{\beta}}(\widetilde{z}) &= \deg(\gamma_{\widetilde{\beta}}, \widetilde{z}) \quad \textit{for} \quad \widetilde{z} \in \widetilde{\beta} \setminus \{\widetilde{x}, \widetilde{y}\}.
\end{aligned}$$

Define

$$\widetilde{A} := \gamma_{\widetilde{\alpha}}([0,1]), \qquad \widetilde{B} := \gamma_{\widetilde{\beta}}([0,1]).$$

Then, for every $g \in \Gamma$, we have

$$g\widetilde{x} \in \widetilde{A} \qquad \Longleftrightarrow \qquad g^{-1}\widetilde{y} \in \widetilde{A}, \tag{23}$$

$$g\widetilde{x} \notin \widetilde{A} \quad \textit{and} \quad g\widetilde{y} \notin \widetilde{A} \qquad \Longleftrightarrow \qquad \widetilde{A} \cap g\widetilde{A} = \emptyset, \tag{24}$$

$$g\widetilde{x} \in \widetilde{A} \quad \textit{and} \quad g\widetilde{y} \in \widetilde{A} \qquad \Longleftrightarrow \qquad g = \mathrm{id}. \tag{25}$$

The same holds with $\widetilde{A}$ replaced by $\widetilde{B}$.

Proof. If α is a contractible embedded circle or not an embedded circle at all we have $\widetilde{A} \cap g\widetilde{A} = \emptyset$ whenever $g \neq \mathrm{id}$ and this implies (23), (24) and (25). Hence assume α is a noncontractible embedded circle. Then we may also assume, without loss of generality, that $\pi(\mathbb{R}) = \alpha$, the map $\widetilde{z} \mapsto \widetilde{z} + 1$ is a deck transformation, π maps the interval $[0,1)$ bijectively onto α, and $\widetilde{x}, \widetilde{y} \in \mathbb{R} = \widetilde{\alpha}$ with $\widetilde{x} < \widetilde{y}$. Thus $\widetilde{A} = [\widetilde{x}, \widetilde{y}]$ and, for every $k \in \mathbb{Z}$,

$$\widetilde{x} + k \in [\widetilde{x}, \widetilde{y}] \qquad \Longleftrightarrow \qquad 0 \leq k \leq \widetilde{y} - \widetilde{x} \qquad \Longleftrightarrow \qquad \widetilde{y} - k \in [\widetilde{x}, \widetilde{y}].$$

Similarly, we have

$$\widetilde{x} + k, \widetilde{y} + k \notin [\widetilde{x}, \widetilde{y}] \qquad \Longleftrightarrow \qquad [\widetilde{x} + k, \widetilde{y} + k] \cap [\widetilde{x}, \widetilde{y}] = \emptyset$$

and

$$\widetilde{x} + k, \widetilde{y} + k \in [\widetilde{x}, \widetilde{y}] \qquad \Longleftrightarrow \qquad [\widetilde{x} + k, \widetilde{y} + k] \subset [\widetilde{x}, \widetilde{y}] \qquad \Longleftrightarrow \qquad k = 0.$$

This proves (23), (24), and (25) for the deck transformation $\widetilde{z} \mapsto \widetilde{z} + k$. If g is any other deck transformation, then we have $\widetilde{\alpha} \cap g\widetilde{\alpha} = \emptyset$ and so (23), (24), and (25) are trivially satisfied. This proves Lemma 5.3. □

Lemma 5.4 (**Winding Number Comparison**). *Suppose Σ is not diffeomorphic to the 2-sphere. Let Λ, π, Γ, $\widetilde{\Lambda}$ be as in Proposition 5.1, and let $\widetilde{A}, \widetilde{B} \subset \mathbb{C}$ be as in Lemma 5.3. Then the following holds.*

(i) *Equation (21) holds for every $g \in \Gamma$ that satisfies $g\widetilde{x}, g\widetilde{y} \notin \widetilde{A} \cup \widetilde{B}$.*

(ii) *If Λ satisfies the arc condition then it also satisfies the cancellation formula (20) for every $g \in \Gamma \setminus \{\mathrm{id}\}$.*

Proof. We prove (i). Let $g \in \Gamma$ such that $g\widetilde{x}, g\widetilde{y} \notin \widetilde{A} \cup \widetilde{B}$ and let $\gamma_{\widetilde{\alpha}}, \gamma_{\widetilde{\beta}}$ be as in Lemma 5.3. Then $\widetilde{\mathrm{w}}(\widetilde{z})$ is the winding number of the loop $\gamma_{\widetilde{\alpha}} - \gamma_{\widetilde{\beta}}$ about the point $\widetilde{z} \in \mathbb{C} \setminus (\widetilde{A} \cup \widetilde{B})$. Moreover, the paths $g\gamma_{\widetilde{\alpha}}, g\gamma_{\widetilde{\beta}} : [0,1] \to \mathbb{C}$ connect the points $g\widetilde{x}, g\widetilde{y} \in \mathbb{C} \setminus (\widetilde{A} \cup \widetilde{B})$. Hence

$$\widetilde{\mathrm{w}}(g\widetilde{y}) - \widetilde{\mathrm{w}}(g\widetilde{x}) = (\gamma_{\widetilde{\alpha}} - \gamma_{\widetilde{\beta}}) \cdot g\gamma_{\widetilde{\alpha}} = (\gamma_{\widetilde{\alpha}} - \gamma_{\widetilde{\beta}}) \cdot g\gamma_{\widetilde{\beta}}.$$

Similarly with g replaced by g^{-1}. Moreover, it follows from Lemma 5.3, that

$$\widetilde{A} \cap g\widetilde{A} = \emptyset, \qquad \widetilde{B} \cap g^{-1}\widetilde{B} = \emptyset.$$

Hence

$$\begin{aligned}
\widetilde{\mathrm{w}}(g\widetilde{y}) - \widetilde{\mathrm{w}}(g\widetilde{x}) &= \Big(\gamma_{\widetilde{\alpha}} - \gamma_{\widetilde{\beta}}\Big) \cdot g\gamma_{\widetilde{\alpha}} \\
&= g\gamma_{\widetilde{\alpha}} \cdot \gamma_{\widetilde{\beta}} \\
&= \gamma_{\widetilde{\alpha}} \cdot g^{-1}\gamma_{\widetilde{\beta}} \\
&= \Big(\gamma_{\widetilde{\alpha}} - \gamma_{\widetilde{\beta}}\Big) \cdot g^{-1}\gamma_{\widetilde{\beta}} \\
&= \widetilde{\mathrm{w}}(g^{-1}\widetilde{y}) - \widetilde{\mathrm{w}}(g^{-1}\widetilde{x})
\end{aligned}$$

Here we have used the fact that every $g \in \Gamma$ is an orientation preserving diffeomorphism of $\mathbb{C}$. Thus we have proved that

$$\widetilde{\mathrm{w}}(g\widetilde{x}) + \widetilde{\mathrm{w}}(g^{-1}\widetilde{y}) = \widetilde{\mathrm{w}}(g\widetilde{y}) + \widetilde{\mathrm{w}}(g^{-1}\widetilde{x}).$$

Since $g\widetilde{x}, g\widetilde{y} \notin \widetilde{A} \cup \widetilde{B}$, we have

$$m_{g\widetilde{x}}(\widetilde{\Lambda}) = 4\widetilde{\mathrm{w}}(g\widetilde{x}), \qquad m_{g^{-1}\widetilde{y}}(\widetilde{\Lambda}) = 4\widetilde{\mathrm{w}}(g^{-1}\widetilde{y}),$$

and the same identities hold with g replaced by g^{-1}. This proves (i).

We prove (ii). If Λ satisfies the arc condition then $g\widetilde{A} \cap \widetilde{A} = \emptyset$ and $g\widetilde{B} \cap \widetilde{B} = \emptyset$ for every $g \in \Gamma \setminus \{\mathrm{id}\}$. In particular, for every $g \in \Gamma \setminus \{\mathrm{id}\}$, we have $g\widetilde{x}, g\widetilde{y} \notin \widetilde{A} \cup \widetilde{B}$ and hence (21) holds by (i). Hence it follows from Lemma 5.2 that the cancellation formula (20) holds for every $g \in \Gamma \setminus \{\mathrm{id}\}$. This proves Lemma 5.4. □

The next lemma deals with (α,β)-traces connecting a point $x \in \alpha\cap\beta$ to itself. An example on the annulus is depicted in Figure 5.

Lemma 5.5 (Isotopy Argument). *Suppose Σ is not diffeomorphic to the 2-sphere. Let Λ, π, Γ, $\widetilde{\Lambda}$ be as in Proposition 5.1. Suppose that there is a deck transformation $g_0 \in \Gamma\setminus\{\mathrm{id}\}$ such that $\widetilde{y} = g_0\widetilde{x}$. Then Λ has Viterbo–Maslov index zero and $m_{g\widetilde{x}}(\widetilde{\Lambda}) = 0$ for every $g \in \Gamma\setminus\{\mathrm{id}, g_0\}$.*

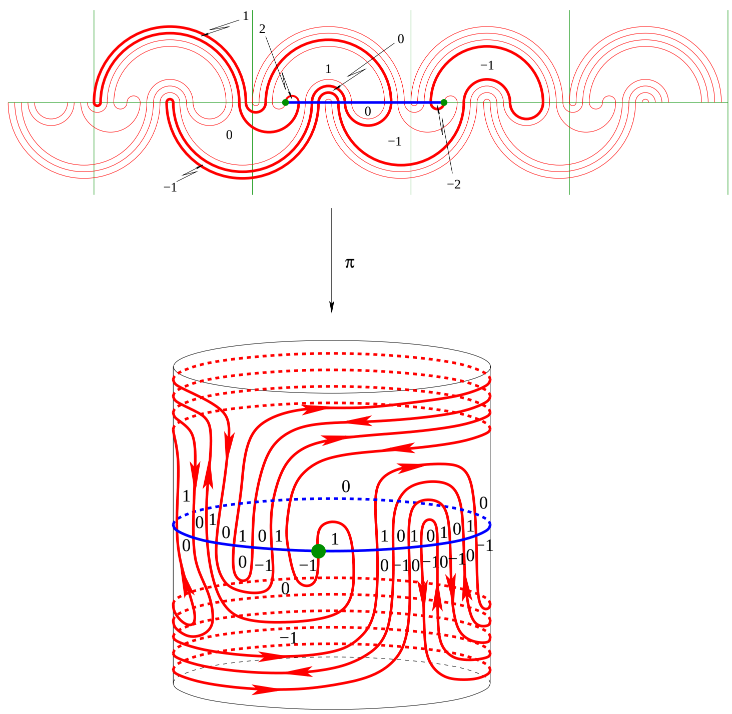

Figure 5: An (α,β)-trace on the annulus with $x = y$.

Proof. By hypothesis, we have $\widetilde{\alpha} = g_0\widetilde{\alpha}$ and $\widetilde{\beta} = g_0\widetilde{\beta}$. Hence α and β are noncontractible embedded circles and some iterate of α is homotopic to some iterate of β. Hence, by Lemma A.4, α must be homotopic to β (with some orientation). Hence we may assume, without loss of generality, that $\pi(\mathbb{R}) = \alpha$, the map $\widetilde{z}\mapsto\widetilde{z}+1$ is a deck transformation, π maps the interval $[0,1)$ bijectively onto α, $\mathbb{R} = \widetilde{\alpha}$, $\widetilde{x} = 0 \in \widetilde{\alpha}\cap\widetilde{\beta}$, $\widetilde{\beta} = \widetilde{\beta}+1$, and that $\widetilde{y} = \ell > 0$ is an integer. Then g_0 is the translation

$$g_0(\widetilde{z}) = \widetilde{z} + \ell.$$

Let $\widetilde{A} := [0,\ell] \subset \widetilde{\alpha}$ and let $\widetilde{B} \subset \widetilde{\beta}$ be the arc connecting 0 to ℓ. Then, for $\widetilde{z} \in \mathbb{C} \setminus (\widetilde{A} \cup \widetilde{B})$, the integer $\widetilde{\mathrm{w}}(\widetilde{z})$ is the winding number of $\widetilde{A} - \widetilde{B}$ about $\widetilde{z}$. Define the projection $\pi_0 : \mathbb{C} \to \mathbb{C}$ by

$$\pi_0(\widetilde{z}) := e^{2\pi \mathbf{i} \widetilde{z}/\ell},$$

denote $\alpha_0 := \pi_0(\widetilde{\alpha}) = S^1$ and $\beta_0 := \pi(\widetilde{\beta})$, and let $\Lambda_0 = (1, 1, \mathrm{w}_0)$ be the induced (α_0, β_0)-trace in $\mathbb{C}$ with $\mathrm{w}_0(z) := \sum_{\widetilde{z} \in \pi^{-1}(z)} \widetilde{\mathrm{w}}(\widetilde{z})$. Then Λ_0 satisfies the conditions of Step 8, Case 3 in the proof of Proposition 4.1 and its boundary is given by $\nu_{\alpha_0} = \partial \mathrm{w}_0|_{\alpha_0 \setminus \beta_0} \equiv 1$ and $\nu_{\beta_0} = \partial \mathrm{w}_0|_{\beta_0 \setminus \alpha_0} \equiv 1$. Hence Λ_0 and $\widetilde{\Lambda}$ have Viterbo–Maslov index zero.

It remains to prove that $m_{g\widetilde{x}}(\widetilde{\Lambda}) = 0$ for every $g \in \Gamma \setminus \{\mathrm{id}, g_0\}$. To see this we use the fact that the embedded loops α and β are homotopic with fixed endpoint x. Hence, by a Theorem of Epstein, they are isotopic with fixed basepoint x (see [8, Theorem 4.1]). Thus there exists a smooth map $f : \mathbb{R}/\mathbb{Z} \times [0,1] \to \Sigma$ such that

$$f(s,0) \in \alpha, \qquad f(s,1) \in \beta, \qquad f(0,t) = x,$$

for all $s \in \mathbb{R}/\mathbb{Z}$ and $t \in [0,1]$, and the map $\mathbb{R}/\mathbb{Z} \to \Sigma : s \mapsto f(s,t)$ is an embedding for every $t \in [0,1]$. Lift this homotopy to the universal cover to obtain a map $\widetilde{f} : \mathbb{R} \times [0,1] \to \mathbb{C}$ such that $\pi \circ \widetilde{f} = f$ and

$$\widetilde{f}(s,0) \in [0,1], \quad \widetilde{f}(s,1) \in \widetilde{B}_1, \quad \widetilde{f}(0,t) = \widetilde{x}, \quad \widetilde{f}(s+k,t) = \widetilde{f}(s,t) + k$$

for all $s,t \in [0,1]$ and $k \in \mathbb{Z}$. Here $\widetilde{B}_1 \subset \widetilde{B}$ denotes the arc in $\widetilde{B}$ from 0 to 1. Since the map $\mathbb{R}/\mathbb{Z} \to \Sigma : s \mapsto f(s,t)$ is injective for every t, we have

$$g\widetilde{x} \notin \{\widetilde{x}, \widetilde{x}+1, \ldots, \widetilde{x}+\ell\} \qquad \Longrightarrow \qquad g\widetilde{x} \notin \widetilde{f}([0,\ell] \times [0,1])$$

for every every $g \in \Gamma$. Now choose a smooth map $\widetilde{u} : \mathbb{D} \to \mathbb{C}$ with $\Lambda_{\widetilde{u}} = \widetilde{\Lambda}$ (see Theorem 2.4). Define the homotopy $F_{\widetilde{u}} : [0,\ell] \times [0,1] \to \mathbb{C}$ by $F_{\widetilde{u}}(s,t) := \widetilde{u}(-\cos(\pi s/\ell), t\sin(\pi s/\ell))$. Then, by Theorem 2.4, $F_{\widetilde{u}}$ is homotopic to $\widetilde{f}|_{[0,\ell]\times[0,1]}$ subject to the boundary conditions $\widetilde{f}(s,0) \in \widetilde{\alpha} = \mathbb{R}$, $\widetilde{f}(s,1) \in \widetilde{\beta}$, $\widetilde{f}(0,t) = \widetilde{x}$, $\widetilde{f}(\ell,t) = \widetilde{y}$. Hence, for every $\widetilde{z} \in \mathbb{C} \setminus (\widetilde{\alpha} \cup \widetilde{\beta})$, we have

$$\widetilde{\mathrm{w}}(\widetilde{z}) = \deg(\widetilde{u}, z) = \deg(F_{\widetilde{u}}, \widetilde{z}) = \deg(\widetilde{f}, \widetilde{z}).$$

In particular, choosing $\widetilde{z}$ near $g\widetilde{x}$, we find $m_{g\widetilde{x}}(\widetilde{\Lambda}) = 4\deg(\widetilde{f}, g\widetilde{x}) = 0$ for every $g \in \Gamma$ that is not one of the translations $\widetilde{z} \mapsto \widetilde{z} + k$ for $k = 0, 1, \ldots, \ell$. This proves the assertion in the case $\ell = 1$.

If $\ell > 1$ it remains to prove $m_k(\widetilde{\Lambda}) = 0$ for $k = 1, \dots, \ell - 1$. To see this, let $\widetilde{A}_1 := [0,1]$, $\widetilde{B}_1 \subset \widetilde{B}$ be the arc from 0 to 1, $\widetilde{\mathrm{w}}_1(\widetilde{z})$ be the winding number of $\widetilde{A}_1 - \widetilde{B}_1$ about $\widetilde{z} \in \mathbb{C} \setminus (\widetilde{A}_1 \cup \widetilde{B}_1)$, and define $\widetilde{\Lambda}_1 := (0, 1, \widetilde{\mathrm{w}}_1)$. Then, by what we have already proved, the $(\widetilde{\alpha}, \widetilde{\beta})$-trace $\widetilde{\Lambda}_1$ satisfies $m_{g\widetilde{x}}(\widetilde{\Lambda}_1) = 0$ for every $g \in \Gamma$ other than the translations by 0 or 1. In particular, we have $m_j(\widetilde{\Lambda}_1) = 0$ for every $j \in \mathbb{Z} \setminus \{0,1\}$ and also $m_0(\widetilde{\Lambda}_1) + m_1(\widetilde{\Lambda}_1) = 2\mu(\widetilde{\Lambda}_1) = 0$. Since $\widetilde{\mathrm{w}}(\widetilde{z}) = \sum_{j=0}^{\ell-1} \widetilde{\mathrm{w}}_1(\widetilde{z} - j)$ for $\widetilde{z} \in \mathbb{C} \setminus (\widetilde{A} \cup \widetilde{B})$, we obtain

$$m_k(\widetilde{\Lambda}) = \sum_{j=0}^{\ell-1} m_{k-j}(\widetilde{\Lambda}_1) = 0$$

for every $k \in \mathbb{Z} \setminus \{0, \ell\}$. This proves Lemma 5.5. $\square$

The next example shows that Lemma 5.4 cannot be strengthened to assert the identity $m_{g\widetilde{x}}(\widetilde{\Lambda}) = 0$ for every $g \in \Gamma$ with $g\widetilde{x}, g\widetilde{y} \notin \widetilde{A} \cup \widetilde{B}$.

Example 5.6. Figure 6 depicts an (α, β)-trace $\Lambda = (x, y, \mathrm{w})$ on the annulus $\Sigma = \mathbb{C}/\mathbb{Z}$ that has Viterbo–Maslov index one and satisfies the arc condition. The lift satisfies $m_{\widetilde{x}}(\widetilde{\Lambda}) = -3$, $m_{\widetilde{x}+1}(\widetilde{\Lambda}) = 4$, $m_{\widetilde{y}}(\widetilde{\Lambda}) = 5$, and $m_{\widetilde{y}-1}(\widetilde{\Lambda}) = -4$. Thus $m_x(\Lambda) = m_y(\Lambda) = 1$.

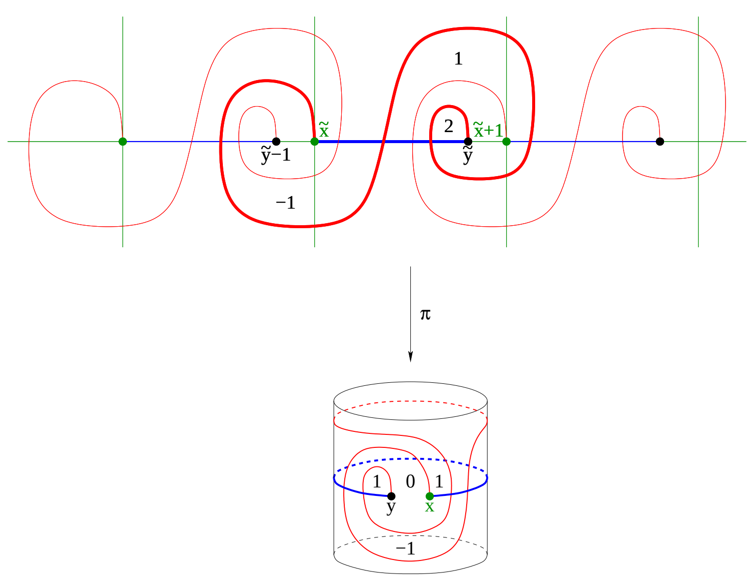

Figure 6: An (α, β)-trace on the annulus satisfying the arc condition.

Proof of Proposition 5.1. The proof has five steps.

Step 1. *Let $\widetilde{A}, \widetilde{B} \subset \mathbb{C}$ be as in Lemma 5.3 and let $g \in \Gamma$ such that*

$$g\widetilde{x} \in \widetilde{A} \setminus \widetilde{B}, \qquad g\widetilde{y} \notin \widetilde{A} \cup \widetilde{B}.$$

(An example is depicted in Figure 7.) Then (21) holds.

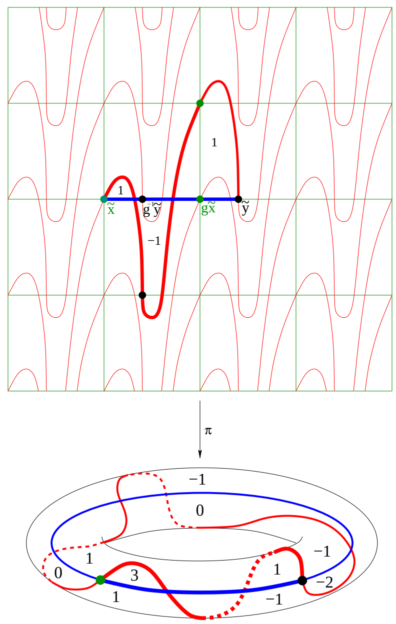

Figure 7: An (α, β)-trace on the torus not satisfying the arc condition.

The proof is a refinement of the winding number comparison argument in Lemma 5.4. Since $g\widetilde{x} \notin \widetilde{B}$ we have $g \neq \mathrm{id}$ and, since $\widetilde{x}, g\widetilde{x} \in \widetilde{A} \subset \widetilde{\alpha}$, it follows that α is a noncontractible embedded circle. Hence we may choose the universal covering $\pi : \mathbb{C} \to \Sigma$ and the lifts $\widetilde{\alpha}$, $\widetilde{\beta}$, $\widetilde{\Lambda}$ such that $\pi(\mathbb{R}) = \alpha$, the map $\widetilde{z} \mapsto \widetilde{z} + 1$ is a deck transformation, the projection π maps the interval $[0, 1)$ bijectively onto α, and

$$\widetilde{\alpha} = \mathbb{R}, \qquad \widetilde{x} = 0 \in \widetilde{\alpha} \cap \widetilde{\beta}, \qquad \widetilde{y} > 0.$$

By hypothesis and Lemma 5.3 there is an integer k such that

$$0 < k < \widetilde{y}, \qquad g\widetilde{x} = k, \qquad g^{-1}\widetilde{y} = \widetilde{y} - k.$$

Thus g is the deck transformation $\widetilde{z} \mapsto \widetilde{z} + k$.

Since $g\widetilde{x} \notin \widetilde{B}$ and $g\widetilde{y} \notin \widetilde{B}$ it follows from Lemma 5.3 that $g^{-1}\widetilde{y} \notin \widetilde{B}$ and $g^{-1}\widetilde{x} \notin \widetilde{B}$ and hence, again by Lemma 5.3, we have

$$\widetilde{B} \cap g\widetilde{B} = \widetilde{B} \cap g^{-1}\widetilde{B} = \emptyset.$$

With $\gamma_{\widetilde{\alpha}}$ and $\gamma_{\widetilde{\beta}}$ chosen as in Lemma 5.3, this implies

$$\gamma_{\widetilde{\beta}} \cdot (\gamma_{\widetilde{\beta}} - k) = (\gamma_{\widetilde{\beta}} + k) \cdot \gamma_{\widetilde{\beta}} = 0. \tag{26}$$

Since $k, -k, \widetilde{y}+k, \widetilde{y}-k \notin \widetilde{B}$, there exists a constant $\varepsilon > 0$ such that

$$-\varepsilon \le t \le \varepsilon \qquad \Longrightarrow \qquad k + \mathbf{i}t,\ \ -k+\mathbf{i}t,\ \ \widetilde{y}-k+\mathbf{i}t,\ \ \widetilde{y}+k+\mathbf{i}t \notin \widetilde{B}.$$

The paths $g\gamma_{\widetilde{\alpha}} \pm \mathbf{i}\varepsilon$ and $g\gamma_{\widetilde{\beta}} \pm \mathbf{i}\varepsilon$ both connect the point $g\widetilde{x} \pm \mathbf{i}\varepsilon$ to $g\widetilde{y} \pm \mathbf{i}\varepsilon$. Likewise, the paths $g^{-1}\gamma_{\widetilde{\alpha}} \pm \mathbf{i}\varepsilon$ and $g^{-1}\gamma_{\widetilde{\beta}} \pm \mathbf{i}\varepsilon$ both connect the point $g^{-1}\widetilde{x} \pm \mathbf{i}\varepsilon$ to $g^{-1}\widetilde{y} \pm \mathbf{i}\varepsilon$. Hence

$$\begin{aligned}
\widetilde{\mathrm{w}}(g\widetilde{y} \pm \mathbf{i}\varepsilon) - \widetilde{\mathrm{w}}(g\widetilde{x} \pm \mathbf{i}\varepsilon) &= (\gamma_{\widetilde{\alpha}} - \gamma_{\widetilde{\beta}}) \cdot (g\gamma_{\widetilde{\alpha}} \pm \mathbf{i}\varepsilon) \\
&= (\gamma_{\widetilde{\alpha}} - \gamma_{\widetilde{\beta}}) \cdot (\gamma_{\widetilde{\alpha}} + k \pm \mathbf{i}\varepsilon) \\
&= (\gamma_{\widetilde{\alpha}} + k \pm \mathbf{i}\varepsilon) \cdot \gamma_{\widetilde{\beta}} \\
&= \gamma_{\widetilde{\alpha}} \cdot (\gamma_{\widetilde{\beta}} - k \mp \mathbf{i}\varepsilon) \\
&= (\gamma_{\widetilde{\alpha}} - \gamma_{\widetilde{\beta}}) \cdot (\gamma_{\widetilde{\beta}} - k \mp \mathbf{i}\varepsilon) \\
&= (\gamma_{\widetilde{\alpha}} - \gamma_{\widetilde{\beta}}) \cdot (g^{-1}\gamma_{\widetilde{\beta}} \mp \mathbf{i}\varepsilon) \\
&= \widetilde{\mathrm{w}}(g^{-1}\widetilde{y} \mp \mathbf{i}\varepsilon) - \widetilde{\mathrm{w}}(g^{-1}\widetilde{x} \mp \mathbf{i}\varepsilon).
\end{aligned}$$

Here the last but one equation follows from (26). Thus we have proved

$$\begin{aligned}
\widetilde{\mathrm{w}}(g\widetilde{x} + \mathbf{i}\varepsilon) + \widetilde{\mathrm{w}}(g^{-1}\widetilde{y} - \mathbf{i}\varepsilon) = \widetilde{\mathrm{w}}(g^{-1}\widetilde{x} - \mathbf{i}\varepsilon) + \widetilde{\mathrm{w}}(g\widetilde{y} + \mathbf{i}\varepsilon), \\
\widetilde{\mathrm{w}}(g\widetilde{x} - \mathbf{i}\varepsilon) + \widetilde{\mathrm{w}}(g^{-1}\widetilde{y} + \mathbf{i}\varepsilon) = \widetilde{\mathrm{w}}(g^{-1}\widetilde{x} + \mathbf{i}\varepsilon) + \widetilde{\mathrm{w}}(g\widetilde{y} - \mathbf{i}\varepsilon).
\end{aligned} \tag{27}$$

Since

$$\begin{aligned}
m_{g\widetilde{x}}(\widetilde{\Lambda}) &= 2\widetilde{\mathrm{w}}(g\widetilde{x} + \mathbf{i}\varepsilon) + 2\widetilde{\mathrm{w}}(g\widetilde{x} - \mathbf{i}\varepsilon), \\
m_{g\widetilde{y}}(\widetilde{\Lambda}) &= 2\widetilde{\mathrm{w}}(g\widetilde{y} + \mathbf{i}\varepsilon) + 2\widetilde{\mathrm{w}}(g\widetilde{y} - \mathbf{i}\varepsilon), \\
m_{g^{-1}\widetilde{x}}(\widetilde{\Lambda}) &= 2\widetilde{\mathrm{w}}(g^{-1}\widetilde{x} + \mathbf{i}\varepsilon) + 2\widetilde{\mathrm{w}}(g^{-1}\widetilde{x} - \mathbf{i}\varepsilon), \\
m_{g^{-1}\widetilde{y}}(\widetilde{\Lambda}) &= 2\widetilde{\mathrm{w}}(g^{-1}\widetilde{y} + \mathbf{i}\varepsilon) + 2\widetilde{\mathrm{w}}(g^{-1}\widetilde{y} - \mathbf{i}\varepsilon),
\end{aligned}$$

Step 1 follows by taking the sum of the two equations in (27).

Step 2. *Let $\widetilde{A}, \widetilde{B} \subset \mathbb{C}$ be as in Lemma 5.3 and let $g \in \Gamma$. Suppose that either $g\widetilde{x}, g\widetilde{y} \notin \widetilde{A}$ or $g\widetilde{x}, g\widetilde{y} \notin \widetilde{B}$. Then (21) holds.*

If $g\widetilde{x}, g\widetilde{y} \notin \widetilde{A} \cup \widetilde{B}$ the assertion follows from Lemma 5.4. If $g\widetilde{x} \in \widetilde{A} \setminus \widetilde{B}$ and $g\widetilde{y} \notin \widetilde{A} \cup \widetilde{B}$ the assertion follows from Step 1. If $g\widetilde{x} \notin \widetilde{A} \cup \widetilde{B}$ and $g\widetilde{y} \in \widetilde{A} \setminus \widetilde{B}$ the assertion follows from Step 1 by interchanging $\widetilde{x}$ and $\widetilde{y}$. Namely, (21) holds for $\widetilde{\Lambda}$ if and only if it holds for the $(\widetilde{\alpha}, \widetilde{\beta})$-trace $-\widetilde{\Lambda} := (\widetilde{y}, \widetilde{x}, -\widetilde{\mathrm{w}})$. This covers the case $g\widetilde{x}, g\widetilde{y} \notin \widetilde{B}$. If $g\widetilde{x}, g\widetilde{y} \notin \widetilde{A}$ the assertion follows by interchanging $\widetilde{A}$ and $\widetilde{B}$. Namely, (21) holds for $\widetilde{\Lambda}$ if and only if it holds for the $(\widetilde{\beta}, \widetilde{\alpha})$-trace $\widetilde{\Lambda}^* := (\widetilde{x}, \widetilde{y}, -\widetilde{\mathrm{w}})$. This proves Step 2.

Step 3. *Let $\widetilde{A}, \widetilde{B} \subset \mathbb{C}$ be as in Lemma 5.3 and let $g \in \Gamma$ such that*

$$g\widetilde{x} \in \widetilde{A} \setminus \widetilde{B}, \qquad g\widetilde{y} \in \widetilde{B} \setminus \widetilde{A}.$$

(An example is depicted in Figure 8.) Then the cancellation formula (20) holds for g and g^{-1}.

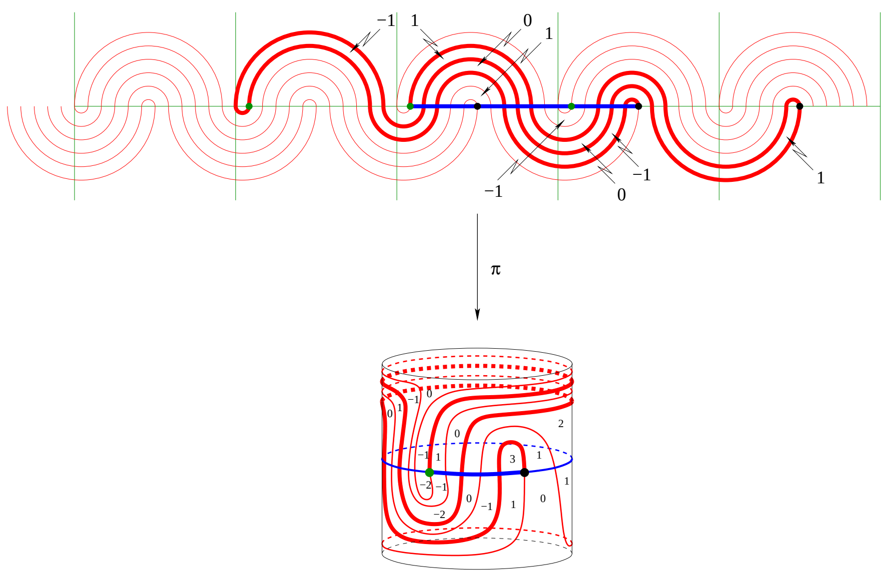

Figure 8: An (α, β)-trace on the annulus with $g\widetilde{x} \in \widetilde{A}$ and $g\widetilde{y} \in \widetilde{B}$.

Since $g\widetilde{x} \notin \widetilde{B}$ (and $g\widetilde{y} \notin \widetilde{A}$) we have $g \neq \mathrm{id}$ and, since $\widetilde{x}, g\widetilde{x} \in \widetilde{A} \subset \widetilde{\alpha}$ and $\widetilde{y}, g\widetilde{y} \in \widetilde{B} \subset \widetilde{\beta}$, it follows that $g\widetilde{\alpha} = \widetilde{\alpha}$ and $g\widetilde{\beta} = \widetilde{\beta}$. Hence α and β are noncontractible embedded circles and some iterate of α is homotopic to some iterate of β. So α is homotopic to β (with some orientation), by Lemma A.4.

Choose the universal covering $\pi : \mathbb{C} \to \Sigma$ and the lifts $\widetilde{\alpha}, \widetilde{\beta}, \widetilde{\Lambda}$ such that $\pi(\mathbb{R}) = \alpha$, the map $\widetilde{z} \mapsto \widetilde{z} + 1$ is a deck transformation, π maps the interval $[0, 1)$ bijectively onto α, and

$$\widetilde{\alpha} = \mathbb{R}, \qquad \widetilde{x} = 0 \in \widetilde{\alpha} \cap \widetilde{\beta}, \qquad \widetilde{y} > 0.$$

Thus $\widetilde{A} = [0, \widetilde{y}]$ is the arc in $\widetilde{\alpha}$ from 0 to $\widetilde{y}$ and $\widetilde{B}$ is the arc in $\widetilde{\beta}$ from 0 to $\widetilde{y}$. Moreover, since α is homotopic to β, we have

$$\widetilde{\beta} = \widetilde{\beta} + 1$$

and the arc in $\widetilde{\beta}$ from 0 to 1 is a fundamental domain for β. Since $g\widetilde{\alpha} = \widetilde{\alpha}$, the deck transformation g is given by $\widetilde{z} \mapsto \widetilde{z} + \ell$ for some integer ℓ. Since $g\widetilde{x} \in \widetilde{A} \setminus \widetilde{B}$ and $g\widetilde{y} \in \widetilde{B}$, we have $g^{-1}\widetilde{y} \notin \widetilde{B}$ and $g^{-1}\widetilde{x} \in \widetilde{B}$ by Lemma 5.3. Hence

$$0 < \ell < \widetilde{y}, \qquad \ell \notin \widetilde{B}, \qquad \widetilde{y} + \ell \in \widetilde{B}, \qquad \widetilde{y} - \ell \notin \widetilde{B}, \qquad -\ell \in \widetilde{B}.$$

This shows that, walking along $\widetilde{\beta}$ from 0 to $\widetilde{y}$ (traversing $\widetilde{B}$) one encounters some negative integer and therefore no positive integers. Hence

$$\widetilde{A} \cap \mathbb{Z} = \left\{0, 1, 2, \cdots, k_{\widetilde{A}}\right\}, \qquad \widetilde{B} \cap \mathbb{Z} = \left\{0, -1, -2, \cdots, -k_{\widetilde{B}}\right\},$$

where $k_{\widetilde{A}}$ is the number of fundamental domains of $\widetilde{\alpha}$ contained in $\widetilde{A}$ and $k_{\widetilde{B}}$ is the number of fundamental domains of $\widetilde{\beta}$ contained in $\widetilde{B}$ (see Figure 8). For $0 \le k \le k_{\widetilde{A}}$ let $\widetilde{A}_k \subset \widetilde{\alpha}$ and $\widetilde{B}_k \subset \widetilde{\beta}$ be the arcs from 0 to $\widetilde{y} - k$. Thus $\widetilde{A}_k$ is obtained from $\widetilde{A}$ by removing k fundamental domains at the end, while $\widetilde{B}_k$ is obtained from $\widetilde{B}$ by attaching k fundamental domains at the end. Consider the $(\widetilde{\alpha}, \widetilde{\beta})$-trace

$$\widetilde{\Lambda}_k := (0, \widetilde{y} - k, \widetilde{\mathrm{w}}_k), \qquad \partial\widetilde{\Lambda}_k := (0, \widetilde{y} - k, \widetilde{A}_k, \widetilde{B}_k),$$

where $\widetilde{\mathrm{w}}_k : \mathbb{C} \setminus (\widetilde{A}_k \cup \widetilde{B}_k) \to \mathbb{Z}$ is the winding number of $\widetilde{A}_k - \widetilde{B}_k$. Then

$$\widetilde{B}_k \cap \mathbb{Z} = \left\{0, -1, -2, \cdots, -k_{\widetilde{B}} - k\right\}$$

and $\widetilde{\Lambda}_0 = \widetilde{\Lambda}$. We prove that, for each k, the $(\widetilde{\alpha}, \widetilde{\beta})$-trace $\widetilde{\Lambda}_k$ satisfies

$$m_j(\widetilde{\Lambda}_k) + m_{\widetilde{y}-k-j}(\widetilde{\Lambda}_k) = 0 \qquad \forall\, j \in \mathbb{Z} \setminus \{0\}. \tag{28}$$

If $\widetilde{y}$ is an integer, then (28) follows from Lemma 5.5. Hence we may assume that $\widetilde{y}$ is not an integer.

We prove equation (28) by reverse induction on k. First let $k = k_{\widetilde{A}}$. Then we have $j, \widetilde{y} - k + j \notin \widetilde{A}_k$ for every $j \in \mathbb{N}$. Hence it follows from Step 2 that

$$m_j(\widetilde{\Lambda}_k) + m_{\widetilde{y}-k-j}(\widetilde{\Lambda}_k) = m_{-j}(\widetilde{\Lambda}_k) + m_{\widetilde{y}-k+j}(\widetilde{\Lambda}) \qquad \forall\, j \in \mathbb{N}. \tag{29}$$

Thus we can apply Lemma 5.2 to the projection of $\widetilde{\Lambda}_k$ to the quotient $\mathbb{C}/\mathbb{Z}$. Hence $\widetilde{\Lambda}_k$ satisfies (28).

Now fix an integer $k \in \{0, 1, \ldots, k_{\widetilde{A}} - 1\}$ and suppose, by induction, that $\widetilde{\Lambda}_{k+1}$ satisfies (28). Denote by $\widetilde{A}' \subset \widetilde{\alpha}$ and $\widetilde{B}' \subset \widetilde{\beta}$ the arcs from $\widetilde{y} - k - 1$ to 1, and by $\widetilde{A}'' \subset \widetilde{\alpha}$ and $\widetilde{B}'' \subset \widetilde{\beta}$ the arcs from 1 to $\widetilde{y} - k$. Then $\widetilde{\Lambda}_k$ is the catenation of the $(\widetilde{\alpha}, \widetilde{\beta})$-traces

$$\begin{aligned}
\widetilde{\Lambda}_{k+1} &:= (0, \widetilde{y} - k - 1, \widetilde{\mathrm{w}}_{k+1}), & \partial\widetilde{\Lambda}_{k+1} &= (0, \widetilde{y} - k - 1, \widetilde{A}_{k+1}, \widetilde{B}_{k+1}), \\
\widetilde{\Lambda}' &:= (\widetilde{y} - k - 1, 1, \widetilde{\mathrm{w}}'), & \partial\widetilde{\Lambda}' &= (\widetilde{y} - k - 1, 1, \widetilde{A}', \widetilde{B}'), \\
\widetilde{\Lambda}'' &:= (1, \widetilde{y} - k, \widetilde{\mathrm{w}}''), & \partial\widetilde{\Lambda}'' &= (1, \widetilde{y} - k, \widetilde{A}'', \widetilde{B}'').
\end{aligned}$$

Here $\widetilde{\mathrm{w}}'(\widetilde{z})$ is the winding number of the loop $\widetilde{A}' - \widetilde{B}'$ about $\widetilde{z} \in \mathbb{C} \setminus (\widetilde{A}' \cup \widetilde{B}')$ and simiarly for $\widetilde{\mathrm{w}}''$. Note that $\widetilde{\Lambda}''$ is the shift of $\widetilde{\Lambda}_{k+1}$ by 1. The catenation of $\widetilde{\Lambda}_{k+1}$ and $\widetilde{\Lambda}'$ is the $(\widetilde{\alpha}, \widetilde{\beta})$-trace from 0 to 1. Hence it has Viterbo–Maslov index zero, by Lemma 5.5, and satisfies

$$m_j(\widetilde{\Lambda}_{k+1}) + m_j(\widetilde{\Lambda}') = 0 \qquad \forall j \in \mathbb{Z} \setminus \{0, 1\}. \tag{30}$$

Since the catenation of $\widetilde{\Lambda}'$ and $\widetilde{\Lambda}''$ is the $(\widetilde{\alpha}, \widetilde{\beta})$-trace from $\widetilde{y} - k - 1$ to $\widetilde{y} - k$, it also has Viterbo–Maslov index zero and satisfies

$$m_{\widetilde{y}-k-j}(\widetilde{\Lambda}') + m_{\widetilde{y}-k-j}(\widetilde{\Lambda}'') = 0 \qquad \forall j \in \mathbb{Z} \setminus \{0, 1\}. \tag{31}$$

Moreover, by the induction hypothesis, we have

$$m_j(\widetilde{\Lambda}_{k+1}) + m_{\widetilde{y}-k-1-j}(\widetilde{\Lambda}_{k+1}) = 0 \qquad \forall j \in \mathbb{Z} \setminus \{0\}. \tag{32}$$

Combining the equations (30), (31), (32) we find that, for $j \in \mathbb{Z} \setminus \{0, 1\}$,

$$\begin{aligned}
m_j(\widetilde{\Lambda}_k) + m_{\widetilde{y}-k-j}(\widetilde{\Lambda}_k) &= m_j(\widetilde{\Lambda}_{k+1}) + m_j(\widetilde{\Lambda}') + m_j(\widetilde{\Lambda}'') \\
&\quad + m_{\widetilde{y}-k-j}(\widetilde{\Lambda}_{k+1}) + m_{\widetilde{y}-k-j}(\widetilde{\Lambda}') + m_{\widetilde{y}-k-j}(\widetilde{\Lambda}'') \\
&= m_j(\widetilde{\Lambda}_{k+1}) + m_j(\widetilde{\Lambda}') \\
&\quad + m_{\widetilde{y}-k-j}(\widetilde{\Lambda}') + m_{\widetilde{y}-k-j}(\widetilde{\Lambda}'') \\
&\quad + m_{j-1}(\widetilde{\Lambda}_{k+1}) + m_{\widetilde{y}-k-j}(\widetilde{\Lambda}_{k+1}) \\
&= 0.
\end{aligned}$$

For $j = 1$ we obtain

$$\begin{aligned}
m_1(\widetilde{\Lambda}_k) + m_{\widetilde{y}-k-1}(\widetilde{\Lambda}_k) &= m_1(\widetilde{\Lambda}_{k+1}) + m_1(\widetilde{\Lambda}') + m_1(\widetilde{\Lambda}'') \\
&\quad + m_{\widetilde{y}-k-1}(\widetilde{\Lambda}_{k+1}) + m_{\widetilde{y}-k-1}(\widetilde{\Lambda}') + m_{\widetilde{y}-k-1}(\widetilde{\Lambda}'') \\
&= m_1(\widetilde{\Lambda}_{k+1}) + m_{\widetilde{y}-k-2}(\widetilde{\Lambda}_{k+1}) \\
&\quad + m_0(\widetilde{\Lambda}_{k+1}) + m_{\widetilde{y}-k-1}(\widetilde{\Lambda}_{k+1}) \\
&\quad + m_{\widetilde{y}-k-1}(\widetilde{\Lambda}') + m_1(\widetilde{\Lambda}') \\
&= 2\mu(\widetilde{\Lambda}_{k+1}) + 2\mu(\widetilde{\Lambda}') \\
&= 0.
\end{aligned}$$

Here the last but one equation follows from equation (32) and Proposition 4.1, and the last equation follows from Lemma 5.5. Hence $\widetilde{\Lambda}_k$ satisfies (28). This completes the induction argument for the proof of Step 3.

Step 4. *Let $\widetilde{A}, \widetilde{B} \subset \mathbb{C}$ be as in Lemma 5.3 and let $g \in \Gamma$ such that*

$$g\widetilde{x} \in \widetilde{A} \cap \widetilde{B}, \qquad g\widetilde{y} \notin \widetilde{A} \cup \widetilde{B}.$$

Then the cancellation formula (20) holds for g and g^{-1}.

The proof is by induction and catenation based on Step 2 and Lemma 5.5. Since $g\widetilde{y} \notin \widetilde{A} \cup \widetilde{B}$ we have $g \neq \mathrm{id}$. Since $g\widetilde{x} \in \widetilde{A} \cap \widetilde{B}$ we have $\widetilde{\alpha} = g\widetilde{\alpha}$ and $\widetilde{\beta} = g\widetilde{\beta}$. Hence α and β are noncontractible embedded circles, and they are homotopic (with some orientation) by Lemma A.4. Thus we may choose $\pi : \mathbb{C} \to \Sigma$, $\widetilde{\alpha}$, $\widetilde{\beta}$, $\widetilde{\Lambda}$ as in Step 3. By hypothesis there is an integer $k \in \widetilde{A} \cap \widetilde{B}$. Hence $\widetilde{A}$ and $\widetilde{B}$ do not contain any negative integers. Choose $k_{\widetilde{A}}, k_{\widetilde{B}} \in \mathbb{N}$ such that

$$\widetilde{A} \cap \mathbb{Z} = \left\{0, 1, \ldots, k_{\widetilde{A}}\right\}, \qquad \widetilde{B} \cap \mathbb{Z} = \left\{0, 1, \ldots, k_{\widetilde{B}}\right\}.$$

Assume without loss of generality that

$$k_{\widetilde{A}} \leq k_{\widetilde{B}}.$$

For $0 \leq k \leq k_{\widetilde{A}}$ denote by $\widetilde{A}_k \subset \widetilde{A}$ and $\widetilde{B}_k \subset \widetilde{B}$ the arcs from 0 to $\widetilde{y} - k$ and consider the $(\widetilde{\alpha}, \widetilde{\beta})$-trace

$$\widetilde{\Lambda}_k := (0, \widetilde{y} - k, \widetilde{\mathrm{w}}_k), \qquad \partial\widetilde{\Lambda}_k := (0, \widetilde{y} - k, \widetilde{A}_k, \widetilde{B}_k).$$

In this case

$$\widetilde{B}_k \cap \mathbb{Z} = \{0, 1, \dots, k_{\widetilde{B}} - k\}.$$

As in Step 3, it follows by reverse induction on k that $\widetilde{\Lambda}_k$ satisfies (28) for every k. We assume again that $\widetilde{y}$ is not an integer. (Otherwise (28) follows from Lemma 5.5). If $k = k_{\widetilde{A}}$ then $j, \widetilde{y} - k + j \notin \widetilde{A}_k$ for every $j \in \mathbb{N}$, hence it follows from Step 2 that $\widetilde{\Lambda}_k$ satisfies (29), and hence it follows from Lemma 5.2 for the projection of $\widetilde{\Lambda}_k$ to the annulus $\mathbb{C}/\mathbb{Z}$ that $\widetilde{\Lambda}_k$ also satisfies (28). The induction step is verbatim the same as in Step 3 and will be omitted. This proves Step 4.

Step 5. *We prove Proposition 5.1.*

If both points $g\widetilde{x}, g\widetilde{y}$ are contained in $\widetilde{A}$ (or in $\widetilde{B}$) then $g = \mathrm{id}$ by Lemma 5.3, and in this case equation (21) is a tautology. If both points $g\widetilde{x}, g\widetilde{y}$ are not contained in $\widetilde{A} \cup \widetilde{B}$, equation (21) has been established in Lemma 5.4. Moreover, we can interchange $\widetilde{x}$ and $\widetilde{y}$ or $\widetilde{A}$ and $\widetilde{B}$ as in the proof of Step 2. Thus Steps 1 and 4 cover the case where precisely one of the points $g\widetilde{x}, g\widetilde{y}$ is contained in $\widetilde{A} \cup \widetilde{B}$ while Step 3 covers the case where $g \neq \mathrm{id}$ and both points $g\widetilde{x}, g\widetilde{y}$ are contained in $\widetilde{A} \cup \widetilde{B}$. This shows that equation (21) holds for every $g \in \Gamma \setminus \{\mathrm{id}\}$. Hence, by Lemma 5.2, the cancellation formula (20) holds for every $g \in \Gamma \setminus \{\mathrm{id}\}$. This proves Proposition 5.1. □

Proof of Theorem 3.4 in the Non Simply Connected Case. Choose a universal covering $\pi : \mathbb{C} \to \Sigma$ and let Γ, $\widetilde{\alpha}$, $\widetilde{\beta}$, and $\widetilde{\Lambda} = (\widetilde{x}, \widetilde{y}, \widetilde{\mathrm{w}})$ be as in Proposition 5.1. Then

$$m_x(\Lambda) + m_y(\Lambda) - m_{\widetilde{x}}(\widetilde{\Lambda}) - m_{\widetilde{y}}(\widetilde{\Lambda}) = \sum_{g \neq \mathrm{id}} \Big(m_{g\widetilde{x}}(\widetilde{\Lambda}) + m_{g^{-1}\widetilde{y}}(\widetilde{\Lambda}) \Big) = 0.$$

Here the last equation follows from the cancellation formula in Proposition 5.1. Hence, by Proposition 4.1, we have

$$\mu(\Lambda) = \mu(\widetilde{\Lambda}) = \frac{m_{\widetilde{x}}(\widetilde{\Lambda}) + m_{\widetilde{y}}(\widetilde{\Lambda})}{2} = \frac{m_x(\Lambda) + m_y(\Lambda)}{2}.$$

This proves the trace formula in the case where Σ is not simply connected. □

II. Combinatorial Lunes

6 Lunes and Traces

We denote the universal covering of Σ by

$$\pi : \widetilde{\Sigma} \to \Sigma$$

and, when Σ is not diffeomorphic to the 2-sphere, we assume $\widetilde{\Sigma} = \mathbb{C}$.

Definition 6.1 (Smooth Lunes). Assume (H). A **smooth** (α, β)**-lune** is an orientation preserving immersion $u : \mathbb{D} \to \Sigma$ such that

$$u(\mathbb{D} \cap \mathbb{R}) \subset \alpha, \qquad u(\mathbb{D} \cap S^1) \subset \beta,$$

Three examples of smooth lunes are depicted in Figure 9. Two lunes are said to be **equivalent** iff there is an orientation preserving diffeomorphism $\varphi : \mathbb{D} \to \mathbb{D}$ such that

$$\varphi(-1) = -1, \qquad \varphi(1) = 1, \qquad u' = u \circ \varphi.$$

The equivalence class of u is denoted by $[u]$. That u is an immersion means that u is smooth and du is injective in all of $\mathbb{D}$, even at the corners ± 1. The set $u(\mathbb{D} \cap \mathbb{R})$ is called the **bottom boundary** of the lune, and the set $u(\mathbb{D} \cap S^1)$ is called the **top boundary**. The points

$$x = u(-1), \qquad y = u(1)$$

are called respectively the left and right **endpoints** of the lune. The locally constant function

$$\Sigma \setminus u(\partial\mathbb{D}) \to \mathbb{N} : z \mapsto \#u^{-1}(z)$$

is called the **counting function** of the lune. (This function is locally constant because a proper local homeomorphism is a covering projection.) A smooth lune is said to be **embedded** iff the map u is injective. These notions depend only on the equivalence class $[u]$ of the smooth lune u.

Our objective is to characterize smooth lunes in terms of their boundary behavior, i.e. to say when a pair of immersions $u_\alpha : (\mathbb{D} \cap \mathbb{R}, -1, 1) \to (\alpha, x, y)$ and $u_\beta : (\mathbb{D} \cap S^1, -1, 1) \to (\beta, x, y)$ extends to a smooth (α, β)-lune u. Recall the following definitions and theorems from Part I.

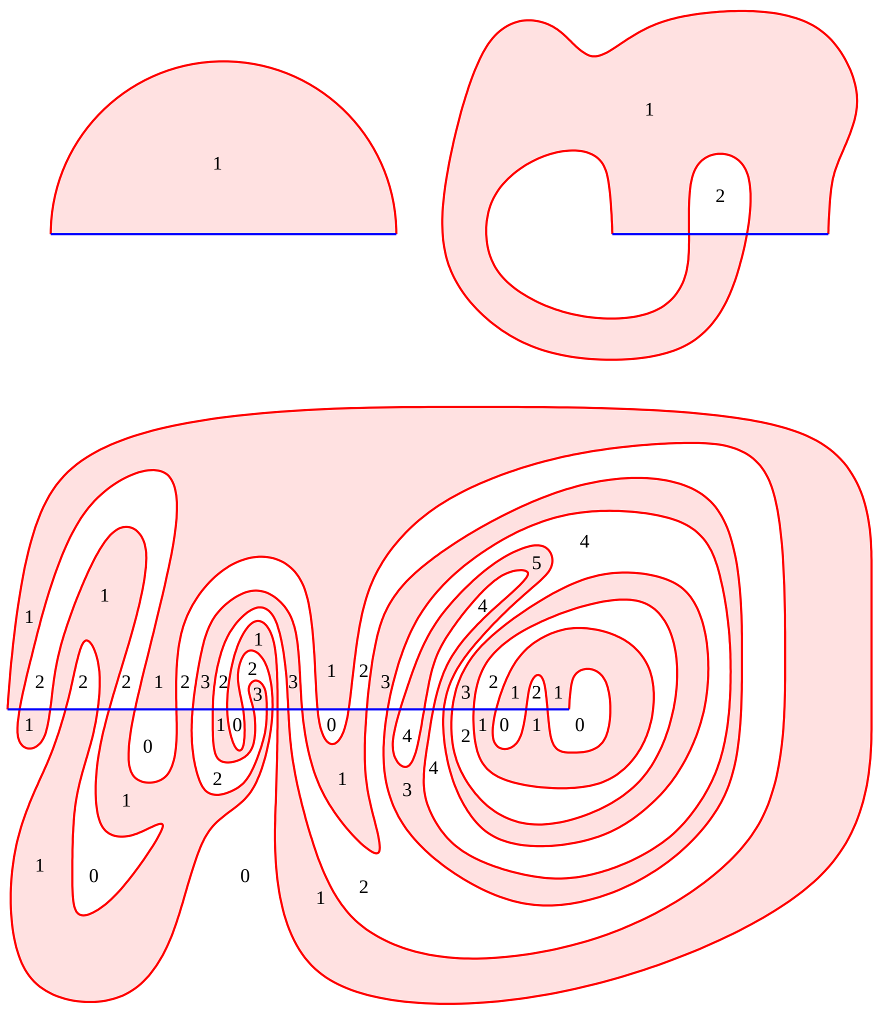

Figure 9: Three lunes.

Definition 6.2 (Traces). Assume (H). An (α, β)**-trace** is a triple

$$\Lambda = (x, y, \mathrm{w})$$

such that $x, y \in \alpha \cap \beta$ and $\mathrm{w} : \Sigma \setminus (\alpha \cup \beta) \to \mathbb{Z}$ is a locally constant function such that there exists a smooth map $u : \mathbb{D} \to \Sigma$ satisfying

$$u(\mathbb{D} \cap \mathbb{R}) \subset \alpha, \qquad u(\mathbb{D} \cap S^1) \subset \beta, \tag{33}$$

$$u(-1) = x, \qquad u(1) = y, \tag{34}$$

$$\mathrm{w}(z) = \deg(u, z), \qquad z \in \Sigma \setminus (\alpha \cup \beta). \tag{35}$$

The (α, β)-trace associated to a smooth map $u : \mathbb{D} \to \Sigma$ satisfying (33) is denoted by Λ_u.

The **boundary of an** (α, β)**-trace** $\Lambda = (x, y, \mathrm{w})$ is the triple

$$\partial\Lambda := (x, y, \partial\mathrm{w}).$$

Here

$$\partial\mathrm{w} : (\alpha \setminus \beta) \cup (\beta \setminus \alpha) \to \mathbb{Z}$$

is the locally constant function that assigns to $z \in \alpha \setminus \beta$ the value of w slightly to the left of α minus the value of w slightly to the right of α near z, and to $z \in \beta \setminus \alpha$ the value of w slightly to the right of β minus the value of w slightly to the left of β near z.

In Lemma 2.3 above it was shown that, if $\Lambda = (x, y, \mathrm{w})$ is the (α, β)-trace of a smooth map $u : \mathbb{D} \to \Sigma$ that satisfies (33), then $\partial\Lambda_u = (x, y, \nu)$, where the function $\nu := \partial\mathrm{w} : (\alpha \setminus \beta) \cup (\beta \setminus \alpha) \to \mathbb{Z}$ is given by

$$\nu(z) = \begin{cases} \deg(u|_{\partial\mathbb{D}\cap\mathbb{R}} : \partial\mathbb{D} \cap \mathbb{R} \to \alpha, z), & \text{for } z \in \alpha \setminus \beta, \\ -\deg(u|_{\partial\mathbb{D}\cap S^1} : \partial\mathbb{D} \cap S^1 \to \beta, z), & \text{for } z \in \beta \setminus \alpha. \end{cases} \tag{36}$$

Here we orient the one-manifolds $\mathbb{D} \cap \mathbb{R}$ and $\mathbb{D} \cap S^1$ from -1 to $+1$. Moreover, in Theorem 2.4 above it was shown that the homotopy class of a smooth map $u : \mathbb{D} \to \Sigma$ satisfying the boundary condition (33) is uniquely determined by its trace $\Lambda_u = (x, y, \mathrm{w})$. If Σ is not diffeomorphic to the 2-sphere then its universal cover is diffeomorphic to the 2-plane. In this situation it was also shown in Theorem 2.4 that the homotopy class of u and the degree function w are uniquely determined by the triple $\partial\Lambda_u = (x, y, \nu)$.

Remark 6.3 (**The Viterbo–Maslov index).** Let $\Lambda = (x, y, \mathrm{w})$ be an (α, β)-trace and denote by $\mu(\Lambda)$ its Viterbo–Maslov index as defined in 3.1 above (see also [39]). For $z \in \alpha \cap \beta$ let $m_z(\Lambda)$ be the sum of the four values of the function w encountered when walking along a small circle surrounding z. In Theorem 3.4 it was shown that the Viterbo–Maslov index of Λ is given by the trace formula

$$\mu(\Lambda) = \frac{m_x(\Lambda) + m_y(\Lambda)}{2}. \tag{37}$$

Let $\Lambda' = (y, z, \mathrm{w}')$ be another (α, β)-trace. The **catenation of** Λ **and** Λ' is defined by

$$\Lambda \# \Lambda' := (x, z, \mathrm{w} + \mathrm{w}').$$

It is again an (α, β)-trace and has Viterbo–Maslov index

$$\mu(\Lambda \# \Lambda') = \mu(\Lambda) + \mu(\Lambda'). \tag{38}$$

For a proof see [39, 30].

Definition 6.4 (**Arc Condition).** Let $\Lambda = (x, y, \mathrm{w})$ be an (α, β)-trace and

$$\nu_\alpha := \partial \mathrm{w}|_{\alpha \setminus \beta}, \qquad \nu_\beta := -\partial \mathrm{w}|_{\beta \setminus \alpha}.$$

Λ is said to satisfy the **arc condition** if

$$x \neq y, \qquad \min |\nu_\alpha| = \min |\nu_\beta| = 0. \tag{39}$$

When Λ satisfies the arc condition there are arcs $A \subset \alpha$ and $B \subset \beta$ from x to y such that

$$\nu_\alpha(z) = \begin{cases} \pm 1, & \text{if } z \in A, \\ 0, & \text{if } z \in \alpha \setminus \overline{A}, \end{cases} \qquad \nu_\beta(z) = \begin{cases} \pm 1, & \text{if } z \in B, \\ 0, & \text{if } z \in \beta \setminus \overline{B}. \end{cases} \tag{40}$$

Here the plus sign is chosen iff the orientation of A from x to y agrees with that of α, respectively the orientation of B from x to y agrees with that of β. In this situation the quadruple (x, y, A, B) and the triple $(x, y, \partial \mathrm{w})$ determine one another and we also write

$$\partial \Lambda = (x, y, A, B)$$

for the boundary of Λ. When $u : \mathbb{D} \to \Sigma$ is a smooth map satisfying (33) and $\Lambda_u = (x, y, \mathrm{w})$ satisfies the arc condition and $\partial \Lambda_u = (x, y, A, B)$ then the path $s \mapsto u(-\cos(\pi s), 0)$ is homotopic in α to a path traversing A and the path $s \mapsto u(-\cos(\pi s), \sin(\pi s))$ is homotopic in β to a path traversing B.

Theorem 6.5. *Assume (H). If $u : \mathbb{D} \to \Sigma$ is a smooth (α, β)-lune then its (α, β)-trace Λ_u satisfies the arc condition.*

Proof. See Section 7 page 55. □

Definition 6.6 (Combinatorial Lunes). *Assume (H). A* **combinatorial** (α, β)**-lune** *is an (α, β)-trace $\Lambda = (x, y, \mathrm{w})$ with boundary $\partial\Lambda =: (x, y, A, B)$ that satisfies the arc condition and the following.*

(I) $\mathrm{w}(z) \geq 0$ *for every* $z \in \Sigma \setminus (\alpha \cup \beta)$.

(II) *The intersection index of A and B at x is $+1$ and at y is -1.*

(III) $\mathrm{w}(z) \in \{0, 1\}$ *for z sufficiently close to x or y.*

Condition (II) says that the angle from A to B at x is between zero and π and the angle from B to A at y is also between zero and π.

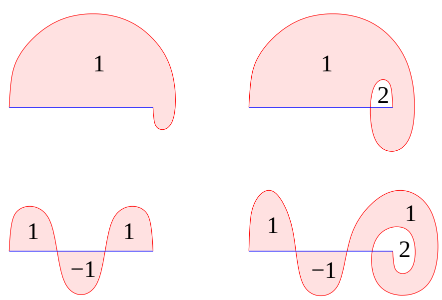

Figure 10: (α, β)-traces which satisfy the arc condition but are not lunes.

Theorem 6.7 (Existence). *Assume (H) and let $\Lambda = (x, y, \mathrm{w})$ be an (α, β)-trace. Consider the following three conditions.*

(i) *There exists a smooth (α, β)-lune u such that $\Lambda_u = \Lambda$.*

(ii) $\mathrm{w} \geq 0$ *and* $\mu(\Lambda) = 1$.

(iii) Λ *is a combinatorial (α, β)-lune.*

Then (i) $\Longrightarrow$ (ii) $\Longleftrightarrow$ (iii). If Σ is simply connected then all three conditions are equivalent.

Proof. See Section 8 page 63. □

Theorem 6.8 (**Uniqueness**). *Assume (H). If two smooth (α, β)-lunes have the same trace then they are equivalent.*

Proof. See Section 8 page 63. $\square$

Corollary 6.9. *Assume (H) and let*

$$\Lambda = (x, y, \mathrm{w})$$

be an (α, β)-trace. Choose a universal covering $\pi : \widetilde{\Sigma} \to \Sigma$, a point

$$\widetilde{x} \in \pi^{-1}(x),$$

and lifts $\widetilde{\alpha}$ and $\widetilde{\beta}$ of α and β such that

$$\widetilde{x} \in \widetilde{\alpha} \cap \widetilde{\beta}.$$

Let

$$\widetilde{\Lambda} = (\widetilde{x}, \widetilde{y}, \widetilde{\mathrm{w}})$$

be the lift of Λ to the universal cover.

(i) *If $\widetilde{\Lambda}$ is a combinatorial $(\widetilde{\alpha}, \widetilde{\beta})$-lune then Λ is a combinatorial (α, β)-lune.*

(ii) *$\widetilde{\Lambda}$ is a combinatorial $(\widetilde{\alpha}, \widetilde{\beta})$-lune if and only if there exists a smooth (α, β)-lune u such that $\Lambda_u = \Lambda$.*

Proof. Lifting defines a one-to-one correspondence between smooth (α, β)-lunes with trace Λ and smooth $(\widetilde{\alpha}, \widetilde{\beta})$-lunes with trace $\widetilde{\Lambda}$. Hence the assertions follow from Theorem 6.7. $\square$

Remark 6.10. Assume (H) and let Λ be an (α, β)-trace. We conjecture that the three conditions in Theorem 6.7 are equivalent, even when Σ is not simply connected, i.e.

If Λ is a combinatorial (α, β)-lune
then there exists a smooth (α, β)-lune u such that $\Lambda = \Lambda_u$.

Theorem 6.7 shows that this conjecture is equivalent to the following.

If Λ is a combinatorial (α, β)-lune
then $\widetilde{\Lambda}$ is a combinatorial $(\widetilde{\alpha}, \widetilde{\beta})$-lune.

The hard part is to prove that $\widetilde{\Lambda}$ satisfies (I), i.e. that the winding numbers are nonnegative.

Remark 6.11. Assume (H). Corollary 6.9 and Theorem 6.8 suggest the following algorithm for finding a smooth (α, β)-lune.

1. Fix two points $x, y \in \alpha \cap \beta$ with opposite intersection indices, and two oriented embedded arcs $A \subset \alpha$ and $B \subset \beta$ from x to y so that (II) holds.

2. If A is not homotopic to B with fixed endpoints discard this pair.[1] Otherwise (x, y, A, B) is the boundary of an (α, β)-trace $\Lambda = (x, y, \mathrm{w})$ satisfying the arc condition and (II) (for a suitable function w to be chosen below).

3a. If Σ is diffeomorphic to the 2-sphere let $\mathrm{w} : \Sigma \setminus (A \cup B) \to \mathbb{Z}$ be the winding number of the loop $A - B$ in $\Sigma \setminus \{z_0\}$, where $z_0 \in \alpha \setminus A$ is chosen close to x. Check if w satisfies (I) and (III). If yes, then $\Lambda = (x, y, \mathrm{w})$ is a combinatorial (α, β)-lune and hence, by Theorems 6.7 and 6.8, gives rise to a smooth (α, β)-lune u, unique up to isotopy.

3b. If Σ is not diffeomorphic to the 2-sphere choose lifts $\widetilde{A}$ of A and $\widetilde{B}$ of B to a universal covering $\pi : \mathbb{C} \to \Sigma$ connecting $\widetilde{x}$ and $\widetilde{y}$ and let

$$\widetilde{\mathrm{w}} : \mathbb{C} \setminus (\widetilde{A} \cup \widetilde{B}) \to \mathbb{Z}$$

be the winding number of $\widetilde{A} - \widetilde{B}$. Check if $\widetilde{\mathrm{w}}$ satisfies (I) and (III). If yes, then $\widetilde{\Lambda} := (\widetilde{x}, \widetilde{y}, \widetilde{\mathrm{w}})$ is a combinatorial $(\widetilde{\alpha}, \widetilde{\beta})$-lune and hence, by Theorem 6.7, gives rise to a smooth (α, β)-lune u such that

$$\Lambda_u = \Lambda := (x, y, \mathrm{w}), \qquad \mathrm{w}(z) := \sum_{\widetilde{z} \in \pi^{-1}(z)} \widetilde{\mathrm{w}}(\widetilde{z}).$$

By Theorem 6.8, the (α, β)-lune u is uniquely determined by Λ up to isotopy.

Proposition 6.12. *Assume (H) and let $\Lambda = (x, y, \mathrm{w})$ be an (α, β)-trace that satisfies the arc condition and let $\partial\Lambda =: (x, y, A, B)$. Let S be a connected component of $\Sigma \setminus (A \cup B)$ such that $\mathrm{w}|_S \not\equiv 0$. Then S is diffeomorphic to the open unit disc in $\mathbb{C}$.*

Proof. By Definition 6.2, there is a smooth map $u : \mathbb{D} \to \Sigma$ satisfying (33) such that $\Lambda_u = \Lambda$. By a homotopy argument we may assume, without loss of generality, that $u(\mathbb{D} \cap \mathbb{R}) = A$ and $u(\mathbb{D} \cap S^1) = B$. Let S be a connected component of $\Sigma \setminus (A \cup B)$ such that w does not vanish on S. We prove in two steps that S is diffeomorphic to the open unit disc in $\mathbb{C}$.

[1] This problem is solvable via Dehn's algorithm. See the wikipedia article Small Cancellation Theory and the references cited therein.

Step 1. *If S is not diffeomorphic to the open unit disc in $\mathbb{C}$ then there is an embedded loop $\gamma : \mathbb{R}/\mathbb{Z} \to S$ and a loop $\gamma' : \mathbb{R}/\mathbb{Z} \to \Sigma$ with intersection number $\gamma \cdot \gamma' = 1$.*

If S has positive genus there are in fact two embedded loops in S with intersection number one. If S has genus zero but is not diffeomorphic to the disc it is diffeomorphic to a multiply connected subset of $\mathbb{C}$, i.e. a disc with at least one hole cut out. Let $\gamma : \mathbb{R}/\mathbb{Z} \to S$ be an embedded loop encircling one of the holes and choose an arc in $\overline{S}$ which connects two boundary points and has intersection number one with γ. (For an elegant construction of such a loop in the case of an open subset of $\mathbb{C}$ see Ahlfors [3].) Since $\Sigma \setminus S$ is connected the arc can be completed to a loop in Σ which still has intersection number one with γ. This proves Step 1.

Step 2. *S is diffeomorphic to the open unit disc in $\mathbb{C}$.*

Assume, by contradiction, that this is false and choose γ and γ' as in Step 1. By transversality theory we may assume that u is transverse to γ. Since $C := \gamma(\mathbb{R}/\mathbb{Z})$ is disjoint from $u(\partial\mathbb{D}) = A \cup B$ it follows that $\Gamma := u^{-1}(C)$ is a disjoint union of embedded circles in $\Delta := u^{-1}(S) \subset \mathbb{D}$. Orient Γ such that the degree of $u|_\Gamma : \Gamma \to C$ agrees with the degree of $u|_\Delta : \Delta \to S$. More precisely, let $z \in \Gamma$ and $t \in \mathbb{R}/\mathbb{Z}$ such that $u(z) = \gamma(t)$. Call a nonzero tangent vector $\hat{z} \in T_z\Gamma$ positive if the vectors $\dot{\gamma}(t), du(z)\mathbf{i}\hat{z}$ form a positively oriented basis of $T_{u(z)}\Sigma$. Then, if $z \in \Gamma$ is a regular point of both $u|_\Delta : \Delta \to S$ and $u|_\Gamma : \Gamma \to C$, the linear map $du(z) : \mathbb{C} \to T_{u(z)}\Sigma$ has the same sign as its restriction $du(z) : T_z\Gamma \to T_{u(z)}C$. Thus $u|_\Gamma : \Gamma \to C$ has nonzero degree. Choose a connected component Γ_0 of Γ such that $u|_{\Gamma_0} : \Gamma_0 \to C$ has degree $d \neq 0$. Since Γ_0 is a loop in $\mathbb{D}$ it follows that the d-fold iterate of γ is contractible. Hence γ is contractible by A.3 in Appendix A. This proves Step 2 and Proposition 6.12. $\square$

7 Arcs

In this section we prove Theorem 6.5. The first step is to prove the arc condition under the hypothesis that α and β are not contractible (Proposition 7.1). The second step is to characterize embedded lunes in terms of their traces (Proposition 7.4). The third step is to prove the arc condition for lunes in the two-sphere (Proposition 7.7).

Proposition 7.1. *Assume (H), suppose Σ is not simply connected, and choose a universal covering $\pi : \mathbb{C} \to \Sigma$. Let $\Lambda = (x, y, \mathrm{w})$ be an (α, β)-trace and denote*

$$\nu_\alpha := \partial \mathrm{w}|_{\alpha \setminus \beta}, \qquad \nu_\beta := -\partial \mathrm{w}|_{\beta \setminus \alpha}.$$

Choose lifts $\widetilde{\alpha}$, $\widetilde{\beta}$, and $\widetilde{\Lambda} = (\widetilde{x}, \widetilde{y}, \widetilde{\mathrm{w}})$ of α, β, and Λ such that $\widetilde{\Lambda}$ is an $(\widetilde{\alpha}, \widetilde{\beta})$-trace. Thus $\widetilde{x}, \widetilde{y} \in \widetilde{\alpha} \cap \widetilde{\beta}$ and the path from $\widetilde{x}$ to $\widetilde{y}$ in $\widetilde{\alpha}$ (respecively $\widetilde{\beta}$) determined by $\partial\widetilde{\mathrm{w}}$ is the lift of the path from x to y in α (respectively β) determined by $\partial \mathrm{w}$. Assume

$$\widetilde{\mathrm{w}} \geq 0, \qquad \widetilde{\mathrm{w}} \not\equiv 0.$$

Then the following holds

(i) *If α is a noncontractible embedded circle then there exists an oriented arc $A \subset \alpha$ from x to y (equal to $\{x\}$ in the case $x = y$) such that*

$$\nu_\alpha(z) = \begin{cases} \pm 1, & \text{for } z \in A \setminus \beta, \\ 0, & \text{for } z \in \alpha \setminus (A \cup \beta). \end{cases} \tag{41}$$

Here the plus sign is chosen if and only if the orientations of A and α agree. If β is a noncontractible embedded circle the same holds for ν_β.

(ii) *If α and β are both noncontractible embedded circles then Λ satisfies the arc condition.*

Proof. We prove (i). The universal covering $\pi : \mathbb{C} \to \Sigma$ and the lifts $\widetilde{\alpha}$, $\widetilde{\beta}$, and $\widetilde{\Lambda} = (\widetilde{x}, \widetilde{y}, \widetilde{\mathrm{w}})$ can be chosen such that

$$\widetilde{\alpha} = \mathbb{R}, \qquad \widetilde{x} = 0, \qquad \widetilde{y} = a \geq 0, \qquad \pi(\widetilde{z} + 1) = \pi(\widetilde{z}),$$

and π maps the interval $[0, 1)$ bijectively onto α. Denote by

$$\widetilde{B} \subset \widetilde{\beta}$$

the closure of the support of

$$\nu_{\widetilde{\beta}} := -\partial\widetilde{\mathrm{w}}|_{\widetilde{\beta} \setminus \widetilde{\alpha}}.$$

If β is noncontractible then $\widetilde{B}$ is the unique arc in $\widetilde{\beta}$ from 0 to a. If β is contractible then $\widetilde{\beta} \subset \mathbb{C}$ is an embedded circle and $\widetilde{B}$ is either an arc in $\widetilde{\beta}$

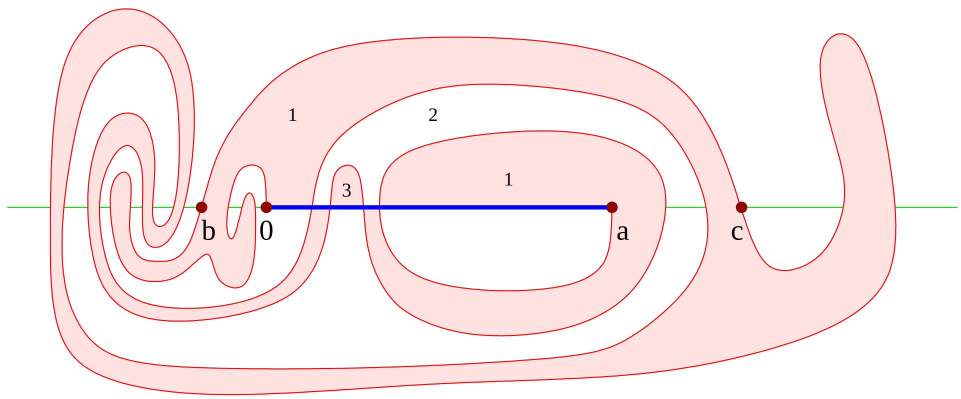

Figure 11: The lift of an (α, β)-trace with $\widetilde{\mathrm{w}} \geq 0$.

from 0 to a or is equal to $\widetilde{\beta}$. We must prove that $A := \pi([0, a])$ is an arc or, equivalently, that $a < 1$.

Let Γ be the set of connected components γ of $\widetilde{B} \cap (\mathbb{R} \times [0, \infty))$ such that the function $\widetilde{\mathrm{w}}$ is zero on one side of γ and positive on the other. If $\gamma \in \Gamma$, neither end point of γ can lie in the open interval $(0, a)$ since the function $\widetilde{\mathrm{w}}$ is at least one above this interval. We claim that there exists a connected component $\gamma \in \Gamma$ whose endpoints b and c satisfy

$$b \leq 0 \leq a \leq c, \qquad \partial\gamma = \{b, c\}. \tag{42}$$

(See Figure 11.) To see this walk slightly above the real axis towards zero, starting at $-\infty$. Just before the first crossing b_1 with $\widetilde{B}$ turn left and follow the arc in $\widetilde{B}$ until it intersects the real axis again at c_1. The two intersections b_1 and c_1 are the endpoints of an element γ_1 of Γ. Obviously $b_1 \leq 0$ and, as noted above, c_1 cannot lie in the interval $(0, a)$. For the same reason c_1 cannot be equal to zero. Hence either $c_1 < 0$ or $c_1 \geq a$. In the latter case γ_1 is the required arc γ. In the former case we continue walking towards zero along the real axis until the next intersection with $\widetilde{B}$ and repeat the above procedure. Because the set of intersection points of $\widetilde{B}$ with $\widetilde{\alpha} = \mathbb{R}$ is finite the process must terminate after finitely many steps. Thus we have proved the existence of an arc $\gamma \in \Gamma$ satisfying (42).

Assume that

$$c \geq b + 1.$$

If $c = b + 1$ then $c \in \widetilde{\beta} \cap (\widetilde{\beta} + 1)$ and hence $\widetilde{\beta} = \widetilde{\beta} + 1$. It follows that the intersection numbers of $\mathbb{R}$ and $\widetilde{\beta}$ at b and c agree. But this contradicts the fact that b and c are the endpoints of an arc in $\widetilde{\beta}$ contained in the closed upper halfplane. Thus we have $c > b+1$. When this holds the arc γ and its translate $\gamma+1$ must intersect and their intersection does not contain the endpoints b and c. We denote by $\zeta \in \gamma \setminus \{b,c\}$ the first point in $\gamma+1$ we encounter when walking along γ from b to c. Let

$$U_0 \subset \widetilde{\beta}, \qquad U_1 \subset \widetilde{\beta}+1$$

be sufficiently small connected open neighborhoods of ζ, so that $\pi : U_0 \to \beta$ and $\pi : U_1 \to \beta$ are embeddings and their images agree. Thus

$$\pi(U_0) = \pi(U_1) \subset \beta$$

is an open neighborhood of $z := \pi(\zeta)$ in β. Hence it follows from a lifting argument that $U_0 = U_1 \subset \gamma + 1$ and this contradicts our choice of ζ. This contradiction shows that our hypothesis $c \geq b+1$ must have been wrong. Thus we have proved that

$$b \leq 0 \leq a \leq c < b+1 \leq 1.$$

Hence $0 \leq a < 1$ and so $A = \pi([0,a])$ is an arc, as claimed. In the case $a = 0$ we obtain the trivial arc from $x = y$ to itself. This proves (i).

We prove (ii). Assume that α and β are noncontractible embedded circles. Then it follows from (i) that there exist oriented arcs $A \subset \alpha$ and $B \subset \beta$ from x to y such that ν_α and ν_β are given by (40). If $x = y$ it follows also from (i) that $A = B = \{x\}$, hence

$$\nu_{\widetilde{\alpha}} \equiv 0, \qquad \nu_{\widetilde{\beta}} \equiv 0,$$

and hence $\widetilde{\mathrm{w}} \equiv 0$, in contradiction to our hypothesis. Thus $x \neq y$ and so Λ satisfies the arc condition. This proves (ii) and Proposition 7.1. $\square$

Example 7.2. Let $\alpha \subset \Sigma$ be a noncontractible embedded circle and $\beta \subset \Sigma$ be a contractible embedded circle intersecting α transversally. Suppose β is oriented as the boundary of an embedded disc $\Delta \subset \Sigma$. Let

$$x = y \in \alpha \cap \beta, \qquad \nu_\alpha \equiv 0, \qquad \nu_\beta \equiv 1,$$

and define

$$\mathrm{w}(z) := \begin{cases} 1, & \text{for } z \in \Delta \setminus (\alpha \cup \beta), \\ 0, & \text{for } z \in \Sigma \setminus (\Delta \cup \alpha \cup \beta). \end{cases}$$

Then $\Lambda = (x, y, \nu_\alpha, \nu_\beta, \mathrm{w})$ is an (α,β)-trace that satisfies the hypotheses of Proposition 7.1 (i) with $x = y$ and $A = \{x\}$.

Definition 7.3. An (α, β)-trace $\Lambda = (x, y, \mathrm{w})$ is called **primitive** if it satisfies the arc condition with boundary $\partial\Lambda =: (x, y, A, B)$ and

$$A \cap \beta = \alpha \cap B = \{x, y\}.$$

A smooth (α, β)-lune u is called **primitive** if its (α, β)-trace Λ_u is primitive. It is called **embedded** if $u : \mathbb{D} \to \Sigma$ is injective.

The next proposition is the special case of Theorems 6.7 and 6.8 for embedded lunes. It shows that isotopy classes of primitive smooth (α, β)-lunes are in one-to-one correspondence with the simply connected components of $\Sigma \setminus (\alpha \cup \beta)$ with two corners. We will also call such a component a **primitive** (α, β)**-lune**.

Proposition 7.4 (**Embedded lunes**). *Assume (H) and let $\Lambda = (x, y, \mathrm{w})$ be an (α, β)-trace. The following are equivalent.*

(i) *Λ is a combinatorial lune and its boundary $\partial\Lambda = (x, y, A, B)$ satisfies*

$$A \cap B = \{x, y\}.$$

(ii) *There exists an embedded (α, β)-lune u such that $\Lambda_u = \Lambda$.*

If Λ satisfies (i) then any two smooth (α, β)-lunes u and v with $\Lambda_u = \Lambda_v = \Lambda$ are equivalent.

Proof. We prove that (ii) implies (i). Let $u : \mathbb{D} \to \Sigma$ be an embedded (α, β)-lune with $\Lambda_u = \Lambda$. Then $u|_{\mathbb{D}\cap\mathbb{R}} : \mathbb{D} \cap \mathbb{R} \to \alpha$ and $u|_{\mathbb{D}\cap S^1} : \mathbb{D} \cap S^1 \to \beta$ are embeddings. Hence Λ satisfies the arc condition and $\partial\Lambda = (x, y, A, B)$ with $A = u(\mathbb{D} \cap \mathbb{R})$ and $B = u(\mathbb{D} \cap S^1)$. Since w is the counting function of u it takes only the values zero and one. If $z \in A \cap B$ then $u^{-1}(z)$ contains a single point which must lie in $\mathbb{D} \cap \mathbb{R}$ and $\mathbb{D} \cap S^1$, hence is either -1 or $+1$, and so $z = x$ or $z = y$. The assertion about the intersection indices follows from the fact that u is an immersion. Thus we have proved that (ii) implies (i).

We prove that (i) implies (ii). This relies on the following.

Claim. *Let $\Lambda = (x, y, \mathrm{w})$ be an (α, β)-trace that satisfies the arc condition and $\partial\Lambda =: (x, y, A, B)$ with $A \cap B = \{x, y\}$. Then $\Sigma \setminus (A \cup B)$ has two components and one of these is homeomorphic to the disc.*

To prove the claim, let $\Gamma \subset \Sigma$ be an embedded circle obtained from $A \cup B$ by smoothing the corners. Then Γ is contractible and hence, by a theorem of Epstein [8], bounds a disc. This proves the claim.

Now suppose that $\Lambda = (x, y, \mathrm{w})$ is an (α, β)-trace that satisfies (i) and let $\partial\Lambda =: (x, y, A, B)$. By the claim, the complement $\Sigma \setminus (A \cup B)$ has two components, one of which is homeomorphic to the disc. Denote the components by Σ_0 and Σ_1. Since Λ is a combinatorial lune, it follows that w only takes the values zero and one. Hence we may choose the indexing such that

$$\mathrm{w}(z) = \begin{cases} 0, & \text{for } z \in \Sigma_0 \setminus (\alpha \cup \beta), \\ 1, & \text{for } z \in \Sigma_1 \setminus (\alpha \cup \beta). \end{cases}$$

We prove that Σ_1 is homeomorphic to the disc. Suppose, by contradiction, that Σ_1 is not homeomorphic to the disc. Then Σ is not diffeomorphic to the 2-sphere and, by the claim, Σ_0 is homeomorphic to the disc. By Definition 6.4, there is a smooth map $u : \mathbb{D} \to \Sigma$ that satisfies the boundary condition (33) such that $\Lambda_u = \Lambda$. Since Σ is not diffeomorphic to the 2-sphere, the homotopy class of u is uniquely determined by the quadruple (x, y, A, B) (see Theorem 2.4 above). Since Σ_0 is homeomorphic to the disc we may choose u such that $u(\mathbb{D}) = \overline{\Sigma}_0$ and hence $\mathrm{w}(z) = \deg(u, z) = 0$ for $z \in \Sigma_1 \setminus (\alpha \cup \beta)$, in contradiction to our choice of indexing. This shows that Σ_1 must be homeomorphic to the disc. Let N denote the closure of Σ_1:

$$N := \overline{\Sigma}_1 = \Sigma_1 \cup A \cup B.$$

Then the orientation of $\partial N = A \cup B$ agrees with the orientation of A and is opposite to the orientation of B, i.e. N lies to the left of A and to the right of B. Since the intersection index of A and B at x is $+1$ and at y is -1, it follows that the angles of N at x and y are between zero and π and hence N is a 2-manifold with two corners. Since N is simply connected there exists a diffeomorphism $u : \mathbb{D} \to N$ such that

$$u(-1) = x, \qquad u(1) = y, \qquad u(\mathbb{D} \cap \mathbb{R}) = A, \qquad u(\mathbb{D} \cap S^1) = B.$$

This diffeomorphism is the required embedded (α, β)-lune.

We prove that the embedded (α, β)-lune in (ii) is unique up to equivalence. Let $v : \mathbb{D} \to \Sigma$ be another smooth (α, β)-lune such that $\Lambda_v = \Lambda$. Then v maps the boundary of $\mathbb{D}$ bijectively onto $A \cup B$, because $A \cap B = \{x, y\}$. Moreover, w is the counting function of v and $\#v^{-1}(z)$ is constant on each component of $\Sigma \setminus (A \cup B)$. Hence $\#v^{-1}(z) = 0$ for $z \in \Sigma_0$ and $\#v^{-1}(z) = 1$ for $z \in \Sigma_1$. This shows that v is injective and $v(\mathbb{D}) = N = u(\mathbb{D})$. Since u and v are embeddings the composition $\varphi := u^{-1} \circ v : \mathbb{D} \to \mathbb{D}$ is an orientation preserving diffeomorphism such that $\varphi(\pm 1) = \pm 1$. Hence $v = u \circ \varphi$ is equivalent to u. This proves Proposition 7.4. $\square$

Lemma 7.5. *Assume (H) and let $u : \mathbb{D} \to \Sigma$ be a smooth (α, β)-lune.*

(i) *Let S be a connected component of $\Sigma \setminus (\alpha \cup \beta)$. If $S \cap u(\mathbb{D}) \neq \emptyset$ then $S \subset u(\mathbb{D})$ and S is diffeomorphic to the open unit disc in $\mathbb{C}$.*

(ii) *Let Δ be a connected component of $\mathbb{D} \setminus u^{-1}(\alpha \cup \beta)$. Then Δ is diffeomorphic to the open unit disc and the restriction of u to Δ is a diffeomorphism onto the open set $S := u(\Delta) \subset \Sigma$.*

Proof. That $S \cap u(\mathbb{D}) \neq \emptyset$ implies $S \subset u(\mathbb{D})$ follows from the fact that u is an immersion. That this implies that S is diffeomorphic to the open unit disc in $\mathbb{C}$ follows as in Proposition 6.12. This proves (i). By (i) the open set $S := u(\Delta)$ in (ii) is diffeomorphic to the disc and hence is simply connected. Since $u : \Delta \to S$ is a proper covering it follows that $u : \Delta \to S$ is a diffeomorphism. This proves Lemma 7.5. □

Let $u : \mathbb{D} \to \Sigma$ be a smooth (α, β)-lune. The image under u of the connected component of $\mathbb{D} \setminus u^{-1}(\alpha \cup \beta)$ whose closure contains -1 is called the **left end of** u. The image under u of the connected component of $\mathbb{D} \setminus u^{-1}(\alpha \cup \beta)$ whose closure contains $+1$ is called the **right end of** u.

Lemma 7.6. *Assume (H) and let u be a smooth (α, β)-lune. If there is a primitive (α, β)-lune with the same left or right end as u it is equivalent to u.*

Proof. If u is not a primitive lune its ends have at least three corners. To see this, walk along $\mathbb{D} \cap \mathbb{R}$ (respectively $\mathbb{D} \cap S^1$) from -1 to 1 and let z_0 (respectively z_1) be the first intersection point with $u^{-1}(\beta)$ (respectively $u^{-1}(\alpha)$). Then $u(-1)$, $u(z_0)$, $u(z_1)$ are corners of the left end of u. Hence the hypotheses of Lemma 7.6 imply that u is a primitive lune. Two primitive lunes with the same ends are equivalent by Proposition 7.4. This proves Lemma 7.6. □

Proposition 7.7. *Assume (H) and suppose that Σ is diffeomorphic to the 2-sphere. If u is a smooth (α, β)-lune then Λ_u satisfies the arc condition.*

Proof. The proof is by induction on the number of intersection points of α and β. It has three steps.

Step 1. *Let u be a smooth (α, β)-lune whose (α, β)-trace*

$$\Lambda = \Lambda_u = (x, y, \mathrm{w})$$

does not satisfy the arc condition. Suppose there is a primitive (α, β)-lune with endpoints in $\Sigma \setminus \{x, y\}$. Then there is an embedded loop β', isotopic to β and transverse to α, and a smooth (α, β')-lune u' with endpoints x, y such that $\Lambda_{u'}$ does not satisfy the arc condition and $\#(\alpha \cap \beta') < \#(\alpha \cap \beta)$.

By Proposition 7.4, there exists a primitive smooth (α, β)-lune $u_0 : \mathbb{D} \to \Sigma$ whose endpoints $x_0 := u_0(-1)$ and $y_0 := u_0(+1)$ are contained in $\Sigma \setminus \{x, y\}$. Use this lune to remove the intersection points x_0 and y_0 by an isotopy of β, supported in a small neighborhood of the image of u_0. More precisely, extend u_0 to an embedding (still denoted by u_0) of the open set

$$\mathbb{D}_\varepsilon := \{z \in \mathbb{C} \,|\, \operatorname{Im} z > -\varepsilon, |z| < 1 + \varepsilon\}$$

for $\varepsilon > 0$ sufficiently small such that

$$u_0(\mathbb{D}_\varepsilon) \cap \beta = u_0(\mathbb{D}_\varepsilon \cap S^1), \qquad u_0(\mathbb{D}_\varepsilon) \cap \alpha = u_0(\mathbb{D}_\varepsilon \cap \mathbb{R}).$$

Choose a smooth cutoff function $\rho : \mathbb{D}_\varepsilon \to [0, 1]$ which vanishes near the boundary and is equal to one on $\mathbb{D}$. Consider the vector field ξ on Σ that vanishes outside $u_0(\mathbb{D}_\varepsilon)$ and satisfies

$$u_0^*\xi(z) = -\rho(z)\mathbf{i}.$$

Let $\psi_t : \Sigma \to \Sigma$ be the isotopy generated by ξ and, for $T > 0$ sufficiently large, define

$$\beta' := \psi_T(\beta), \qquad \Lambda' := (x, y, \nu_\alpha, \nu_{\beta'}, \mathrm{w}').$$

Here $\nu_{\beta'} : \beta' \setminus \alpha \to \mathbb{Z}$ is the unique one-chain equal to ν_β on $\beta \setminus u_0(\mathbb{D}_\varepsilon)$ and $\mathrm{w}' : \Sigma \setminus (\alpha \cup \beta') \to \mathbb{Z}$ is the unique two-chain equal to w on $\Sigma \setminus u_0(\mathbb{D}_\varepsilon)$. Since Λ does not satisfy the arc condition, neither does Λ'. Let $U \subset \mathbb{D}$ be the union of the components of $u^{-1}(u_0(\mathbb{D}_\varepsilon))$ that contain an arc in $\mathbb{D} \cap S^1$ and define the map $u' : \mathbb{D} \to \Sigma$ by

$$u'(z) := \begin{cases} \psi_T(u(z)), & \text{if } z \in U, \\ u(z), & \text{if } z \in \mathbb{D} \setminus U. \end{cases}$$

We prove that $U \cap \mathbb{R} = \emptyset$. To see this, note that the restriction of u to each connected component of U is a diffeomorphism onto its image which is either equal to $u_0(\{z \in \mathbb{D}_\varepsilon \,|\, |z| \geq 1\})$ or equal to $u_0(\{z \in \mathbb{D}_\varepsilon \,|\, |z| \leq 1\})$ (see Figure 16 below). Thus

$$u(U) \cap (\alpha \setminus \{x_0, y_0\}) \subset \operatorname{int}(u(U))$$

and hence $U \cap \mathbb{R} = \emptyset$ as claimed. This implies that u' is a smooth (α, β')-lune such that $\Lambda_{u'} = \Lambda'$. Hence $\Lambda_{u'}$ does not satisfy the arc condition. This proves Step 1.

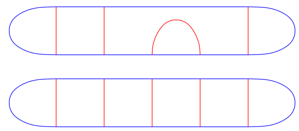

Figure 12: α encircles at least two primitive lunes.

Step 2. *Let u be a smooth (α, β)-lune with endpoints x, y and suppose that every primitive (α, β)-lune has x or y as one of its endpoints. Then Λ_u satisfies the arc condition.*

Both connected components of $\Sigma \backslash \alpha$ are discs, and each of these discs contains at least two primitive (α, β)-lunes. If it contains more than two there is one with endpoints in $\Sigma \setminus \{x, y\}$. Hence, under the assumptions of Step 2, each connected component of $\Sigma \setminus \alpha$ contains precisely two primitive (α, β)-lunes. (See Figure 12.) Thus each connected component of $\Sigma \setminus (\alpha \cup \beta)$ is either a quadrangle or a primitive (α, β)-lune and there are precisely four primitive (α, β)-lunes, two in each connected component of $\Sigma \backslash \alpha$. At least two primitive (α, β)-lunes contain x and at least two contain y. (See Figure 13.)

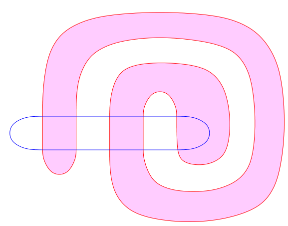

Figure 13: Four primitive lunes in the 2-sphere.

As Σ is diffeomorphic to S^2, the number of intersection points of α and β is even. Write $\alpha \cap \beta = \{x_0, \ldots, x_{2n-1}\}$, where the ordering is chosen along α, starting at $x_0 = x$. Then x_0 is contained in two primitive (α, β)-lunes, one with endpoints x_0, x_{2n-1} and one with endpoints x_0, x_1. Each connected component of $\Sigma \setminus \alpha$ determines an equivalence relation on $\alpha \cap \beta$: distinct points are equivalent iff they are connected by a β-arc in this component. Let A be the connected component containing the β-arc from x_0 to x_{2n-1} and B be the connected component containing the β-arc from x_0 to x_1. Then $x_{j-1} \sim_A x_{2n-j}$ and $x_{j+1} \sim_B x_{2n-j}$ for $j = 1, \ldots, n$. Thus the only other intersection point contained in two primitive (α, β)-lunes is $y = x_n$. Moreover, α and β have opposite intersection indices at x_i and x_{i+1} for each i, because the arcs in α from x_{i-1} to x_i and from x_i to x_{i+1} are contained in different connected components of $\Sigma \setminus \beta$. Since α and β have opposite intersection indices at x and y it follows that n is odd. Now the image of a neighborhood of $\mathbb{R} \cap \mathbb{D}$ under u is contained in either $\overline{A}$ or $\overline{B}$. Hence, when $n = 2k+1 \geq 3$, Figure 14 shows that one of the ends of u is a quadrangle and the other end is a primitive (α, β)-lune, in contradiction to Lemma 7.6. Hence the number of intersection points is $2n = 2$, each component of $\Sigma \setminus (\alpha \cup \beta)$ is a primitive (α, β)-lune, and all four primitive (α, β)-lunes contain x and y. By Lemma 7.6, one of them is equivalent to u. This proves Step 2.

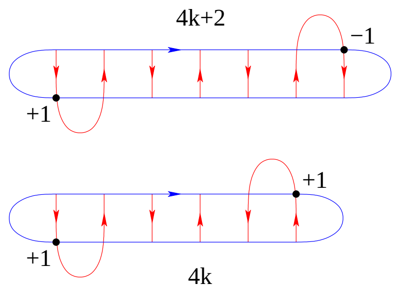

Figure 14: α intersects β in $4k$ or $4k + 2$ points.

Step 3. *We prove the proposition.*

Assume, by contradiction, that there is a smooth (α, β)-lune u such that Λ_u does not satisfy the arc condition. By Step 1 we can reduce the number of intersection points of α and β until there are no primitive (α, β)-lunes with endpoints in $\Sigma \setminus \{x, y\}$. Once this algorithm terminates the resulting lune still does not satisfy the arc condition, in contradiction to Step 2. This proves Step 3 and Proposition 7.7. $\square$

Proof of Theorem 6.5. Let $u : \mathbb{D} \to \Sigma$ be a smooth (α, β)-lune with (α, β)-trace $\Lambda_u =: (x, y, \mathrm{w})$ and denote $A := u(\mathbb{D} \cap \mathbb{R})$ and $B := u(\mathbb{D} \cap S^1)$. Since u is an immersion, α and β have opposite intersection indices at x and y, and hence $x \neq y$. We must prove that A and B are arcs. It is obvious that A is an arc whenever α is not compact, and B is an arc whenever β is not compact. It remains to show that A and B are arcs in the remaining cases. We prove this in four steps.

Step 1. *If α is not a contractible embedded circle then A is an arc.*

This follows immediately from Proposition 7.1.

Step 2. *If α and β are contractible embedded circles then A and B are arcs.*

If Σ is diffeomorphic to S^2 this follows from Proposition 7.7. Hence assume that Σ is not diffeomorphic to S^2. Then the universal cover of Σ is diffeomorphic to the complex plane. Choose a universal covering $\pi : \mathbb{C} \to \Sigma$ and a point $\widetilde{x} \in \pi^{-1}(x)$. Choose lifts $\widetilde{\alpha}, \widetilde{\beta} \subset \mathbb{C}$ of α, β such that $\widetilde{x} \in \widetilde{\alpha} \cap \widetilde{\beta}$. Then $\widetilde{\alpha}$ and $\widetilde{\beta}$ are embedded loops in $\mathbb{C}$ and u lifts to a smooth $(\widetilde{\alpha}, \widetilde{\beta})$-lune $\widetilde{u} : \mathbb{D} \to \mathbb{C}$ such that $\widetilde{u}(-1) = \widetilde{x}$. Compactify $\mathbb{C}$ to get the 2-sphere. Then, by Proposition 7.7, the subsets $\widetilde{A} := \widetilde{u}(\mathbb{D} \cap \mathbb{R}) \subset \widetilde{\alpha}$ and $\widetilde{B} := \widetilde{u}(\mathbb{D} \cap S^1) \subset \widetilde{\beta}$ are arcs. Since the restriction of π to $\widetilde{\alpha}$ is a diffeomorphism from $\widetilde{\alpha}$ to α it follows that $A \subset \alpha$ is an arc. Similarly for B. This proves Step 2.

Step 3. *If α is not a contractible embedded circle and β is a contractible embedded circle then A and B are arcs.*

That A is an arc follows from Step 1. To prove that B is an arc choose a universal covering $\pi : \mathbb{C} \to \Sigma$ with $\pi(0) = x$ and lifts $\widetilde{\alpha}$, $\widetilde{\beta}$, $\widetilde{u}$ with $0 \in \widetilde{\alpha} \cap \widetilde{\beta}$ and $\widetilde{u}(-1) = 0$ as in the proof of Step 2. Then $\widetilde{\beta} \subset \mathbb{C}$ is an embedded loop and we may assume without loss of generality that $\widetilde{\alpha} = \mathbb{R}$ and $\widetilde{A} = [0, a]$ with $0 < a < 1$. (If α is a noncontractible embedded circle we choose the lift such that $\widetilde{z} \mapsto \widetilde{z} + 1$ is a covering transformation and π maps the interval $[0, 1)$ bijectively onto α; if α is not compact we choose the universal covering such that π maps the interval $[0, a]$ bijectively onto A and $\widetilde{\beta}$ is transverse to $\mathbb{R}$, and then replace $\widetilde{\alpha}$ by $\mathbb{R}$.) In the Riemann sphere $S^2 \cong \overline{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ the real axis $\widetilde{\alpha} = \mathbb{R}$ compactifies to a great circle. Hence it follows from Proposition 7.7 that $\widetilde{B}$ is an arc. Since $\pi : \widetilde{\beta} \to \beta$ is a diffeomorphism it follows that B is an arc. This proves Step 3.

Step 4. *If β is not a contractible embedded circle then A and B are arcs.*

That B is an arc follows from Step 1 by interchanging α and β and replacing u with the smooth (β, α)-lune

$$v(z) := u\left(\frac{\mathbf{i} - z}{1 - \mathbf{i}z}\right).$$

If α is not a contractible embedded circle then A is an arc by Step 1. If α is a contractible embedded circle then A is an arc by Step 3 with α and β interchanged. This proves Step 4. The assertion of Theorem 6.5 follows from Steps 2, 3, and 4. $\square$

8 Combinatorial Lunes

In this section we prove Theorems 6.7 and 6.8. Proposition 7.4 establishes the equivalence of (i) and (iii) in Theorem 6.7 under the additional hypothesis that $\Lambda = (x, y, A, B, \mathrm{w})$ satisfies the arc condition and $A \cap B = \{x, y\}$. In this case the hypothesis that Σ is simply connected can be dropped. The induction argument for the proof of Theorems 6.7 and 6.8 is the content of the next three lemmas.

Lemma 8.1. *Assume (H) and suppose that Σ is simply connected. Let $\Lambda = (x, y, \mathrm{w})$ be a combinatorial (α, β)-lune with boundary $\partial\Lambda = (x, y, A, B)$ such that*

$$A \cap B \neq \{x, y\}.$$

Then there exists a combinatorial (α, β)-lune $\Lambda_0 = (x_0, y_0, \mathrm{w}_0)$ with boundary $\partial\Lambda_0 = (x_0, y_0, A_0, B_0)$ such that $\mathrm{w} \geq \mathrm{w}_0$ and

$$A_0 \subset A \setminus \{x, y\}, \quad B_0 \subset B \setminus \{x, y\}, \quad A_0 \cap B = A \cap B_0 = \{x_0, y_0\}. \tag{43}$$

Proof. Let $\prec$ denote the order relation on A determined by the orientation from x to y. Denote the intersection points of A and B by

$$x = x_0 \prec x_1 \prec \cdots \prec x_{n-1} \prec x_n = y.$$

Define a function $\sigma : \{0, \ldots, n-1\} \to \{1, \ldots, n\}$ as follows. Walk along B towards y, starting at x_i and denote the next intersection point encountered by $x_{\sigma(i)}$. This function σ is bijective. Let $\varepsilon_i \in \{\pm 1\}$ be the intersection index of A and B at x_i. Thus

$$\varepsilon_0 = 1, \qquad \varepsilon_n = -1, \qquad \sum_{i=0}^{n} \varepsilon_i = 0.$$

Consider the set

$$I := \{i \in \mathbb{N} \,|\, 0 \leq i \leq n-1,\, \varepsilon_i = 1,\, \varepsilon_{\sigma(i)} = -1\}.$$

We prove that this set has the following properties.

(a) $I \neq \emptyset$.

(b) If $i \in I$, $i < j < \sigma(i)$, and $\varepsilon_j = 1$, then $j \in I$ and $i < \sigma(j) < \sigma(i)$.

(c) If $i \in I$, $\sigma(i) < j < i$, and $\varepsilon_j = 1$, then $j \in I$ and $\sigma(i) < \sigma(j) < i$.

(d) $0 \in I$ if and only if $n \in \sigma(I)$ if and only if $n = 1 = \sigma(0)$.

To see this, denote by m_i the value of w in the right upper quadrant near x_i. Thus

$$m_j = m_0 + \sum_{i=1}^{j} \varepsilon_i$$

for $j = 1, \ldots, n$ and

$$m_{\sigma(i)} = m_i + \varepsilon_{\sigma(i)} \tag{44}$$

for $i = 0, \ldots, n-1$. (See Figure 15.)

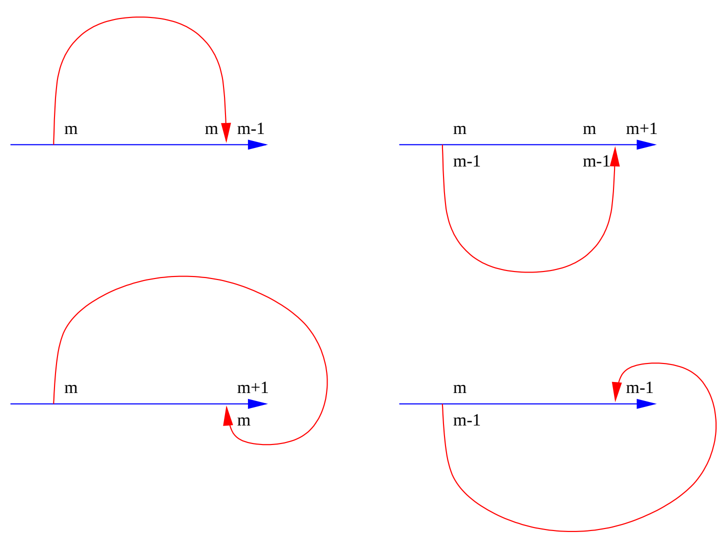

Figure 15: Simple arcs.

We prove that I satisfies (a). Consider the sequence

$$i_0 := 0, \qquad i_1 := \sigma(i_0), \qquad i_2 := \sigma(i_1), \ldots.$$

Thus the points x_i are encountered in the order

$$x = x_0 = x_{i_0}, x_{i_1}, \ldots, x_{i_{n-1}}, x_{i_n} = x_n = y$$

when walking along B from x to y. By (44), we have

$$\varepsilon_{i_0} = 1, \qquad \varepsilon_{i_n} = -1, \qquad m_{i_k} = m_{i_{k-1}} + \varepsilon_{i_k}.$$

Let $k \in \{0, \ldots, n-1\}$ be the largest integer such that $\varepsilon_{i_k} = 1$. Then we have $\varepsilon_{\sigma(i_k)} = \varepsilon_{i_{k+1}} = -1$ and hence $i_k \in I$. Thus I is nonempty.

We prove that I satisfies (b) and (c). Let $i \in I$ such that $\sigma(i) > i$. Then $\varepsilon_i = 1$ and $\varepsilon_{\sigma(i)} = -1$. Hence

$$m_{\sigma(i)} = m_i + \varepsilon_{\sigma(i)} = m_i - 1,$$

and hence, in the interval $i < j < \sigma(i)$, the numbers of intersection points with positive and with negative intersection indices agree. Consider the arcs $A_i \subset A$ and $B_i \subset B$ that connect x_i to $x_{\sigma(i)}$. Then $A \cap B_i = \{x_i, x_{\sigma(i)}\}$. Since Σ is simply connected the piecewise smooth embedded loop $A_i - B_i$ is contractible. This implies that the complement $\Sigma \setminus (A_i \cup B_i)$ has two connected components. Let Σ_i be the connected component of $\Sigma \setminus (A_i \cup B_i)$ that contains the points slightly to the left of A_i. Then any arc on B that starts at $x_j \in A_i$ with $\varepsilon_j = 1$ is trapped in Σ_i and hence must exit it through A_i. Hence

$$x_j \in A_i, \quad \varepsilon_j = 1 \qquad \Longrightarrow \qquad x_{\sigma(j)} \in A_i, \quad \varepsilon_{\sigma(j)} = -1.$$

Thus we have proved that I satisfies (b). That it satisfies (c) follows by a similar argument.

We prove that I satisfies (d). Here we use the fact that Λ satisfies (III) or, equivalently, $m_0 = 1$ and $m_n = 0$. If $0 \in I$ then $m_{\sigma(0)} = m_0 + \varepsilon_{\sigma(0)} = 0$. Since $m_i > 0$ for $i < n$ this implies $\sigma(0) = n = 1$. Conversely, suppose that $n \in \sigma(I)$ and let $i := \sigma^{-1}(n) \in I$. Then $m_i = m_n - \varepsilon_{\sigma(i)} = 1$. Since $m_i > 1$ for $i \in I \setminus \{0\}$ this implies $i = 0$. Thus I satisfies (d).

It follows from (a), (b), and (c) by induction that there exists a point $i \in I$ such that $\sigma(i) \in \{i-1, i+1\}$. Assume first that $\sigma(i) = i+1$, denote by A_i the arc in A from x_i to x_{i+1}, and denote by B_i the arc in B from x_i to x_{i+1}. If $i = 0$ it follows from (d) that $x_i = x_0 = x$ and $x_{i+1} = x_n = y$, in contradiction to $A \cap B \neq \{x, y\}$. Hence $i \neq 0$ and it follows from (d) that $0 < i < i+1 < n$. The arcs A_i and B_i satisfy

$$A_i \cap B = A \cap B_i = \{x_i, x_{i+1}\}.$$

Let D_i be the connected component of $\Sigma \setminus (A \cup B)$ that contains the points slightly to the left of A_i. This component is bounded by A_i and B_i. Moreover, the function w is positive on D_i. Hence it follows from Propostion 6.12 that D_i is diffeomorphic to the open unit disc in $\mathbb{C}$. Let $\mathrm{w}_i(z) := 1$ for $z \in D_i$ and $\mathrm{w}_i(z) := 0$ for $z \in \Sigma \setminus \overline{D}_i$. Then the combinatorial lune

$$\Lambda_i := (x_i, x_{i+1}, A_i, B_i, \mathrm{w}_i)$$

satisfies (43) and $\mathrm{w}_i \leq \mathrm{w}$.

Now assume $\sigma(i) = i - 1$, denote by A_i the arc in A from x_{i-1} to x_i, and denote by B_i the arc in B from x_{i-1} to x_i. Thus the orientation of A_i (from x_{i-1} to x_i) agrees with the orientation of A while the orientation of B_i is opposite to the orientation of B. Moreover, we have $0 < i - 1 < i < n$. The arcs A_i and B_i satisfy

$$A_i \cap B = A \cap B_i = \{x_{i-1}, x_i\}.$$

Let D_i be the connected component of $\Sigma \setminus (A \cup B)$ that contains the points slightly to the left of A_i. This component is again bounded by A_i and B_i, the function w is positive on D_i, and so D_i is diffeomorphic to the open unit disc in $\mathbb{C}$ by Propostion 6.12. Let $\mathrm{w}_i(z) := 1$ for $z \in D_i$ and $\mathrm{w}_i(z) := 0$ for $z \in \Sigma \setminus \overline{D}_i$. Then the combinatorial lune

$$\Lambda_i := (x_{i-1}, x_i, A_i, B_i, \mathrm{w}_i)$$

satisfies (43) and $\mathrm{w}_i \leq \mathrm{w}$. This proves Lemma 8.1. $\square$

Lemma 8.2. *Assume (H). Let u be a smooth (α, β)-lune whose (α, β)-trace $\Lambda_u = (x, y, \mathrm{w})$ is a combinatorial (α, β)-lune. Let $\gamma : [0, 1] \to \mathbb{D}$ be a smooth path such that*

$$\gamma(0) \in (\mathbb{D} \cap \mathbb{R}) \setminus \{\pm 1\}, \quad \gamma(1) \in (\mathbb{D} \cap S^1) \setminus \{\pm 1\}, \quad u(\gamma(t)) \notin A$$

for $0 < t < 1$. Then $\mathrm{w}(u(\gamma(t))) = 1$ for t near 1.

Proof. Denote $A := u(\mathbb{D} \cap \mathbb{R})$. Since Λ is a combinatorial (α, β)-lune we have $x, y \notin u(\mathrm{int}(\mathbb{D}))$. Hence $u^{-1}(A)$ is a union of embedded arcs, each connecting two points in $\mathbb{D} \cap S^1$. If $\mathrm{w}(u(\gamma(t)) \geq 2$ for t close to 1, then $\gamma(1) \in \mathbb{D} \cap S^1$ is separated from $\mathbb{D} \cap \mathbb{R}$ by one these arcs in $\mathbb{D} \setminus \mathbb{R}$. This proves Lemma 8.2. $\square$

For each combinatorial (α, β)-lune Λ the integer $\nu(\Lambda)$ denotes the number of equivalence classes of smooth (α, β)-lunes u with $\Lambda_u = \Lambda$.

Lemma 8.3. *Assume (H) and suppose that Σ is simply connected. Let $\Lambda = (x, y, \mathrm{w})$ be a combinatorial (α, β)-lune with boundary $\partial\Lambda = (x, y, A, B)$ such that $A \cap B \neq \{x, y\}$. Then there exists an embedded loop β', isotopic to β and transverse to α, and a combinatorial (α, β')-lune $\Lambda' = (x, y, \mathrm{w}')$ with boundary $\partial\Lambda' = (x, y, A, B')$ such that*

$$\#(A' \cap B') < \#(A \cap B), \qquad \nu(\Lambda) = \nu(\Lambda').$$

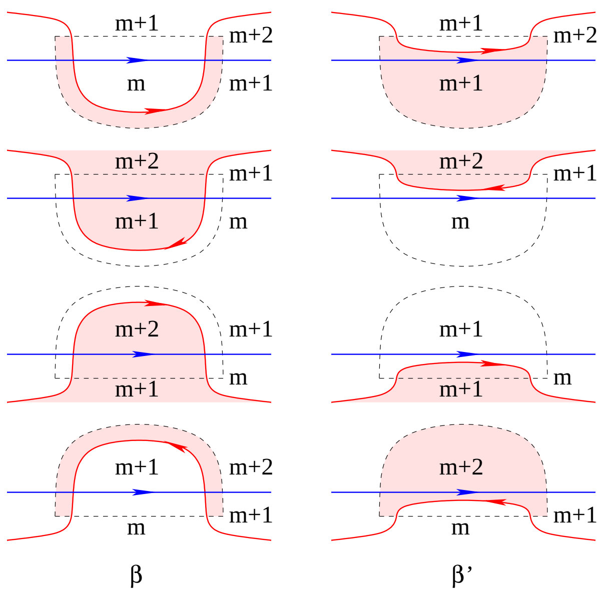

Figure 16: Deformation of a lune.

Proof. By Lemma 8.1 there exists a combinatorial (α, β)-lune

$$\Lambda_0 = (x_0, y_0, \mathrm{w}_0), \qquad \partial\Lambda_0 = (x_0, y_0, A_0, B_0),$$

that satisfies $\mathrm{w} \geq \mathrm{w}_0$ and (43). In particular, we have

$$A_0 \cap B_0 = \{x_0, y_0\}$$

and so, by Proposition 7.4, there is an embedded smooth lune $u_0 : \mathbb{D} \to \Sigma$ with bottom boundary A_0 and top boundary B_0. As in the proof of Step 1 in Proposition 7.7 we use this lune to remove the intersection points x_0 and y_0 by an isotopy of B, supported in a small neighborhood of the image of u_0. This isotopy leaves the number $\nu(\Lambda)$ unchanged. More precisely, extend u_0 to an embedding (still denoted by u_0) of the open set

$$\mathbb{D}_\varepsilon := \{z \in \mathbb{C} \,|\, \mathrm{Im}\, z > -\varepsilon, |z| < 1 + \varepsilon\}$$

for $\varepsilon > 0$ sufficiently small such that

$$u_0(\mathbb{D}_\varepsilon) \cap B = u_0(\mathbb{D}_\varepsilon \cap S^1), \qquad u_0(\mathbb{D}_\varepsilon) \cap A = u_0(\mathbb{D}_\varepsilon \cap \mathbb{R}),$$

$$u(\{z \in \mathbb{D}_\varepsilon \,|\, |z| > 1\}) \cap \beta = \emptyset, \quad u(\{z \in \mathbb{D}_\varepsilon \,|\, \mathrm{Re}\, z < 0\}) \cap \alpha = \emptyset.$$

Choose a smooth cutoff function $\rho : \mathbb{D}_\varepsilon \to [0,1]$ that vanishes near the boundary of $\mathbb{D}_\varepsilon$ and is equal to one on $\mathbb{D}$. Consider the vector field ξ on Σ that vanishes outside $u_0(\mathbb{D}_\varepsilon)$ and satisfies

$$u_0^*\xi(z) = -\rho(z)\mathbf{i}.$$

Let $\psi_t : \Sigma \to \Sigma$ be the isotopy generated by ξ and, for $T > 0$ sufficiently large, define

$$\beta' := \psi_T(\beta), \qquad B' := \psi_T(B), \qquad \Lambda' := (x, y, A, B', \mathrm{w}'),$$

where $\mathrm{w}' : \Sigma \setminus (\alpha \cup \beta') \to \mathbb{Z}$ is the unique two-chain that agrees with w on $\Sigma \setminus u_0(\mathbb{D}_\varepsilon)$. Thus w' corresponds to the homotopy from A to B determined by w followed by the homotopy ψ_t from B to B'. Then Λ' is a combinatorial (α, β')-lune. If $u : \mathbb{D} \to \Sigma$ is a smooth (α, β)-lune let $U \subset \mathbb{D}$ be the unique component of $u^{-1}(u_0(\mathbb{D}_\varepsilon))$ that contains an arc in $\mathbb{D} \cap S^1$. Then U does not intersect $\mathbb{D} \cap \mathbb{R}$. (See Figure 16.) Hence the map $u' : \mathbb{D} \to \Sigma$, defined by

$$u'(z) := \begin{cases} \psi_T(u(z)), & \text{if } z \in U, \\ u(z), & \text{if } z \in \mathbb{D} \setminus U, \end{cases}$$

is a smooth (α, β')-lune such that $\Lambda_{u'} = \Lambda'$.

We claim that the map $u \mapsto u'$ defines a one-to-one correspondence between smooth (α, β)-lunes u such that $\Lambda_u = \Lambda$ and smooth (α, β')-lunes u' such that $\Lambda_{u'} = \Lambda'$. The map $u \mapsto u'$ is obviously injective. To prove that it is surjective we choose a smooth (α, β')-lune u' such that $\Lambda_{u'} = \Lambda'$. Denote by

$$U' \subset \mathbb{D}$$

the unique connected component of ${u'}^{-1}(u_0(\mathbb{D}_\varepsilon))$ that contains an arc in $\mathbb{D} \cap S^1$. There are four cases as depicted in Figure 16. In two of these cases (second and third row) we have $u'(U') \cap \alpha = \emptyset$ and hence $U' \cap \mathbb{R} = \emptyset$. In the casees where $u'(U') \cap \alpha \neq \emptyset$ it follows from an orientation argument (fourth row) and from Lemma 8.2 (first row) that U' cannot intersect $\mathbb{D} \cap \mathbb{R}$. Thus we have shown that U' does not intersect $\mathbb{D} \cap \mathbb{R}$ in all four cases. This implies that u' is in the image of the map $u \mapsto u'$. Hence the map $u \mapsto u'$ is bijective as claimed, and hence

$$\nu(\Lambda) = \nu(\Lambda').$$

This proves Lemma 8.3. $\square$

Proof of Theorems 6.7 and 6.8. Assume first that Σ is simply connected. We prove that (iii) implies (i) in Theorem 6.7. Let Λ be a combinatorial (α,β)-lune. By Lemma 8.3, reduce the number of intersection points of Λ, while leaving the number $\nu(\Lambda)$ unchanged. Continue by induction until reaching an embedded combinatorial lune in Σ. By Proposition 7.4, such a lune satisfies $\nu = 1$. Hence $\nu(\Lambda) = 1$. In other words, there is a smooth (α,β)-lune $u : \mathbb{D} \to \Sigma$, unique up to equivalence, such that $\Lambda_u = \Lambda$. Thus we have proved that (iii) implies (i). We have also proved, in the simply connected case, that u is uniquely determined by Λ_u up to equivalence. From now on we drop the hypothesis that Σ is simply connected.

We prove Theorem 6.8. Let $u : \mathbb{D} \to \Sigma$ and $u' : \mathbb{D} \to \Sigma$ be smooth (α,β)-lunes such that

$$\Lambda_u = \Lambda_{u'}.$$

Let $\widetilde{u} : \mathbb{D} \to \widetilde{\Sigma}$ and $\widetilde{u}' : \mathbb{D} \to \widetilde{\Sigma}$ be lifts to the universal cover such that

$$\widetilde{u}(-1) = \widetilde{u}'(-1).$$

Then $\Lambda_{\widetilde{u}} = \Lambda_{\widetilde{u}'}$. Hence, by what we have already proved, $\widetilde{u}$ is equivalent to $\widetilde{u}'$ and hence u is equivalent to u'. This proves Theorem 6.8.

We prove that (i) implies (ii) in Theorem 6.7. Let $u : \mathbb{D} \to \Sigma$ be a smooth (α,β)-lune and denote by $\Lambda_u =: (x, y, A, B, \mathrm{w})$ be its (α,β)-trace. Then $\mathrm{w}(z) = \#u^{-1}(z)$ is the counting function of u and hence is nonnegative. For $0 < t \le 1$ define the curve $\lambda_t : [0,1] \to \mathbb{R}\mathrm{P}^1$ by

$$du(-\cos(\pi s), t\sin(\pi s))\lambda_t(s) := \mathbb{R}\frac{\partial}{\partial s}u(-\cos(\pi s), t\sin(\pi s)), \quad 0 \le s \le 1.$$

For $t = 0$ use the same definition for $0 < s < 1$ and extend the curve continuously to the closed interval $0 \le s \le 1$. Then

$$\begin{aligned} \lambda_0(s) &= du(-\cos(\pi s), 0)^{-1} T_{u(-\cos(\pi s),0)}\alpha, \\ \lambda_1(s) &= du(-\cos(\pi s), \sin(\pi s))^{-1} T_{u(-\cos(\pi s),\sin(\pi s))}\beta, \end{aligned} \qquad 0 \le s \le 1.$$

The Viterbo–Maslov index $\mu(\Lambda_u)$ is, by definition, the relative Maslov index of the pair of Lagrangian paths (λ_0, λ_1), denoted by $\mu(\lambda_0, \lambda_1)$ (see Definition 3.1 above or [39, 30]). Hence it follows from the homotopy axiom for the relative Maslov index that

$$\mu(\Lambda_u) = \mu(\lambda_0, \lambda_1) = \mu(\lambda_0, \lambda_t)$$

for every $t > 0$. Choosing t sufficiently close to zero we find that $\mu(\Lambda_u) = 1$.

We prove that (iii) implies (ii) in Theorem 6.7. If $\Lambda = (x, y, \mathrm{w})$ is a combinatorial (α, β)-lune, then $\mathrm{w} \geq 0$ by (I) in Definition 6.6. Moreover, by the trace formula (37), we have

$$\mu(\Lambda) = \frac{m_x(\Lambda) + m_y(\Lambda)}{2}.$$

It follows from (II) and (III) in Definition 6.6 that $m_x(\Lambda) = m_y(\Lambda) = 1$ and hence $\mu(\Lambda) = 1$. Thus we have proved that (iii) implies (ii).

We prove that (ii) implies (iii) in Theorem 6.7. Let $\Lambda = (x, y, \mathrm{w})$ be an (α, β)-trace such that $\mathrm{w} \geq 0$ and $\mu(\Lambda) = 1$. Denote $\nu_\alpha := \partial \mathrm{w}|_{\alpha \setminus \beta}$ and $\nu_\beta := -\partial \mathrm{w}|_{\beta \setminus \alpha}$. Reversing the orientation of α or β, if necessary, we may assume that $\nu_\alpha \geq 0$ and $\nu_\beta \geq 0$. Let $\varepsilon_x, \varepsilon_y \in \{\pm 1\}$ be the intersection indices of α and β at x, y with these orientations, and let

$$n_\alpha := \min \nu_\alpha \geq 0, \qquad n_\beta := \min \nu_\beta \geq 0.$$

As before, denote by m_x (respectively m_y) the sum of the four values of w encountered when walking along a small circle surrounding x (respectvely y). Since the Viterbo–Maslov index of Λ is odd, we have $\varepsilon_x \neq \varepsilon_y$ and thus $x \neq y$. This shows that Λ satisfies the arc condition if and only if $n_\alpha = n_\beta = 0$.

We prove that Λ satisfies (II). Suppose, by contradiction, that Λ does not satisfy (II). Then $\varepsilon_x = -1$ and $\varepsilon_y = 1$. This implies that the values of w near x are given by k, $k + n_\alpha + 1$, $k + n_\alpha + n_\beta + 1$, $k + n_\beta + 1$ for some integer k. Since $\mathrm{w} \geq 0$ these numbers are all nonnegative. Hence $k \geq 0$ and hence $m_x \geq 3$. The same argument shows that $m_y \geq 3$ and, by the trace formula (37), we have $\mu(\Lambda) = (m_x(\Lambda) + m_y(\Lambda))/2 \geq 3$, in contradiction to our hypothesis. This shows that Λ satisfies (II).

We prove that Λ satisfies the arc condition and (III). By (II) we have $\varepsilon_x = 1$. Hence the values of w near x in counterclockwise order are given by k_x, $k_x + n_\alpha + 1$, $k_x + n_\alpha - n_\beta$, $k_x - n_\beta$ for some integer $k_x \geq n_\beta \geq 0$. This implies

$$m_x(\Lambda) = 4k_x - 2n_\beta + 2n_\alpha + 1$$

and, similarly, $m_y(\Lambda) = 4k_y - 2n_\beta + 2n_\alpha + 1$ for some integer $k_y \geq n_\beta \geq 0$. Hence, by the trace formula (37), we have

$$1 = \mu(\Lambda) = \frac{m_x(\Lambda) + m_y(\Lambda)}{2} = k_x + (k_x - n_\beta) + k_y + (k_y - n_\beta) + 2n_\alpha + 1.$$

Hence $k_x = k_y = n_\alpha = n_\beta = 0$ and so Λ satisfies the arc condition and (III). Thus we have shown that (ii) implies (i). This proves Theorem 6.7. $\square$

Example 8.4. The arguments in the proof of Theorem 6.7 can be used to show that, if Λ is an (α, β)-trace with $\mu(\Lambda) = 1$, then (I) $\Longrightarrow$ (III) $\Longrightarrow$ (II). Figure 17 shows three (α, β)-traces that satisfy the arc condition and have Viterbo–Maslov index one but do not satisfy (I); one that still satisfies (II) and (III), one that satisfies (II) but not (III), and one that satisfies neither (II) nor (III). Figure 18 shows an (α, β)-trace of Viterbo–Maslov index two that satisfies (I) and (III) but not (II). Figure 19 shows an (α, β)-trace of Viterbo–Maslov index three that satisfies (I) and (II) but not (III).

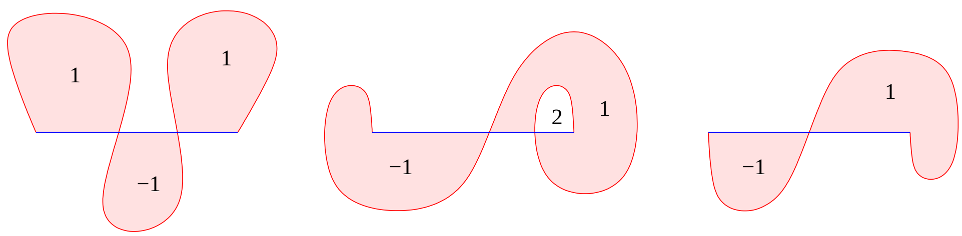

Figure 17: Three (α, β)-traces with Viterbo–Maslov index one.

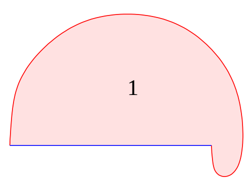

Figure 18: An (α, β)-trace with Viterbo–Maslov index two.

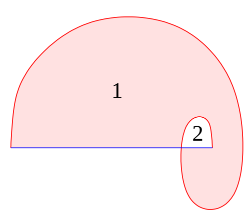

Figure 19: An (α, β)-trace with Viterbo–Maslov index three.

We close this section with two results about lunes that will be useful below.

Proposition 8.5. *Assume (H) and suppose that α and β are noncontractible nonisotopic transverse embedded circles and let $x, y \in \alpha \cap \beta$. Then there is at most one (α, β)-trace from x to y that satisfies the arc condition. Hence, by Theorems 6.5 and 6.8, there is at most one equivalence class of smooth (α, β)-lunes from x to y.*

Proof. Let

$$\alpha = \alpha_1 \cup \alpha_2, \qquad \beta = \beta_1 \cup \beta_2,$$

where α_1 and α_2 are the two arcs of α with endpoints x and y, and similarly for β. Assume that the quadruple $(x, y, \alpha_1, \beta_1)$ is an (α, β)-trace. Then α_1 is homotopic to β_1 with fixed endpoints. Since α is not contractible, α_2 is not homotopic to β_1 with fixed endpoints. Since β is not contractible, β_2 is not homotopic to α_1 with fixed endpoints. Since α is not isotopic to β, α_2 is not homotopic to β_2 with fixed endpoints. Hence the quadruple $(x, y, \alpha_i, \beta_j)$ is not an (α, β)-trace unless $i = j = 1$. This proves Proposition 8.5. □

The hypotheses that the loops α and β are not contractible and not isotopic to each other cannot be removed in Proposition 8.5. A pair of isotopic circles with precisely two intersection points is an example. Another example is a pair consisting of a contractible and a non-contractible loop, again with precisely two intersection points.

Proposition 8.6. *Assume (H). If there is a smooth (α, β)-lune then there is a primitive (α, β)-lune.*

Proof. The proof has three steps.

Step 1. *If α or β is a contractible embedded circle and $\alpha \cap \beta \neq \emptyset$ then there exists a primitive (α, β)-lune.*

Assume α is a contractible embedded circle. Then, by a theorem of Epstein [9], there exists an embedded closed disc $D \subset \Sigma$ with boundary $\partial D = \alpha$. Since α and β intersect transversally, the set $D \cap \beta$ is a finite union of arcs. Let $\mathcal{A}$ be the set of all arcs $A \subset \alpha$ which connect the endpoints of an arc $B \subset D \cap \beta$. Then $\mathcal{A}$ is a nonempty finite set, partially ordered by inclusion. Let $A_0 \subset \alpha$ be a minimal element of $\mathcal{A}$ and $B_0 \subset D \cap \beta$ be the arc with the same endpoints as A_0. Then A_0 and B_0 bound a primitive (α, β)-lune. This proves Step 1 when α is a contractible embedded circle. When β is a contractible embedded circle the proof is analogous.

Step 2. *Assume α and β are not contractible embedded circles. If there exists a smooth (α, β)-lune then there exists an embedded (α, β)-lune u such that $u^{-1}(\alpha) = \mathbb{D} \cap \mathbb{R}$.*

Let $v : \mathbb{D} \to \Sigma$ be a smooth (α, β)-lune. Then the set

$$X := v^{-1}(\alpha) \subset \mathbb{D}$$

is a smooth 1-manifold with boundary $\partial X = v^{-1}(\alpha) \cap S^1$. The interval $\mathbb{D} \cap \mathbb{R}$ is one component of X and no component of X is a circle. (If $X_0 \subset X$ is a circle, then $v|_{X_0} : X_0 \to \Sigma$ is a contractible loop covering α finitely many times. Hence, by Lemma A.3 in the appendix, it would follow that α is a contractible embedded circle, in contradiction to the hypothesis of Step 2.) Write

$$\partial X = \{e^{\mathbf{i}\theta_1}, \ldots, e^{\mathbf{i}\theta_n}\}, \qquad \pi = \theta_1 > \theta_2 > \cdots > \theta_{n-1} > \theta_n = 0.$$

Then there is a permutation $\sigma \in S_n$ such that the arc of $v^{-1}(\alpha)$ that starts at $e^{\mathbf{i}\theta_j}$ ends at $e^{\mathbf{i}\theta_{\sigma(j)}}$. This permutation satisfies $\sigma \circ \sigma = \mathrm{id}$, $\sigma(1) = n$, and

$$j < k < \sigma(j) \qquad \Longrightarrow \qquad j < \sigma(k) < \sigma(j).$$

Hence, by induction, there exists a $j \in \{1, \ldots, n-1\}$ such that $\sigma(j) = j + 1$. Let $X_0 \subset X$ be the submanifold with boundary points $e^{\mathbf{i}\theta_j}$ and $e^{\mathbf{i}\theta_{j+1}}$ and denote

$$Y_0 := \left\{ e^{\mathbf{i}\theta} \,|\, \theta_{j+1} \le \theta \le \theta_j \right\}.$$

Then the closure of the domain $\Delta \subset \mathbb{D}$ bounded by X_0 and Y_0 is diffeomorphic to the half disc and $\overline{\Delta} \cap v^{-1}(\alpha) = X_0$. Hence there exists an orientation preserving embedding $\varphi : \mathbb{D} \to \mathbb{D}$ that maps $\mathbb{D} \cap \mathbb{R}$ onto X_0 and maps $\mathbb{D} \cap S^1$ onto Y_0. It follows that

$$u := v \circ \varphi : \mathbb{D} \to \Sigma$$

is a smooth (α, β)-lune such that

$$u^{-1}(\alpha) = \varphi^{-1}(\overline{\Delta} \cap v^{-1}(\alpha)) = \varphi^{-1}(X_0) = \mathbb{D} \cap \mathbb{R}.$$

Moreover, $\Lambda_u = (x, y, A, B, \mathrm{w})$ with

$$x := v(e^{\mathbf{i}\theta_j}), \qquad y := v(e^{\mathbf{i}\theta_{j+1}}), \qquad A := v(X_0), \qquad B := v(Y_0).$$

Since $A \cap B = \alpha \cap B = \{x, y\}$, it follows from Proposition 7.4 that u is an embedding. This proves Step 2.

Step 3. *Assume α and β are not contractible embedded circles. If there exists an embedded (α, β)-lune u such that $u^{-1}(\alpha) = \mathbb{D} \cap \mathbb{R}$ then there exists a primitive (α, β)-lune.*

Repeat the argument in the proof of Step 2 with v replaced by u and the set $v^{-1}(\alpha)$ replaced by the 1-manifold

$$Y := u^{-1}(\beta) \subset \mathbb{D}$$

with boundary

$$\partial Y = u^{-1}(\beta) \cap \mathbb{R}.$$

The argument produces an arc $Y_0 \subset Y$ with boundary points $a < b$ such that the closed interval $X_0 := [a, b]$ intersects Y only in the endpoints. Hence the arcs $A_0 := u(X_0)$ and $B_0 := u(Y_0)$ bound a primitive (α, β)-lune. This proves Step 3 and Proposition 8.6. $\square$

III. Floer Homology

9 Combinatorial Floer Homology

We assume throughout this section that Σ is an oriented 2-manifold without boundary and that $\alpha, \beta \subset \Sigma$ are noncontractible nonisotopic transverse embedded circles. We orient α and β. There are three ways we can count the number of points in their intersection:

- The **numerical intersection number** $\mathrm{num}\,(\alpha, \beta)$ is the actual number of intersection points.

- The **geometric intersection number** $\mathrm{geo}\,(\alpha, \beta)$ is defined as the minimum of the numbers $\mathrm{num}\,(\alpha, \beta')$ over all embedded loops β' that are transverse to α and isotopic to β.

- The **algebraic intersection number** $\mathrm{alg}\,(\alpha, \beta)$ is the sum

  $$\alpha \cdot \beta = \sum_{x \in \alpha \cap \beta} \pm 1$$

  where the plus sign is chosen iff the orientations match in the direct sum $T_x\Sigma = T_x\alpha \oplus T_x\beta$.

Note that $|\mathrm{alg}\,(\alpha, \beta)| \leq \mathrm{geo}\,(\alpha, \beta) \leq \mathrm{num}\,(\alpha, \beta)$.

Theorem 9.1. *Define a chain complex $\partial : \mathrm{CF}(\alpha, \beta) \to \mathrm{CF}(\alpha, \beta)$ by*

$$\mathrm{CF}(\alpha, \beta) = \bigoplus_{x \in \alpha \cap \beta} \mathbb{Z}_2 x, \qquad \partial x = \sum_y n(x, y) y, \tag{45}$$

where $n(x, y)$ denotes the number modulo two of equivalence classes of smooth (α, β)-lunes from x to y. Then

$$\partial \circ \partial = 0.$$

The homology group of this chain complex is denoted by

$$\mathrm{HF}(\alpha, \beta) := \ker \partial / \mathrm{im} \partial$$

and is called the **Combinatorial Floer Homology** *of the pair (α, β).*

Proof. See Section 10 page 81. □

Theorem 9.2. *Combinatorial Floer homology is invariant under isotopy: If $\alpha', \beta' \subset \Sigma$ are noncontractible transverse embedded circles such that α is isotopic to α' and β is isotopic to β' then*

$$\mathrm{HF}(\alpha, \beta) \cong \mathrm{HF}(\alpha', \beta').$$

Proof. See Section 11 page 84. ☐

Theorem 9.3. *Combinatorial Floer homology is isomorphic to the original analytic Floer homology. In fact, the two chain complexes agree.*

Proof. See Section 12 page 100. ☐

Corollary 9.4. *If* $\mathrm{geo}\,(\alpha, \beta) = \mathrm{num}\,(\alpha, \beta)$ *there is no smooth* (α, β)*-lune.*

Proof. If there exists a smooth (α, β)-lune then, by Proposition 8.6, there exists a primitive (α, β)-lune and hence there exists an embedded curve β' that is isotopic to β and satisfies $\mathrm{num}\,(\alpha, \beta') < \mathrm{num}\,(\alpha, \beta)$. This contradicts our hypothesis. ☐

Corollary 9.5. $\dim \mathrm{HF}(\alpha, \beta) = \mathrm{geo}\,(\alpha, \beta)$.

Proof. By Theorem 9.2 we may assume that $\mathrm{num}\,(\alpha, \beta) = \mathrm{geo}\,(\alpha, \beta)$. In this case there is no (α, β)-lune by Corollary 9.4. Hence the Floer boundary operator is zero, and hence the dimension of

$$\mathrm{HF}(\alpha, \beta) \cong \mathrm{CF}(\alpha, \beta)$$

is the geometric intersection number $\mathrm{geo}\,(\alpha, \beta)$. ☐

Corollary 9.6. *If* $\mathrm{geo}\,(\alpha, \beta) < \mathrm{num}\,(\alpha, \beta)$ *there is a primitive* (α, β)*-lune.*

Proof. By Corollary 9.5, the Floer homology group has dimension

$$\dim\ \mathrm{HF}(\alpha, \beta) = \mathrm{geo}\,(\alpha, \beta).$$

Since the Floer chain complex has dimension

$$\dim\ \mathrm{CF}(\alpha, \beta) = \mathrm{num}\,(\alpha, \beta)$$

it follows that the Floer boundary operator is nonzero. Hence there exists a smooth (α, β)-lune and hence, by Proposition 8.6, there exists a primitive (α, β)-lune. ☐

Remark 9.7 (Action Filtration). Consider the space

$$\Omega_{\alpha,\beta} := \{x \in C^\infty([0,1],\Sigma) \,|\, x(0) \in \alpha,\ x(1) \in \beta\}$$

of paths connecting α to β. Every intersection point $x \in \alpha \cap \beta$ determines a constant path in $\Omega_{\alpha,\beta}$ and hence a component of $\Omega_{\alpha,\beta}$. In general, $\Omega_{\alpha,\beta}$ is not connected and different intersection points may determine different components (see [29] for the case $\Sigma = \mathbb{T}^2$). By Proposition A.1 in Appendix A, each component of $\Omega_{\alpha,\beta}$ is simply connected. Now fix a positive area form ω on Σ and define a 1-form Θ on $\Omega_{\alpha,\beta}$ by

$$\Theta(x;\xi) := \int_0^1 \omega(\dot{x}(t), \xi(t))\, dt$$

for $x \in \Omega_{\alpha,\beta}$ and $\xi \in T_x\Omega_{\alpha,\beta}$. This form is closed and hence exact. Let

$$\mathcal{A} : \Omega_{\alpha,\beta} \to \mathbb{R}$$

be a function whose differential is Θ. Then the critical points of $\mathcal{A}$ are the zeros of Θ. These are the constant paths and hence the intersection points of α and β. If $x, y \in \alpha \cap \beta$ belong to the same connected component of $\Omega_{\alpha,\beta}$ then

$$\mathcal{A}(x) - \mathcal{A}(y) = \int u^*\omega$$

where $u : [0,1] \times [0,1] \to \Sigma$ is any smooth function that satisfies

$$u(0,t) = x(t), \qquad u(0,1) = y(t), \qquad u(s,0) \in \alpha, \qquad u(s,1) \in \beta$$

for all s and t (i.e. the map $s \mapsto u(s,\cdot)$ is a path in $\Omega_{\alpha,\beta}$ connecting x to y). In particular, if x and y are the endpoints of a smooth lune then $\mathcal{A}(x) - \mathcal{A}(y)$ is the area of that lune. Figure 20 shows that there is no upper bound (independent of α and β in fixed isotopy classes) on the area of a lune.

Proposition 9.8. *Define a relation $\prec$ on $\alpha \cap \beta$ by $x \prec y$ if and only if there is a sequence $x = x_0, x_1, \ldots, x_n = y$ in $\alpha \cap \beta$ such that, for each i, there is a lune from x_i to x_{i-1} (see Figure 21). Then $\prec$ is a strict partial order.*

Proof. Since there is an (α,β)-lune from x_i to x_{i-1} we have $\mathcal{A}(x_{i-1}) < \mathcal{A}(x_i)$ for every i and hence, by induction, $\mathcal{A}(x_0) < \mathcal{A}(x_k)$. $\square$

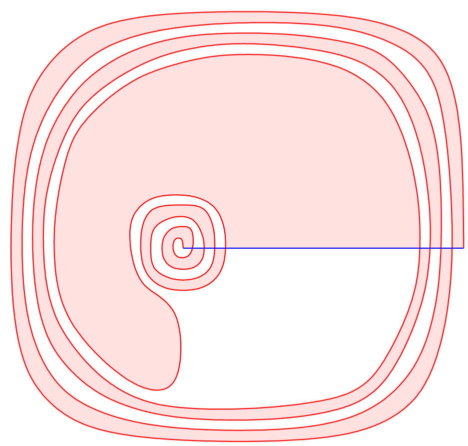

Figure 20: A lune of large area.

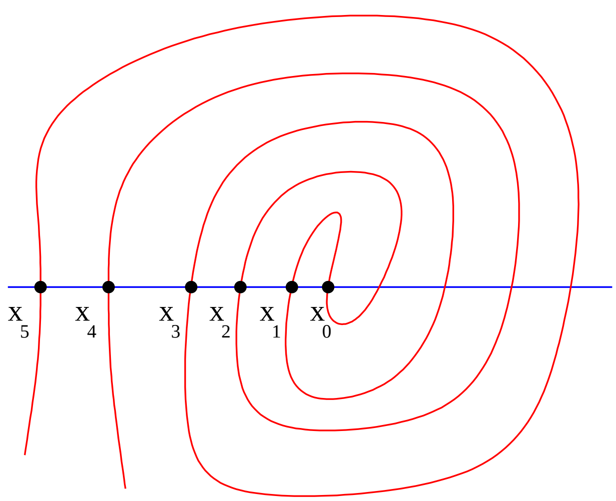

Figure 21: Lunes from x_i to x_{i-1}.

Remark 9.9 (Mod Two Grading). The endpoints of a lune have opposite intersection indices. Thus we may choose a $\mathbb{Z}/2\mathbb{Z}$-grading of $\mathrm{CF}(\alpha,\beta)$ by first choosing orientations of α and β and then defining $\mathrm{CF}_0(\alpha,\beta)$ to be generated by the intersection points with intersection index $+1$ and $\mathrm{CF}_1(\alpha,\beta)$ to be generated by the intersection points with intersection index -1. Then the boundary operator interchanges these two subspaces and we have

$$\mathrm{alg}\,(\alpha,\beta) = \dim \mathrm{HF}_0(\alpha,\beta) - \dim \mathrm{HF}_1(\alpha,\beta).$$

Remark 9.10 (Integer Grading). Since each component of the path space $\Omega_{\alpha,\beta}$ is simply connected the $\mathbb{Z}/2\mathbb{Z}$-grading in Remark 9.9 can be refined to an integer grading. The grading is only well defined up to a global shift and the relative grading is given by the Viterbo–Maslov index. Then we obtain

$$\mathrm{alg}\,(\alpha,\beta) = \chi(\mathrm{HF}(\alpha,\beta)) = \sum_{i\in\mathbb{Z}} (-1)^i \dim \mathrm{HF}_i(\alpha,\beta).$$

Figure 21 shows that there is no upper or lower bound on the relative index in the combinatorial Floer chain complex. Figure 22 shows that there is no upper bound on the dimension of $\mathrm{CF}_i(\alpha,\beta)$.

In the case of the 2-torus $\Sigma = \mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ the shift in the integer grading can be fixed using Seidel's notion of a graded Lagrangian submanifold [33]. Namely, the tangent bundle of $\mathbb{T}^2$ is trivial so that each tangent space is equipped with a canonical isomorphism to $\mathbb{R}^2$. Hence every embedded circle $\alpha \subset \mathbb{T}^2$ determines a map $\alpha \to \mathbb{R}\mathrm{P}^1 : z \mapsto T_z\alpha$. A grading of α is a lift of this map to the universal cover of $\mathbb{R}\mathrm{P}^1$. A choice of gradings for α and β can be used to fix an integer grading of the combinatorial Floer homology.

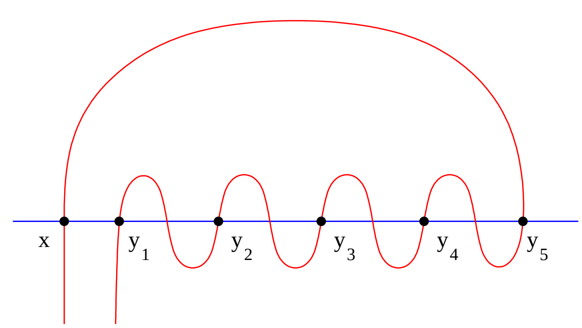

Figure 22: Lunes from x to y_i.

Remark 9.11 (**Integer Coefficients**). One can define combinatorial Floer homology with integer coefficients as follows. Fix an orientation of α. Then each (α,β)-lune $u:\mathbb{D}\to\Sigma$ comes with a sign

$$\nu(u) := \begin{cases} +1, & \text{if the arc } u|_{\mathbb{D}\cap\mathbb{R}}:\mathbb{D}\cap\mathbb{R}\to\alpha \text{ is orientation preserving,} \\ -1, & \text{if the arc } u|_{\mathbb{D}\cap\mathbb{R}}:\mathbb{D}\cap\mathbb{R}\to\alpha \text{ is orientation reversing.} \end{cases}$$

Now define the chain complex by

$$\mathrm{CF}(\alpha,\beta;\mathbb{Z}) = \bigoplus_{x\in\alpha\cap\beta}\mathbb{Z}x,$$

and

$$\partial x := \sum_{y\in\alpha\cap\beta} n(x,y;\mathbb{Z})y, \qquad n(x,y;\mathbb{Z}) := \sum_{[u]}\nu(u),$$

where the sum runs over all equivalence classes $[u]$ of smooth (α,β)-lunes from x to y. The results of Section 10 show that Theorem 9.1 remains valid with this refinement, and the results of Section 11 show that Theorem 9.2 also remains valid. We will not discuss here any orientation issue for the analytic Floer theory and leave it to others to investigate the validity of Theorem 9.3 with integer coefficients.

10 Hearts

Definition 10.1. Let $x,z\in\alpha\cap\beta$. A **broken** (α,β)**-heart** from x to z is a triple

$$h=(u,y,v)$$

such that $y\in\alpha\cap\beta$, u is a smooth (α,β)-lune from x to y, and v is a smooth (α,β)-lune from y to z. The point y is called the **midpoint** of the heart.

By Theorem 6.8 the broken (α,β)-heart h is uniquely determined by the septuple

$$\Lambda_h := (x,y,z,u(\mathbb{D}\cap\mathbb{R}),v(\mathbb{D}\cap\mathbb{R}),u(\mathbb{D}\cap S^1),v(\mathbb{D}\cap S^1)).$$

Two broken (α,β)-hearts $h=(u,y,z)$ and $h'=(u',y',z')$ from x to z are called **equivalent** if $y'=y$, u' is equivalent to u, and v' is equivalent to v. The equivalence class of h is denoted by $[h]=([u],y,[v])$. The set of equivalence classes of broken (α,β)-hearts from x to z will be denoted by $\mathcal{H}(x,z)$.

Proposition 10.2. *Let $h = (u, y, v)$ be a broken (α, β)-heart from x to z and write $\Lambda_h =: (x, y, z, A_{xy}, A_{yz}, B_{xy}, B_{yz})$. Then exactly one of the following four alternatives (see Figure 23) holds:*

(**a**) $A_{xy} \cap A_{yz} = \{y\},\ B_{yz} \subsetneq B_{xy}$. (**b**) $A_{xy} \cap A_{yz} = \{y\},\ B_{xy} \subsetneq B_{yz}$.
(**c**) $B_{xy} \cap B_{yz} = \{y\},\ A_{yz} \subsetneq A_{xy}$. (**d**) $B_{xy} \cap B_{yz} = \{y\},\ A_{xy} \subsetneq A_{yz}$.

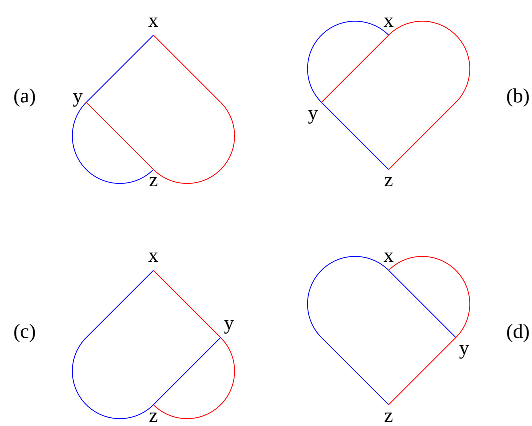

Figure 23: Four broken hearts.

Proof. The combinatorial (α, β)-lunes $\Lambda_{xy} := \Lambda_u$ and $\Lambda_{yz} := \Lambda_v$ have boundaries

$$\partial\Lambda_{xy} = (x, y, A_{xy}, B_{xy}), \qquad \partial\Lambda_{yz} = (y, z, A_{yz}, B_{yz})$$

and their catenation $\Lambda_{xz} := \Lambda_{xy} \# \Lambda_{yz} = (x, z, \mathrm{w}_{xy} + \mathrm{w}_{yz})$ has Viterbo–Maslov index two, by (38). Hence $m_x(\Lambda_{xy}) + m_x(\Lambda_{yz}) + m_z(\Lambda_{xy}) + m_z(\Lambda_{yz}) = 4$, by the trace formula (37). Since $m_x(\Lambda_{xy}) = m_z(\Lambda_{yz}) = 1$ this implies

$$m_x(\Lambda_{yz}) + m_z(\Lambda_{xy}) = 2. \tag{46}$$

By Proposition 9.8 we have $x \neq z$. Hence x cannot be an endpoint of Λ_{yz}. Thus $m_x(\Lambda_{yz}) \geq 2$ whenever $x \in A_{yz} \cup B_{yz}$ and $m_x(\Lambda_{yz}) \geq 4$ whenever $x \in A_{yz} \cap B_{yz}$. The same holds for $m_z(\Lambda_{xy})$. Hence it follows from (46) that Λ_{xy} and Λ_{yz} satisfy precisely one of the following conditions.

(a) $x \notin A_{yz} \cup B_{yz},\ z \in B_{xy} \setminus A_{xy}$. **(b)** $x \in B_{yz} \setminus A_{yz},\ z \notin A_{xy} \cup B_{xy}$.
(c) $x \notin A_{yz} \cup B_{yz},\ z \in A_{xy} \setminus B_{xy}$. **(d)** $x \in A_{yz} \setminus B_{yz},\ z \notin A_{xy} \cup B_{xy}$.

This proves Proposition 10.2. □

Let $N \subset \mathbb{C}$ be an embedded convex half disc such that

$$[0,1] \cup \mathbf{i}[0,\varepsilon) \cup (1+\mathbf{i}[0,\varepsilon)) \subset \partial N, \qquad N \subset [0,1]+\mathbf{i}[0,1]$$

for some $\varepsilon > 0$ and define

$$H := ([0,1]+\mathbf{i}[0,1]) \cup (\mathbf{i}+N) \cup (1+\mathbf{i}-\mathbf{i}N).$$

(See Figure 24.) The boundary of H decomposes as

$$\partial H = \partial_0 H \cup \partial_1 H$$

where $\partial_0 H$ denotes the boundary arc from 0 to $1+\mathbf{i}$ that contains the horizontal interval $[0,1]$ and $\partial_1 H$ denotes the arc from 0 to $1+\mathbf{i}$ that contains the vertical interval $\mathbf{i}[0,1]$.

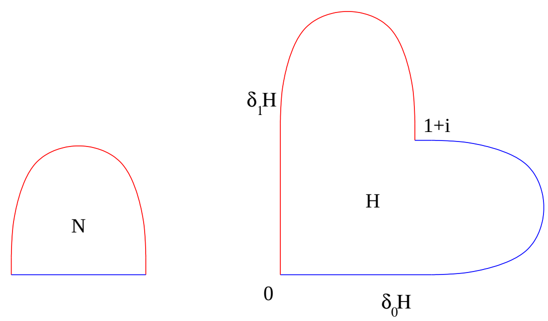

Figure 24: The domains N and H.

Definition 10.3. Let $x, z \in \alpha \cap \beta$. A **smooth** (α,β)**-heart of type (ac) from** x **to** z is an orientation preserving immersion $w : H \to \Sigma$ that satisfies

$$w(0) = x, \qquad w(1+\mathbf{i}) = z, \qquad w(\partial_0 H) \subsetneq \alpha, \qquad w(\partial_1 H) \subsetneq \beta. \qquad (ac)$$

Two smooth (α,β)-hearts $w, w' : H \to \Sigma$ are called **equivalent** iff there exists an orientation preserving diffeomorphism $\chi : H \to H$ such that

$$\chi(0) = 0, \qquad \chi(1+\mathbf{i}) = 1+\mathbf{i}, \qquad w' = w \circ \chi.$$

A **smooth** (α,β)**-heart of type (bd) from** x **to** z is a smooth (β,α)-heart of type (ac) from z to x. Let w be a smooth (α,β)-heart of type (ac) from x to y and $h = (u,y,v)$ be a broken (α,β)-heart from x to y of type (a) or (c).

The broken heart h is called **compatible** with the smooth heart w if there exist orientation preserving embeddings $\varphi : \mathbb{D} \to H$ and $\psi : \mathbb{D} \to H$ such that

$$\varphi(-1) = 0, \qquad \psi(1) = 1 + \mathbf{i}, \tag{47}$$

$$H = \varphi(\mathbb{D}) \cup \psi(\mathbb{D}), \qquad \varphi(\mathbb{D}) \cap \psi(\mathbb{D}) = \varphi(\partial\mathbb{D}) \cap \psi(\partial\mathbb{D}), \tag{48}$$

$$u = w \circ \varphi, \qquad v = w \circ \psi. \tag{49}$$

Lemma 10.4. *Let $h = (u, y, v)$ be a broken (α, β)-heart of type (a), write*

$$\Lambda_h =: (x, y, z, A_{xy}, A_{yz}, B_{xy}, B_{yz}),$$

and define A_{xz} and B_{xz} by

$$A_{xz} := A_{xy} \cup A_{yz}, \qquad B_{xy} =: B_{xz} \cup B_{yz}, \qquad B_{xz} \cap B_{yz} = \{z\}.$$

Let w be a smooth (α, β)-heart of type (ac) from x to z that is compatible with h and let $\varphi, \psi : \mathbb{D} \to H$ be embeddings that satisfy (47), (48), and (49). Then

$$\varphi(e^{\mathbf{i}\theta_1}) = 1 + \mathbf{i}, \tag{50}$$

where $\theta_1 \in [0, \pi]$ is defined by $u(e^{\mathbf{i}\theta_1}) = z$, and

$$\varphi(\mathbb{D}) \cap \psi(\mathbb{D}) = \psi(\mathbb{D} \cap S^1), \tag{51}$$

$$w(\partial_0 H) = A_{xz}, \qquad w(\partial_1 H) = B_{xz}. \tag{52}$$

Proof. By definition of a smooth heart of type (ac), $w(\partial_0 H)$ is arc in α connecting x to z and $w(\partial_1 H)$ is an arc in β connecting x to z. Moreover, by (49),

$$\varphi(\mathbb{D} \cap \mathbb{R}) \cup \psi(\mathbb{D} \cap \mathbb{R}) \subset w^{-1}(\alpha), \qquad \varphi(\mathbb{D} \cap S^1) \cup \psi(\mathbb{D} \cap S^1) \subset w^{-1}(\beta).$$

Now $w^{-1}(\alpha)$ is a union of disjoint embedded arcs, and so is $w^{-1}(\beta)$. One of the arcs in $w^{-1}(\alpha)$ contains $\partial_0 H$ and one of the arcs in $w^{-1}(\beta)$ contains $\partial_1 H$. Since $\varphi(-1) = 0 \in \partial_0 H$ and the arc

$$w \circ \varphi(\mathbb{D} \cap \mathbb{R}) = u(\mathbb{D} \cap \mathbb{R}) = A_{xy}$$

does not contain z we have

$$\varphi(\mathbb{D} \cap \mathbb{R}) \subset \partial_0 H, \qquad A_{xy} \subset w(\partial_0 H).$$

This implies the first equation in (52). Since $w\circ\varphi(\mathbb{D}\cap S^1)=u(\mathbb{D}\cap S^1)=B_{xy}$ is an arc containing z we have

$$\partial_1 H\subset\varphi(\mathbb{D}\cap S^1),\qquad w(\partial_1 H)\subset B_{xy}.$$

This implies the second equation in (52).

We prove (51). Choose $\theta_1\in[0,\pi]$ so that $u(e^{\mathbf{i}\theta_1})=z$ and denote

$$S_0:=\left\{e^{\mathbf{i}\theta}\,|\,0\le\theta\le\theta_1\right\},\qquad S_1:=\left\{e^{\mathbf{i}\theta}\,|\,\theta_1\le\theta\le\pi\right\},$$

So that

$$\mathbb{D}\cap S^1=S_0\cup S_1,\qquad u(S_0)=B_{yz},\qquad u(S_1)=B_{xz}.$$

Hence $w\circ\varphi(S_1)=u(S_1)=B_{xz}=w(\partial_1 H)$ and $0\in\varphi(S_1)\cap\partial_1 H$. Since w is an immersion it follows that

$$\varphi(S_1)=\partial_1 H,\qquad \varphi(e^{\mathbf{i}\theta_1})=1+\mathbf{i}=\psi(1).$$

This proves (50). Moreover, by (49),

$$w\circ\varphi(S_0)=u(S_0)=B_{yz}=v(\mathbb{D}\cap S^1)=w\circ\psi(\mathbb{D}\cap S^1)$$

and $1+\mathbf{i}$ is an endpoint of both arcs $\varphi(S_0)$ and $\psi(\mathbb{D}\cap S^1)$. Since w is an immersion it follows that

$$\psi(\mathbb{D}\cap S^1)=\varphi(S_0)\subset\varphi(\mathbb{D})\cap\psi(\mathbb{D}).$$

To prove the converse inclusion, let $\zeta\in\varphi(\mathbb{D})\cap\psi(\mathbb{D})$. Then, by definition of a smooth heart, $\zeta\in\varphi(\partial\mathbb{D})\cap\psi(\partial\mathbb{D})$. If $\zeta\in\varphi(\mathbb{D}\cap\mathbb{R})\cap\psi(\mathbb{D}\cap\mathbb{R})$ then $w(\zeta)\in A_{xy}\cap A_{yz}=\{y\}$ and hence $\zeta=\psi(-1)\in\psi(\mathbb{D}\cap S^1)$. Now suppose $\zeta=\varphi(e^{\mathbf{i}\theta})\in\psi(\mathbb{D}\cap\mathbb{R})$ for some $\theta\in[0,\pi]$. Then we claim that $\theta\le\theta_1$. To see this, consider the curve $\psi(\mathbb{D}\cap\mathbb{R})$. By (49), this curve is mapped to A_{yz} under w and it contains the point $\psi(1)=1+\mathbf{i}$. Hence $\psi(\mathbb{D}\cap\mathbb{R})\subset\partial_0 H\setminus\{0\}$. But if $\theta>\theta_1$ then $\varphi(e^{\mathbf{i}\theta})\in\partial_1 H\setminus\{1+\mathbf{i}\}$ and this set does not intersect $\partial_0 H\setminus\{0\}$. Thus we have proved that $\theta\le\theta_1$ and hence $\varphi(e^{\mathbf{i}\theta})\in\varphi(S_0)=\psi(\mathbb{D}\cap S^1)$, as claimed. This proves Lemma 10.4. □

Proposition 10.5. **(i)** *Let $h=(u,y,v)$ be a broken (α,β)-heart of type (a) or (c) from x to z. Then there exists a smooth (α,β)-heart w of type (ac) from x to z, unique up to equivalence, that is compatible with h.*

(ii) *Let w be a smooth (α,β)-heart of type (ac) from x to z. Then there exists precisely one equivalence class of broken (α,β)-hearts of type (a) from x to z that are compatible with w, and precisely one equivalence class of broken (α,β)-hearts of type (c) from x to z that are compatible with w.*

Proof. We prove (i). Write

$$\Lambda_h =: (x, y, z, A_{xy}, A_{yz}, B_{xy}, B_{yz})$$

and assume first that Λ_h satisfies (a). We prove the existence of $[w]$. Choose a Riemannian metric on Σ such that the direct sum decompositions

$$T_y\Sigma = T_y\alpha \oplus T_y\beta, \qquad T_z\Sigma = T_z\alpha \oplus T_z\beta$$

are orthogonal and α intersects small neighborhoods of y and z in geodesic arcs. Choose a diffeomorphism $\gamma : [0,1] \to B_{yz}$ such that $\gamma(0) = y$ and $\gamma(1) = z$ and let $\zeta(t) \in T_{\gamma(t)}\Sigma$ be a unit normal vector field pointing to the right. Then there are orientation preserving embeddings $\varphi, \psi : \mathbb{D} \to H$ such that

$$\varphi(\mathbb{D}) = ([0,1] + \mathbf{i}[0,1]) \cup (\mathbf{i} + N), \qquad \varphi(\mathbb{D} \cap \mathbb{R}) = [0,1],$$
$$\psi(\mathbb{D}) = 1 + \mathbf{i} - \mathbf{i}N, \qquad \psi(\mathbb{D} \cap S^1) = 1 + \mathbf{i}[0,1],$$

and

$$u \circ \varphi^{-1}(1 + s + \mathbf{i}t) = \exp_{\gamma(t)}(s\zeta(t))$$

for $0 \le t \le 1$ and small $s \le 0$, and

$$v \circ \psi^{-1}(1 + s + \mathbf{i}t) = \exp_{\gamma(t)}(s\zeta(t))$$

for $0 \le t \le 1$ and small $s \ge 0$. The function $w : H \to \Sigma$, defined by

$$w(z) := \begin{cases} u \circ \varphi^{-1}(z), & \text{if } z \in \varphi(\mathbb{D}), \\ v \circ \psi^{-1}(z), & \text{if } z \in \psi(\mathbb{D}), \end{cases}$$

is a smooth (α, β)-heart of type (ac) from x to z that is compatible with h.

We prove the uniqueness of $[w]$. Suppose that $w' : H \to \Sigma$ is another smooth (α, β)-heart of type (ac) that is compatible with h. Let $\varphi' : \mathbb{D} \to H$ and $\psi' : \mathbb{D} \to H$ be embeddings that satisfy (47) and (48) and suppose that w' is given by (49) with φ and ψ replaced by φ' and ψ'. Then

$$\begin{aligned} w' \circ \varphi' \circ \varphi^{-1}(1 + \mathbf{i}t) &= u \circ \varphi^{-1}(1 + \mathbf{i}t) \\ &= v \circ \psi^{-1}(1 + \mathbf{i}t) \\ &= w' \circ \psi' \circ \psi^{-1}(1 + \mathbf{i}t) \end{aligned}$$

for $0 \le t \le 1$ and, by (47) and (50),

$$\varphi' \circ \varphi^{-1}(1 + \mathbf{i}) = 1 + \mathbf{i} = \psi' \circ \psi^{-1}(1 + \mathbf{i}).$$

Since w' is an immersion it follows that

$$\varphi' \circ \varphi^{-1}(1+\mathbf{i}t) = \psi' \circ \psi^{-1}(1+\mathbf{i}t) \tag{53}$$

for $0 \le t \le 1$. Consider the map $\chi : H \to H$ given by

$$\chi(\zeta) := \left\{ \begin{array}{ll} \varphi' \circ \varphi^{-1}(\zeta), & \text{for } \zeta \in \varphi(\mathbb{D}), \\ \psi' \circ \psi^{-1}(\zeta), & \text{for } \zeta \in \psi(\mathbb{D}). \end{array} \right.$$

By (53), this map is well defined. Since $H = \varphi'(\mathbb{D}) \cup \psi'(\mathbb{D})$, the map χ is surjective. We prove that χ is injective. Let $\zeta, \zeta' \in H$ such that $\chi(\zeta) = \chi(\zeta')$. If $\zeta, \zeta' \in \varphi(\mathbb{D})$ or $\zeta, \zeta' \in \psi(\mathbb{D})$ then it is obvious that $\zeta = \zeta'$. Hence assume $\zeta \in \varphi(\mathbb{D})$ and $\zeta' \in \psi(\mathbb{D})$. Then $\varphi' \circ \varphi^{-1}(\zeta) = \psi' \circ \psi^{-1}(\zeta')$ and hence, by (51),

$$\psi' \circ \psi^{-1}(\zeta') \in \psi'(\mathbb{D} \cap S^1).$$

Hence $\zeta' \in \psi(\mathbb{D} \cap S^1) \subset \varphi(\mathbb{D})$, so ζ and ζ' are both contained in $\varphi(\mathbb{D})$, and it follows that $\zeta = \zeta'$. Thus we have proved that $\chi : H \to H$ is a homeomorphism. Since $w' = w \circ \chi$ it follows that χ is a diffeomorphism. This proves (i) in the case (a). The case (c) follows by reversing the orientation of Σ and replacing u, v, w by

$$u' = u \circ \rho, \qquad v' = v \circ \rho, \qquad w'(\zeta) = w(\mathbf{i}\bar{\zeta}), \qquad \rho(\zeta) = \frac{\mathbf{i} + \bar{\zeta}}{1 + \mathbf{i}\bar{\zeta}}.$$

Thus $\rho : \mathbb{D} \to \mathbb{D}$ is an orientation reversing diffeomorphism with fixed points ± 1 that interchanges $\mathbb{D} \cap \mathbb{R}$ and $\mathbb{D} \cap S^1$. The map $H \to H : \zeta \mapsto \mathbf{i}\bar{\zeta}$ is an orientation reversing diffeomorphism with fixed points 0 and $1 + \mathbf{i}$ that interchanges $\partial_0 H$ and $\partial_1 H$. This proves (i).

We prove (ii). Let $w : H \to \Sigma$ be a smooth (α, β)-heart of type (ac) and denote

$$A_{xz} := w(\partial_0 H), \qquad B_{xz} := w(\partial_1 H).$$

Let $\gamma \subset w^{-1}(\beta)$ be the arc that starts at $1+\mathbf{i}$ and points into the interior of H. Let $\eta \in \partial H$ denote the second endpoint of γ. Since β has no self-intersections we have $y := w(\eta) \in A_{xz}$. The arc γ divides H into two components, each of which is diffeomorphic to $\mathbb{D}$. (See Figure 25.) The component which contains 0 gives rise to a smooth (α, β)-lune u from x to y and the other component gives rise to a smooth (α, β)-lune v from y to z. Let

$$\partial \Lambda_u =: (x, y, A_{xy}, B_{xy}), \qquad \partial \Lambda_v =: (y, z, A_{yz}, B_{yz}).$$

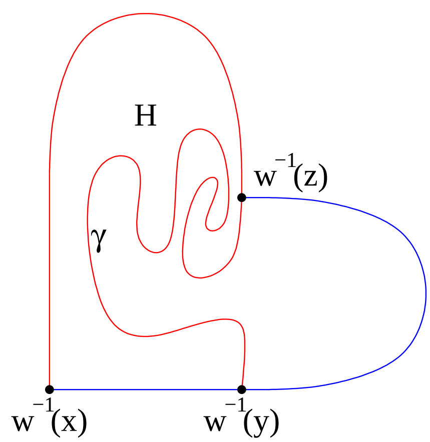

Figure 25: Breaking a heart.

Then

$$B_{xy} = B_{xz} \cup B_{yz}, \qquad B_{yz} = w(\gamma).$$

By Theorem 6.5, B_{xy} is an arc. Hence $B_{yz} \subsetneq B_{xy}$, and hence, by Proposition 10.2, the broken (α, β)-heart $h = (u, y, v)$ from x to z satisfies (a). It is obviously compatible with w. A similar argument, using the arc $\gamma' \subset w^{-1}(\alpha)$ that starts at $1+\mathbf{i}$ and points into the interior of H, proves the existence of a broken (α, β)-heart $h' \in \mathcal{H}(x, z)$ that satisfies (c) and is compatible with w. If $\widetilde{h} = (\widetilde{u}, \widetilde{y}, \widetilde{v})$ is any other broken (α, β)-heart of type (a) that is compatible with w, then it follows from uniqueness in part (i) that $w^{-1}(\widetilde{y}) = \eta$ is the endpoint of γ, hence $\widetilde{y} = y$, and hence, by Proposition 8.5, $\widetilde{h}$ is equivalent to h. This proves (ii) and Proposition 10.5. □

Proof of Theorem 9.1. The square of the boundary operator is given by

$$\partial\partial x = \sum_{z \in \alpha \cap \beta} n_H(x, z) z,$$

where

$$n_H(x, z) := \sum_{y \in \alpha \cap \beta} n(x, y) n(y, z) = \#\mathcal{H}(x, z).$$

By Proposition 10.5, and the analogous result for smooth (α, β)-hearts of type (bd), there is an involution $\tau : \mathcal{H}(x, z) \to \mathcal{H}(x, z)$ without fixed points. Hence $n_H(x, z)$ is even for all x and z and hence $\partial \circ \partial = 0$. This proves Theorem 9.1. □

11 Invariance under Isotopy

Proposition 11.1. *Let $x, y, x', y' \in \alpha \cap \beta$ be distinct intersection points such that*

$$n(x, y) = n(x', y) = n(x, y') = 1.$$

Let $u : \mathbb{D} \to \Sigma$ be a smooth (α, β)-lune from x to y and assume that the boundary $\partial\Lambda_u =: (x, y, A, B)$ of its (α, β)-trace satisfies

$$A \cap \beta = \alpha \cap B = \{x, y\}.$$

Then there is no smooth (α, β)-lune from x' to y', i.e.

$$n(x', y') = 0.$$

Moreover, extending the arc from x to y (in either α or β) beyond y, we encounter x' before y' (see Figures 26 and 27) and the two arcs $A' \subset \alpha$ and $B' \subset \beta$ from x' to y' that pass through x and y form an (α, β)-trace that satisfies the arc condition.

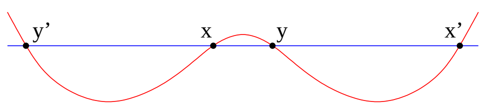

Figure 26: No lune from x' to y'.

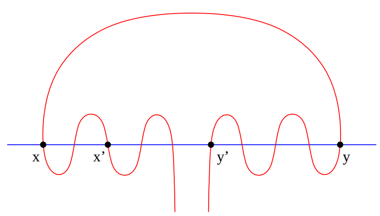

Figure 27: Lunes from x or x' to y or y'.

Proof. The proof has three steps.

Step 1. *There exist (α, β)-traces $\Lambda_{x'y} = (x', y, \mathrm{w}_{x'y})$, $\Lambda_{yx} = (y, x, \mathrm{w}_{yx})$, and $\Lambda_{xy'} = (x, y', \mathrm{w}_{xy'})$ with Viterbo–Maslov indices*

$$\mu(\Lambda_{x'y}) = 1, \qquad \mu(\Lambda_{yx}) = -1, \qquad \mu(\Lambda_{xy'}) = 1$$

such that

$$\mathrm{w}_{x'y} \geq 0, \qquad \mathrm{w}_{yx} \leq 0, \qquad \mathrm{w}_{x,y'} \geq 0.$$

By hypothesis, there exist smooth (α,β)-lunes from x' to y, from x to y, and from x to y'. By Theorem 6.7 this implies the existence of combinatorial (α,β)-lunes $\Lambda_{x'y}=(x',y,\mathrm{w}_{x'y})$, $\Lambda_{xy}=(x,y,\mathrm{w}_{xy})$, and $\Lambda_{xy'}=(x,y',\mathrm{w}_{xy'})$. To prove Step 1, reverse the direction of Λ_{xy} to obtain the required (α,β)-trace $\Lambda_{yx}=(y,x,\mathrm{w}_{yx})$ with $\mathrm{w}_{yx}:=-\mathrm{w}_{xy}$.

Step 2. *Let $\Lambda_{x'y}$, Λ_{yx}, $\Lambda_{xy'}$ be as in Step 1. Then*

$$\begin{aligned} m_{x'}(\Lambda_{yx}) = m_{y'}(\Lambda_{yx}) &= 0,\\ m_{x}(\Lambda_{x'y}) = m_{y}(\Lambda_{xy'}) &= 0,\\ m_{y'}(\Lambda_{x'y}) = m_{x'}(\Lambda_{xy'}) &= 0. \end{aligned} \tag{54}$$

By hypothesis, the combinatorial (α,β)-lune $\Lambda_{xy}=\Lambda_u$ has the boundary $\partial\Lambda_{xy}=(x,y,A,B)$ with $A\cap\beta=\alpha\cap B=\{x,y\}$. Hence $\mathrm{w}_{yx}=-\mathrm{w}_{xy}$ vanishes near every intersection point of α and β other than x and y. This proves the first equation in (54). By (38), the (α,β)-trace $\Lambda_{x'x}:=\Lambda_{x'y}\#\Lambda_{yx}$ has Viterbo–Maslov index index zero. Hence, by the trace formula (37),

$$\begin{aligned} 0 &= m_{x'}(\Lambda_{x'x})+m_x(\Lambda_{x'x})\\ &= m_{x'}(\Lambda_{x'y})+m_{x'}(\Lambda_{yx})+m_x(\Lambda_{x'y})+m_x(\Lambda_{yx})\\ &= m_x(\Lambda_{x'y}). \end{aligned}$$

Here the last equation follows from the fact that $m_{x'}(\Lambda_{yx})=0$, $m_{x'}(\Lambda_{x'y})=1$, and $m_x(\Lambda_{yx})=-1$. The equation $m_y(\Lambda_{xy'})=0$ is proved by an analogous argument, using the fact that $\Lambda_{yy'}:=\Lambda_{yx}\#\Lambda_{xy'}$ has Viterbo–Maslov index zero. This proves the second equation in (54). To prove the last equation in (54) we observe that the catenation

$$\Lambda_{x'y'}:=\Lambda_{x'y}\#\Lambda_{yx}\#\Lambda_{xy'}=(x',y',\mathrm{w}_{x'y}+\mathrm{w}_{yx}+\mathrm{w}_{xy'}) \tag{55}$$

has Viterbo–Maslov index one. Hence, by the trace formula (37),

$$\begin{aligned} 2 &= m_{x'}(\Lambda_{x'y'})+m_{y'}(\Lambda_{x'y'})\\ &= m_{x'}(\Lambda_{x'y})+m_{x'}(\Lambda_{yx})+m_{x'}(\Lambda_{xy'})\\ &\quad +m_{y'}(\Lambda_{x'y})+m_{y'}(\Lambda_{yx})+m_{y'}(\Lambda_{xy'})\\ &= 2+m_{x'}(\Lambda_{xy'})+m_{y'}(\Lambda_{x'y}). \end{aligned}$$

Here the last equation follows from the first equation in (54) and the fact that $m_{x'}(\Lambda_{x'y})=m_{y'}(\Lambda_{xy'})=1$. Since the numbers $m_{x'}(\Lambda_{xy'})$ and $m_{y'}(\Lambda_{x'y})$ are nonnegative, this proves the last equation in (54). This proves Step 2.

Step 3. *We prove the Proposition.*

Let $\Lambda_{x'y}$, Λ_{yx}, $\Lambda_{xy'}$ be as in Step 1 and denote

$$\begin{aligned}\partial\Lambda_{x'y} &=: (x', y, A_{x'y}, B_{x'y}),\\ \partial\Lambda_{yx} &=: (y, x, A_{yx}, B_{yx}),\\ \partial\Lambda_{xy'} &=: (x, y', A_{xy'}, B_{xy'}).\end{aligned}$$

By Step 2 we have

$$x, y' \notin A_{x'y} \cup B_{x'y}, \qquad x', y' \notin A_{yx} \cup B_{yx}, \qquad x', y \notin A_{xy'} \cup B_{xy'}. \tag{56}$$

In particular, the arc in α or β from y to x' contains neither x nor y'. Hence it is the extension of the arc from x to y and we encounter x' before y' as claimed. It follows also from (56) that the catenation $\Lambda_{x'y'}$ in (55) satisfies the arc condition and has boundary arcs

$$A_{x'y'} := A_{x'y} \cup A_{yx} \cup A_{xy'}, \qquad B_{x'y'} := B_{x'y} \cup B_{yx} \cup B_{xy'}.$$

Thus $x \in A_{x'y'}$ and it follows from Step 2 that

$$m_x(\Lambda_{x'y'}) = m_x(\Lambda_{x'y}) + m_x(\Lambda_{yx}) + m_x(\Lambda_{xy'}) = 0.$$

This shows that the function $\mathrm{w}_{x'y'}$ is not everywhere nonnegative, and hence $\Lambda_{x'y'}$ is not a combinatorial (α, β)-lune. By Proposition 8.5 there is no other (α, β)-trace with endpoints x', y' that satisfies the arc condition. Hence $n(x', y') = 0$. This proves Proposition 11.1 □

Proof of Theorem 9.2. By composing with a suitable ambient isotopy assume without loss of generality that $\alpha = \alpha'$. Furthermore assume the isotopy $\{\beta_t\}_{0\le t\le 1}$ with $\beta_0 = \beta$ and $\beta_1 = \beta'$ is generic in the following sense. There exists a finite sequence of pairs $(t_i, z_i) \in [0, 1] \times \Sigma$ such that

$$0 < t_1 < t_2 < \cdots < t_m < 1,$$

$\alpha \pitchfork_z \beta_t$ unless $(t, z) = (t_i, z_i)$ for some i, and for each i there exists a coordinate chart $U_i \to \mathbb{R}^2 : z \mapsto (\xi, \eta)$ at z_i such that

$$\alpha \cap U_i = \{\eta = 0\}, \qquad \beta_t \cap U_i = \left\{\eta = -\xi^2 \pm (t - t_i)\right\} \tag{57}$$

for t near t_i. It is enough to consider two cases.

Case 1 is $m = 0$. In this case there exists an ambient isotopy φ_t such that $\varphi_t(\alpha) = \alpha$ and $\varphi_t(\beta) = \beta_t$. It follows that the map $\mathrm{CF}(\alpha, \beta) \to \mathrm{CF}(\alpha, \beta')$ induced by $\varphi_1 : \alpha \cap \beta \to \alpha \cap \beta'$ is a chain isomorphism that identifies the boundary maps.

In Case 2 we have $m = 1$, the isotopy is supported near U_1, and (57) holds with the minus sign. Thus there are two intersection points in U_1 for $t < t_1$, no intersection points in U_1 for $t > t_1$, and all other intersection points of α and β_t are independent of t. Denote by $x, y \in \alpha \cap \beta$ the intersection points that cancel at time $t = t_1$ and choose the ordering such that

$$n(x, y) = 1. \tag{58}$$

Then $\alpha \cap \beta' = (\alpha \cap \beta) \setminus \{x, y\}$. We prove in seven steps that

$$n'(x', y') = n(x', y') + n(x', y)n(x, y') \tag{59}$$

for $x', y' \in \alpha \cap \beta'$, where $n(x', y')$ denotes the number of (α, β)-lunes from x' to y' and $n'(x', y')$ denotes the number of (α, β')-lunes from x' to y'.

Step 1. *If there is no (α, β)-trace from x' to y' that satisfies the arc condition then (59) holds.*

In this case there is no (α, β_t)-trace from x' to y' that satisfies the arc condition for any t. Hence it follows from Theorem 6.5 that $n(x', y') = n'(x', y') = 0$ and it follows from Proposition 11.1 that $n(x', y)n(x, y') = 0$. Hence (59) holds in this case.

Standing Assumptions, Part 1. *From now on we assume that there is an (α, β)-trace from x' to y' that satisfies the arc condition. Then there is an (α, β_t)-trace from x' to y' that satisfies the arc condition for every t. By Proposition 8.5, this (α, β_t)-trace is uniquely determined by x' and y'. We denote it by $\Lambda_{x'y'}(t)$ and its boundary by*

$$\partial\Lambda_{x'y'}(t) =: (x', y', A_{x'y'}, B_{x'y'}(t)).$$

Choose a universal covering $\pi : \mathbb{C} \to \Sigma$ so that $\widetilde{\alpha} = \mathbb{R}$ is a lift of α, the map $\widetilde{z} \mapsto \widetilde{z} + 1$ is a covering transformation, and π maps the interval $[0, 1)$ bijectively onto α. Let

$$\widetilde{\Lambda}_{x'y'}(t) = (\widetilde{x}', \widetilde{y}', \mathrm{w}_{\widetilde{\Lambda}_{x'y'}(t)})$$

be a continuous family of lifts of the (α, β_t)-traces to the universal cover $\pi : \mathbb{C} \to \Sigma$ with boundaries $\partial\widetilde{\Lambda}_{x'y'}(t) = (\widetilde{x}', \widetilde{y}', \widetilde{A}_{x'y'}, \widetilde{B}_{x'y'}(t))$.

Step 2. *The lift $\widetilde{\Lambda}_{x'y'}(t)$ satisfies condition (II) (that the intersection index of $\widetilde{A}_{x'y'}$ and $\widetilde{B}_{x'y'}(t)$ at $\widetilde{x}'$ is $+1$ and at $\widetilde{y}'$ is -1) and condition (III) (that the winding numbers of $\widetilde{A}_{x'y'} - \widetilde{B}_{x'y'}(t)$ are zero or one near $\widetilde{x}'$ and $\widetilde{y}'$) either for all values of t or for no value of t.*

The intersection indices of $\widetilde{A}_{x'y'}$ and $\widetilde{B}_{x'y'}(t)$ at $\widetilde{x}'$ and $\widetilde{y}'$, and the winding numbers of $\widetilde{A}_{x'y'} - \widetilde{B}_{x'y'}(t)$ near $\widetilde{x}'$ and $\widetilde{y}'$, are obviously independent of t.

Step 3. *If one of the arcs $A_{x'y'}(t)$ and $B_{x'y'}(t)$ does not pass through U_1 then (59) holds.*

In this case the winding numbers do not change sign as t varies and hence, by Step 2, $\nu(\Lambda_{x'y'}(t))$ is independent of t. Hence

$$n(x', y') = n'(x', y').$$

To prove equation (59) in this case, we must show that one of the numbers $n(x', y)$ or $n(x, y')$ vanishes. Suppose otherwise that

$$n(x', y) = n(x, y') = 1.$$

Since $n(x, y) = 1$, by equation (58), it follows from Proposition 11.1 that the two arcs from x' to y' that pass through U_1 form an (α, β)-trace that satisfies the arc condition. Hence there are two (α, β)-traces from x' to y' that satisfy the arc condition, which is impossible by Proposition 8.5. This contradiction proves Step 3.

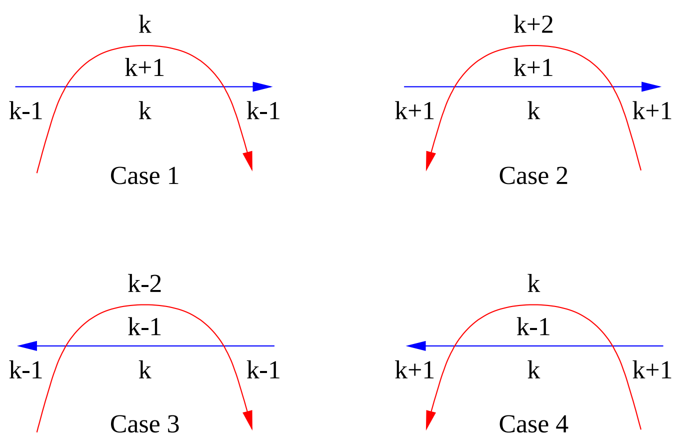

Figure 28: The winding numbers of $\widetilde{\Lambda}_{x'y'}$ in $\widetilde{U}_1$ for $t = 0$.

Standing Assumptions, Part 2. *From now on we assume that $A_{x'y'}$ and $B_{x'y'}(t)$ both pass through U_1. Denote by $\widetilde{x}$ and $\widetilde{y}$ the unique lifts of x and y, respectively, in $\widetilde{A}_{x'y'} \cap \widetilde{B}_{x'y'}(0)$.*

Under this assumption the winding number of $\widetilde{\Lambda}_{x'y'}(t)$ only changes in the area enclosed by the two arcs in the lift $\widetilde{U}_1$ of U_1 which contains $\widetilde{x}$ and $\widetilde{y}$. There are four cases, depending on the orientations of the two arcs from x' to y'. (See Figure 28.) The next step deals with three of these cases.

Step 4. *Assume that the orientation of α from x' to y' does not agree with one of the orientations from x' to y or from x to y', or else that this holds for β (i.e. that one of the Cases 1,2,3 holds in Figure 28). Then (59) holds.*

In Cases 1,2,3 the pattern of winding numbers shows (for any value of k) that $\mathrm{w}_{\widetilde{\Lambda}_{x'y'}(t)}$ is either nonnegative for all values of t or is somewhere negative for all values of t. Hence, by Step 2, $\nu(\Lambda_{x'y'}(t))$ is independent of t, and hence $n(x', y') = n'(x', y')$. Moreover, by Proposition 11.1, we have that in these cases $n(x, y')n(x', y) = 0$. Hence (59) holds in the Cases 1,2,3.

Standing Assumptions, Part 3. *From now on we assume that Case 4 holds in Figure 28, i.e. that the orientations of α and β from x' to y' agree with the orientations from x' to y and with the orientations from x to y'.*

Step 5. *Assume Case 4 and $n(x', y) = n(x, y') = 1$. Then (59) holds.*

By Proposition 11.1, we have $n(x', y') = 0$. We must prove that $n'(x', y') = 1$. Let $\Lambda_{x'y}$ and $\Lambda_{xy'}$ be the (α, β)-traces from x' to y, respectively from x to y', that satisfy the arc condition and denote their lifts by $\widetilde{\Lambda}_{x'y}$ and $\widetilde{\Lambda}_{xy'}$. By (55),

$$\mathrm{w}_{\widetilde{\Lambda}_{x'y'}(0)}(\widetilde{z}) = \mathrm{w}_{\widetilde{\Lambda}_{x'y}}(\widetilde{z}) + \mathrm{w}_{\widetilde{\Lambda}_{xy'}}(\widetilde{z}) \qquad \text{for} \quad \widetilde{z} \in \mathbb{C} \setminus \widetilde{U}_1. \tag{60}$$

Thus $\mathrm{w}_{\widetilde{\Lambda}_{x'y'}(0)} \geq 0$ in $\mathbb{C} \setminus \widetilde{U}_1$. Moreover, by Theorem 6.7, the lifts $\widetilde{\Lambda}_{x'y}, \widetilde{\Lambda}_{xy'}$ have winding numbers zero in the regions labelled by k and $k-1$ in Figure 28, Case 4. Hence, by (60), we have $k = 0$ and hence

$$\mathrm{w}_{\widetilde{\Lambda}_{x'y'}(t)} \geq 0 \qquad \text{for} \qquad t > t_1.$$

Thus we have proved that $\widetilde{\Lambda}_{x'y'}(t)$ satisfies (I) for $t > t_1$. Moreover, the Viterbo–Maslov index is given by

$$\mu(\widetilde{\Lambda}_{x'y'}(t)) = \mu(\widetilde{\Lambda}_{x'y}) + \mu(\widetilde{\Lambda}_{xy'}) - \mu(\widetilde{\Lambda}_{xy}) = 1.$$

Hence, by Theorem 6.7, $\widetilde{\Lambda}_{x'y'}(t)$ is a combinatorial lune for $t > t_1$ and we have $n'(x', y') = 1$.

Step 6. *Assume Case 4 and $n'(x', y') = 1$. Then (59) holds.*

The winding numbers of $\widetilde{\Lambda}_{x'y'}(1)$ are nonnegative and hence we must have $k \geq 0$ in Figure 28, Case 4. If $k > 0$ then the winding numbers of $\widetilde{\Lambda}_{x'y'}(0)$ are also nonnegative and hence, by Step 2, $n(x', y') = 1$. Hence, by (58) and Proposition 11.1, one of the numbers $n(x', y)$ and $n(x, y')$ must vanish, and hence (59) holds when $k > 0$.

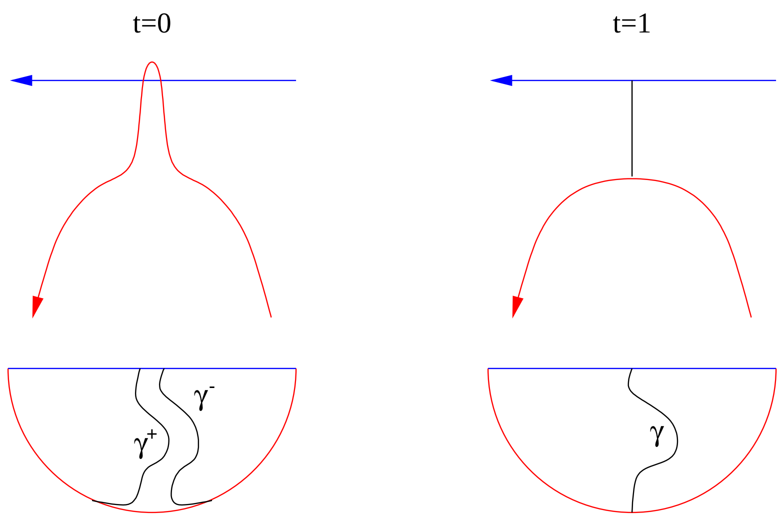

Figure 29: A lune splits.

Now assume $k = 0$. Then $n(x', y') = 0$ and we must prove that

$$n(x', y) = n(x, y') = 1.$$

To see this, we choose a smooth $(\widetilde{\alpha}, \widetilde{\beta}')$-lune

$$\widetilde{u}' : \mathbb{D} \to \mathbb{C}$$

from $\widetilde{x}'$ to $\widetilde{y}'$. Since $k = 0$ this lune has precisely one preimage in the region in $\widetilde{U}_1$ where there winding number is $k + 1 = 1 \neq 0$ for $t = 1$. Choose an embedded arc

$$\widetilde{\gamma} : [0, 1] \to \mathbb{C}$$

in $\widetilde{U}_1$ for $t = 1$ connecting the two branches of $\widetilde{\alpha}$ and $\widetilde{\beta}'$ such that $\widetilde{\gamma}$ intersects $\widetilde{\alpha}$ and $\widetilde{\beta}'$ only at the endpoints (see Figure 29). Then

$$\gamma := \widetilde{u}'^{\,-1}(\widetilde{\gamma})$$

divides $\mathbb{D}$ into two components. Deform the curve $\widetilde{\beta}'$ along $\widetilde{\gamma}$ to a curve $\widetilde{\beta}''$ that intersects $\widetilde{\alpha}$ in two points $\widetilde{x}'', \widetilde{y}'' \in \widetilde{U}_1$. The preimage of $\widetilde{\beta}''$ under $\widetilde{u}'$ contains two arcs γ^- and γ^+, parallel to γ, in the two components of $\mathbb{D}$. This results in two half-discs contained in $\mathbb{D}$, and the restriction of $\widetilde{u}$ to these two half discs gives rise to two $(\widetilde{\alpha}, \widetilde{\beta}'')$-lunes, one from $\widetilde{x}'$ to $\widetilde{y}''$ and one from $\widetilde{x}''$ to $\widetilde{y}'$ (see Figure 29). Moreover $\widetilde{\beta}''$ descends to an embedded loop in Σ that is isotopic to β through loops that are transverse to α. Hence $n(x', y) = n(x, y') = 1$, as claimed.

Step 7. *Assume Case 4 and $n'(x', y') = n(x', y)n(x, y') = 0$. Then (59) holds.*

We must prove that $n(x', y') = 0$. By Step 2, conditions (II) and (III) on $\widetilde{\Lambda}_{x'y'}(t)$ are independent of t and so we have that $n(x', y') = 0$ whenever these conditions are not satisfied. Hence assume that $\widetilde{\Lambda}_{x'y'}(t)$ satisfies (II) and (III) for every t. If $k > 0$ in Figure 28 then condition (I) also holds for every t and hence $n(x', y') = n'(x', y') = 1$, a contradiction. If $k \leq 0$ in Figure 28 then $\widetilde{\Lambda}_{x'y'}(0)$ does not satisfy (I) and hence $n(x', y') = 0$, as claimed.

Thus we have established (59). Hence, by Lemma C.1, $\mathrm{HF}(\alpha, \beta)$ is isomorphic to $\mathrm{HF}(\alpha, \beta')$. This proves Theorem 9.2. $\square$

12 Lunes and Holomorphic Strips

We assume throughout that Σ and $\alpha, \beta \subset \Sigma$ satisfy hypothesis (H). We also fix a complex structure J on Σ. A **holomorphic** (α, β)**-strip** is a holomorphic map $v : \mathbb{S} \to \Sigma$ of finite energy such that

$$v(\mathbb{R}) \subset \alpha, \qquad v(\mathbb{R} + \mathbf{i}) \subset \beta. \tag{61}$$

It follows [32, Theorem A] that the limits

$$x = \lim_{s \to -\infty} v(s + \mathbf{i}t), \qquad y = \lim_{s \to +\infty} v(s + \mathbf{i}t) \tag{62}$$

exist; the convergence is exponential and uniform in t; moreover $\partial_s v$ and all its derivatives converge exponentially to zero as s tends to $\pm\infty$. Call two holomorphic strips **equivalent** if they differ by a time shift. Every holomorphic strip v has a **Viterbo–Maslov index** $\mu(v)$, defined as follows. Trivialize the complex line bundle $v^*T\Sigma \to \mathbb{S}$ such that the trivialization converges to a frame of $T_x\Sigma$ as s tends to $-\infty$ and to a frame of $T_y\Sigma$ as s tends to $+\infty$ (with convergence uniform in t). Then $s \mapsto T_{v(s,0)}\alpha$ and $s \mapsto T_{v(s,1)}\beta$ are Lagrangian paths and their relative Maslov index is $\mu(v)$ (see [39] and [30]).

At this point it is convenient to introduce the notation

$$\mathcal{M}^{\mathrm{Floer}}(x,y;J) := \frac{\left\{ v : \mathbb{S} \to \Sigma \,\middle|\, \begin{array}{l} v \text{ is a holomorphic } (\alpha,\beta)\text{-strip} \\ \text{from } x \text{ to } y \text{ with } \mu(v) = 1 \end{array} \right\}}{\text{time shift}}$$

for the moduli space of index one holomorphic strips from x to y up to time shift. This moduli space depends on the choice of a complex structure J on Σ. We also introduce the notation

$$\mathcal{M}^{\mathrm{comb}}(x,y) := \frac{\left\{ u : \mathbb{D} \to \Sigma \,\middle|\, \begin{array}{l} u \text{ is a smooth } (\alpha,\beta)\text{-lune} \\ \text{from } x \text{ to } y \end{array} \right\}}{\text{isotopy}}$$

for the moduli space of (equivalence classes of) smooth (α,β)-lunes. This space is independent of the choice of J. We show that there is a bijection between these moduli spaces for every pair $x, y \in \alpha \cap \beta$.

Given a smooth (α,β)-lune $u : \mathbb{D} \to \Sigma$, the Riemann mapping theorem gives a unique homeomorphism $\varphi_u : \mathbb{D} \to \mathbb{D}$ such that the restriction of φ_u to $\mathbb{D} \setminus \{\pm 1\}$ is a diffeomorphism and

$$\varphi_u(-1) = -1, \qquad \varphi_u(0) = 0, \qquad \varphi_u(1) = 1, \qquad \varphi_u^* u^* J = \mathbf{i}. \tag{63}$$

Let $g : \mathbb{S} \to \mathbb{D} \setminus \{\pm 1\}$ be the holomorphic diffeomorphism given by

$$g(s + \mathbf{i}t) := \frac{e^{(s+\mathbf{i}t)\pi/2} - 1}{e^{(s+\mathbf{i}t)\pi/2} + 1}. \tag{64}$$

Then, for every (α,β)-lune u, the composition $v := u \circ \varphi_u \circ g$ is a holomorphic (α,β)-strip.

Theorem 12.1. *Assume (H), let $x, y \in \alpha \cap \beta$, and choose a complex structure J on Σ. Then the map $u \mapsto u \circ \varphi_u \circ g$ induces a bijection*

$$\mathcal{M}^{\mathrm{comb}}(x,y) \to \mathcal{M}^{\mathrm{Floer}}(x,y;J) : [u] \mapsto [u \circ \varphi_u \circ g] \tag{65}$$

between the corresponding moduli spaces.

Proof. See page 97 below. □

The proof of Theorem 12.1 relies on the asymptotic analysis of holomorphic strips in [32] (see Appendx D for a summary) and on an explicit formula for the Viterbo–Maslov index. For each intersection point $x \in \alpha \cap \beta$ we denote by $\theta_x \in (0, \pi)$ the angle from $T_x\alpha$ to $T_x\beta$ with respect to our complex structure J. Thus

$$T_x\beta = (\cos(\theta_x) + \sin(\theta_x)J)\, T_x\alpha, \qquad 0 < \theta_x < \pi. \tag{66}$$

Fix a nonconstant holomorphic (α, β)-strip $v : \mathbb{S} \to \Sigma$ from x to y. Choose a holomorphic coordinate chart $\psi_y : U_y \to \mathbb{C}$ on an open neighborhood $U_y \subset \Sigma$ of y such that $\psi_y(y) = 0$. By [32, Theorem C] (see also Corollary D.2) there is a complex number c_y and a integer $\nu_y(v) \geq 1$ such that

$$\psi_y(v(s + \mathbf{i}t)) = c_y e^{-(\nu_y(v)\pi - \theta_y)(s+\mathbf{i}t)} + O(e^{-(\nu_y(v)\pi - \theta_y + \delta)s}) \tag{67}$$

for some $\delta > 0$ and all $s > 0$ sufficiently close to $+\infty$. The complex number c_y belongs to the tangent space $T_0(\psi_y(\alpha \cap U_y))$ and the integer $\nu_y(v)$ is independent of the choice of the coordinate chart.

Now let us interchange α and β as well as x and y, and replace v by the (β, α)-holomorphic strip

$$s + \mathbf{i}t \mapsto v(-s + \mathbf{i}(1 - t))$$

from y to x. Choose a holomorphic coordinate chart $\psi_x : U_x \to \mathbb{C}$ on an open neighborhood $U_x \subset \Sigma$ of x such that $\psi_x(x) = 0$. Using [32, Theorem C] again we find that there is a complex number c_x and an integer $\nu_x(v) \geq 0$ such that

$$\psi_x(v(s + \mathbf{i}t)) = c_x e^{(\nu_x(v)\pi + \theta_x)(s+\mathbf{i}t)} + O(e^{(\nu_x(v)\pi + \theta_x + \delta)s}) \tag{68}$$

for some $\delta > 0$ and all $s < 0$ sufficiently close to $-\infty$. As before, the complex number c_x belongs to the tangent space $T_0(\psi_x(\alpha \cap U_x))$ and the integer $\nu_x(v)$ is independent of the choice of the coordinate chart.

Denote the set of critical points of v by

$$C_v := \{z \in \mathbb{S} \,|\, dv(z) = 0\}.$$

It follows from Corollary D.2 (i) that this is a finite set. For $z \in C_v$ denote by $\nu_z(v) \in \mathbb{N}$ the order to which dv vanishes at z. Thus the first nonzero term in the Taylor expansion of v at z (in a local holomorphic coordinate chart on Σ centered at $v(z)$) has order $\nu_z(v) + 1$.

Theorem 12.2. *Assume (H) and choose a complex structure J on Σ. Let $x, y \in \alpha \cap \beta$ and $v : \mathbb{S} \to \Sigma$ be a nonconstant holomorphic (α, β)-strip from x to y. Then the linearized operator D_v associated to this strip in Floer theory is surjective. Moreover, the Viterbo–Maslov index of v is equal to the Fredholm index of D_v and is given by the* **index formula**

$$\mu(v) = \nu_x(v) + \nu_y(v) + \sum_{z \in C_v \cap \partial \mathbb{S}} \nu_z(v) + 2 \sum_{z \in C_v \cap \mathrm{int}(\mathbb{S})} \nu_z(v). \tag{69}$$

The right hand side in equation (69) is positive because all summands are nonnegative and $\nu_y(v) \geq 1$.

Proof. See page 97 below. $\square$

The surjectivity statement in Theorem 12.2 has been observed by many authors. A proof for holomorphic polygons is contained in Seidel's book [34].

We will need a more general index formula (equation (76) below) which we explain next. Choose a Riemannian metric on Σ that is compatible with the complex structure J. This metric induces a Hermitian structure on the pullback tangent bundle $v^*T\Sigma \to \mathbb{S}$ and the Hilbert spaces $W^{1,2}(\mathbb{S}, v^*T\Sigma)$ and $L^2(\mathbb{S}, v^*T\Sigma)$ are understood with respect to this induced structure. These Hilbert spaces are independent of the choice of the metric on Σ, only their inner products depend on this choice. The linearized operator

$$D_v : W^{1,2}_{\mathrm{BC}}(\mathbb{S}, v^*T\Sigma) \to L^2(\mathbb{S}, v^*T\Sigma)$$

with

$$W^{1,2}_{\mathrm{BC}}(\mathbb{S}, v^*T\Sigma) := \left\{ \hat{v} \in W^{1,2}(\mathbb{S}, v^*T\Sigma) \,\middle|\, \begin{array}{l} \hat{v}(s,0) \in T_{v(s,0)}\alpha \ \forall s \in \mathbb{R} \\ \hat{v}(s,1) \in T_{v(s,1)}\beta \ \forall s \in \mathbb{R} \end{array} \right\}$$

is given by

$$D_v \hat{v} = \nabla_s \hat{v} + J \nabla_t \hat{v}$$

for $\hat{v} \in W^{1,2}_{\mathrm{BC}}(\mathbb{S}, v^*T\Sigma)$, where ∇ denotes the Levi-Civita connection. Here we use the fact that $\nabla J = 0$ because J is integrable. We remark that, first, this is a Fredholm operator for every smooth map $v : \mathbb{S} \to \Sigma$ satisfying (61) and (62) (where the convergence is exponential and uniformly in t, and $\partial_s v$, $\nabla_s \partial_s v$, $\nabla_s \partial_t v$ converge exponentially to zero as s tends to $\pm\infty$). Second, the definition of the Viterbo–Maslov index $\mu(v)$ extends to this setting and it is equal to the Fredholm index of D_v (see [31]) Third, the operator D_v is independent of the choice of the Riemannian metric whenever v is an (α, β)-holomorphic strip.

Next we choose a unitary trivialization

$$\Phi(s,t) : \mathbb{C} \to T_{v(s,t)}\Sigma$$

of the pullback tangent bundle such that

$$\Phi(s,0)\mathbb{R} = T_{v(s,0)}\alpha, \qquad \Phi(s,1)\mathbb{R} = T_{v(s,1)}\beta,$$

and $\Phi(s,t) = \Psi_t(v(s,t))$ for $|s|$ sufficiently large. Here Ψ_t, $0 \le t \le 1$, is a smooth family of unitary trivializations of the tangent bundle over a neighborhood $U_x \subset \Sigma$ of x, respectively $U_y \subset \Sigma$ of y, such that $\Psi_0(z)\mathbb{R} = T_z\alpha$ for $z \in (U_x \cup U_y) \cap \alpha$ and $\Psi_1(z)\mathbb{R} = T_z\beta$ for $z \in (U_x \cup U_y) \cap \beta$. Then

$$\begin{aligned}
\mathcal{W} &:= \Phi^{-1} W^{1,2}_{\mathrm{BC}}(\mathbb{S}, v^*T\Sigma) = \left\{\xi \in W^{1,2}(\mathbb{S},\mathbb{C}) \,|\, \xi(s,0), \xi(s,1) \in \mathbb{R}\ \forall s \in \mathbb{R}\right\}, \\
\mathcal{H} &:= \Phi^{-1} L^2(\mathbb{S}, v^*T\Sigma) = L^2(\mathbb{S},\mathbb{C}).
\end{aligned}$$

The operator $D_S := \Phi^{-1} \circ D_v \circ \Phi : \mathcal{W} \to \mathcal{H}$ has the form

$$D_S\xi = \partial_s\xi + \mathbf{i}\partial_t\xi + S\xi \tag{70}$$

where the function $S : \mathbb{S} \to \mathrm{End}_{\mathbb{R}}(\mathbb{C})$ is given by

$$S(s,t) := \Phi(s,t)^{-1}\Big(\nabla_s\Phi(s,t) + J(v(s,t))\nabla_t\Phi(s,t)\Big).$$

The matrix $\Phi^{-1}\nabla_s\Phi$ is skew-symmetric and the matrix $\Phi^{-1}J(v)\nabla_t\Phi$ is symmetric. Moreover, it follows from our hypotheses on v and the trivialization that S converges exponentially and $\Phi^{-1}\nabla_s\Phi$ as well as $\partial_s S$ converge exponentially to zero as as s tends to $\pm\infty$. The limits of S are the symmetric matrix functions

$$\begin{aligned}
S_x(t) &:= \lim_{s\to-\infty} S(s,t) = \Psi_t(x)^{-1}J(x)\partial_t\Psi_t(x), \\
S_y(t) &:= \lim_{s\to+\infty} S(s,t) = \Psi_t(y)^{-1}J(y)\partial_t\Psi_t(y).
\end{aligned} \tag{71}$$

Thus there exist positive constants c and ε such that

$$\begin{aligned}
|S(s,t) - S_x(t)| + |\partial_s S(s,t)| &\le ce^{\varepsilon s}, \\
|S(s,t) - S_y(t)| + |\partial_s S(s,t)| &\le ce^{-\varepsilon s}
\end{aligned} \tag{72}$$

for every $s \in \mathbb{R}$. This shows that the operator (70) satisfies the hypotheses of [32, Lemma 3.6]. This lemma asserts the following. Let $\xi \in \mathcal{W}$ be a nonzero function in the kernel of D_S:

$$\xi \in \mathcal{W}, \qquad D_S\xi = \partial_s\xi + \mathbf{i}\partial_t\xi + S\xi = 0, \qquad \xi \ne 0.$$

Then there exist nonzero functions $\xi_x, \xi_y : [0,1] \to \mathbb{C}$ and positive real numbers $\lambda_x, \lambda_y, C, \delta$ such that

$$\begin{aligned} \mathbf{i}\dot{\xi}_x(t) + S_x(t)\xi_x(t) &= -\lambda_x \xi_x(t), \qquad \xi_x(0), \xi_x(1) \in \mathbb{R} \\ \mathbf{i}\dot{\xi}_y(t) + S_y(t)\xi_y(t) &= \lambda_y \xi_y(t), \qquad \xi_y(0), \xi_y(1) \in \mathbb{R} \end{aligned} \tag{73}$$

and

$$\begin{aligned} \left|\xi(s,t) - e^{\lambda_x s}\xi_x(t)\right| &\le Ce^{(\lambda_x+\delta)s}, \qquad s \le 0 \\ \left|\xi(s,t) - e^{-\lambda_y s}\xi_y(t)\right| &\le Ce^{-(\lambda_y+\delta)s}, \qquad s \ge 0. \end{aligned} \tag{74}$$

We prove that there exist integers $\iota(x,\xi) \ge 0$ and $\iota(y,\xi) \ge 1$ such that

$$\lambda_x = \iota(x,\xi)\pi + \theta_x, \qquad \lambda_y = \iota(y,\xi)\pi - \theta_y. \tag{75}$$

Here θ_x is chosen as above such that

$$T_x\beta = \exp(\theta_x J(x))T_x\alpha, \qquad 0 < \theta_x < \pi,$$

and the same for θ_y. To prove (75), we observe that the function

$$v_x(t) := \Psi_t(x)\xi_x(t)$$

satisfies

$$J(x)\dot{v}_x(t) = -\lambda_x v_x(t), \qquad v_x(0) \in T_x\alpha, \qquad v_x(1) \in T_x\beta.$$

Hence $v_x(t) = \exp(t\lambda_x J(x))v_x(0)$ and this proves the first equation in (75). Likewise, the function $v_y(t) := \Psi_t(y)\xi_y(t)$ satisfies $J(y)\dot{v}_y(t) = \lambda_y v_y(t)$ and $v_y(0) \in T_y\alpha$ and $v_y(1) \in T_y\beta$. Hence $v_y(t) = \exp(-t\lambda_y J(y))v_y(0)$, and this proves the second equation in (75).

Lemma 12.3. *Suppose S satisfies the asymptotic condition (72) and let $\xi \in \mathcal{W}$ be a smooth function with isolated zeros that satisfies (73), (74), and (75). Then the Fredholm index of D_S is given by the* **linear index formula**

$$\mathrm{index}(D_S) = \iota(x,\xi) + \iota(y,\xi) + \sum_{\substack{z\in\partial\mathbb{S} \\ \xi(z)=0}} \iota(z,\xi) + 2\sum_{\substack{z\in\mathrm{int}(\mathbb{S}) \\ \xi(z)=0}} \iota(z,\xi). \tag{76}$$

In the second sum $\iota(z,\xi)$ denotes the index of z as a zero of ξ. In the first sum $\iota(z,\xi)$ denotes the degree of the loop $[0,\pi] \to \mathbb{R}\mathrm{P}^1 : \theta \mapsto \xi(z+\varepsilon e^{\mathbf{i}\theta})\mathbb{R}$ when $z \in \mathbb{R}$ and of the loop $[0,\pi] \to \mathbb{R}\mathrm{P}^1 : \theta \mapsto \xi(z-\varepsilon e^{\mathbf{i}\theta})\mathbb{R}$ when $z \in \mathbb{R}+\mathbf{i}$; in both cases $\varepsilon > 0$ is chosen so small that the closed ε-neighborhood of z contains no other zeros of ξ.

Proof. Since ξ_x and ξ_y have no zeros, by (73), it follows from equation (74) that the zeros of ξ are confined to a compact subset of $\mathbb{S}$. Moreover the zeros of ξ are isolated and so the right hand side of (76) is a finite sum. Now let $\xi_0 : \mathbb{S} \to \mathbb{C}$ be the unique solution of the equation

$$\mathbf{i}\partial_t \xi_0(s,t) + S(s,t)\xi_0(s,t) = 0, \qquad \xi_0(s,0) = 1. \tag{77}$$

Then

$$\begin{aligned} \xi_{0,x}(t) &:= \lim_{s\to-\infty} \xi_0(s,t) = \Psi_t(x)^{-1}\Psi_0(x)1, \\ \xi_{0,y}(t) &:= \lim_{s\to+\infty} \xi_0(s,t) = \Psi_t(y)^{-1}\Psi_0(y)1. \end{aligned} \tag{78}$$

Thus the Lagrangian path

$$\mathbb{R} \to \mathbb{R}\mathrm{P}^1 : s \mapsto \Lambda_S(s) := \mathbb{R}\xi(s,1)$$

is asymptotic to the subspace $\Psi_1(x)^{-1}T_x\alpha$ as s tends to $-\infty$ and to the subspace $\Psi_1(y)^{-1}T_y\alpha$ as s tends to $+\infty$. These subspaces are both transverse to $\mathbb{R}$. By the spectral-flow-equals-Maslov-index theorem in [31] the Fredholm index of D_S is equal to the relative Maslov index of the pair $(\Lambda_S, \mathbb{R})$:

$$\mathrm{index}(D_S) = \mu(\Lambda_S, \mathbb{R}). \tag{79}$$

It follows from (73) and (78) that

$$\frac{\xi_{0,x}(t)}{\xi_x(t)} = \frac{e^{-\mathbf{i}\lambda_x t}}{\xi_x(0)}, \qquad \frac{\xi_{0,y}(t)}{\xi_y(t)} = \frac{e^{\mathbf{i}\lambda_y t}}{\xi_y(0)}. \tag{80}$$

Now let $U = \bigcup_{\xi(z)=0} U_z \subset \mathbb{S}$ be a union of open discs or half discs U_z of radius less than one half, centered at the zeros z of ξ, whose closures are disjoint. Consider the smooth map $\Lambda : \mathbb{S} \setminus U \to \mathbb{R}\mathrm{P}^1$ defined by

$$\Lambda(s,t) := \xi_0(s,t)\overline{\xi(s,t)}\mathbb{R}.$$

By (80) this map converges, uniformly in t, as s tends to $\pm\infty$ with limits

$$\Lambda_x(t) := \lim_{s\to-\infty} \Lambda(s,t) = e^{-\mathbf{i}\lambda_x t}\mathbb{R}, \qquad \Lambda_y(t) := \lim_{s\to+\infty} \Lambda(s,t) = e^{\mathbf{i}\lambda_y t}\mathbb{R}. \tag{81}$$

Moreover, we have

$$\begin{aligned} \Lambda(s,1) &= \xi_0(s,1)\mathbb{R} = \Lambda_S(s), \qquad && (s,1) \notin U, \\ \Lambda(s,0) &= \xi_0(s,0)\mathbb{R} = \mathbb{R}, && (s,0) \notin U. \end{aligned}$$

If $z \in \mathrm{int}(\mathbb{S})$ with $\xi(z) = 0$ then the map $\Lambda_z := \Lambda|_{\partial U_z}$ is homotopic to the map $\partial U_z \to \mathbb{R}\mathrm{P}^1 : s + \mathbf{i}t \mapsto \overline{\xi(s,t)}\mathbb{R}$. Hence it follows from the definition of the index $\iota(z,\xi)$ that its degree is

$$\deg(\Lambda_z : \partial U_z \to \mathbb{R}\mathrm{P}^1) = -2\iota(z,\xi), \qquad z \in \mathrm{int}(\mathbb{S}), \qquad \xi(z) = 0. \tag{82}$$

If $z \in \partial\mathbb{S}$ with $\xi(z) = 0$, define the map $\Lambda_z : \partial U_z \to \mathbb{R}P^1$ by

$$\Lambda_z(s,t) := \begin{cases} \xi_0(s,t)\overline{\xi(s,t)}\mathbb{R}, & \text{if } (s,t) \in \partial U_z \setminus \partial\mathbb{S}, \\ \xi_0(s,t)\mathbb{R}, & \text{if } (s,t) \in \partial U_z \cap \partial\mathbb{S}. \end{cases}$$

This map is homotopic to the map $(s,t) \mapsto \overline{\xi(s,t)}\mathbb{R}$ for $(s,t) \in \partial U_z \setminus \partial\mathbb{S}$ and $(s,t) \mapsto \mathbb{R}$ for $(s,t) \in \partial U_z \cap \partial\mathbb{S}$. Hence it follows from the definition of the index $\iota(z,\xi)$ that its degree is

$$\deg(\Lambda_z : \partial U_z \to \mathbb{R}\mathrm{P}^1) = -\iota(z,\xi), \qquad z \in \partial\mathbb{S}, \qquad \xi(z) = 0. \tag{83}$$

Abbreviate $\mathbb{S}_T := [-T,T] + \mathbf{i}[0,1]$ for $T > 0$ sufficiently large. Since the map $\Lambda : \partial(\mathbb{S}_T \setminus U) \to \mathbb{R}\mathrm{P}^1$ extends to $\mathbb{S}_T \setminus U$ its degree is zero and it is equal to the relative Maslov index of the pair of Lagrangian loops $(\Lambda|_{\partial(\mathbb{S}_T \cap U)}, \mathbb{R})$. Hence

$$\begin{aligned} 0 &= \lim_{T\to\infty} \mu(\Lambda|_{\partial(\mathbb{S}_T\setminus U)}, \mathbb{R}) \\ &= \mu(\Lambda_y, \mathbb{R}) - \mu(\Lambda_x, \mathbb{R}) - \mu(\Lambda_S, \mathbb{R}) - \sum_{\substack{z\in\mathbb{S} \\ \xi(z)=0}} \mu(\Lambda_z, \mathbb{R}) \\ &= \iota(y,\xi) + \iota(x,\xi) - \mu(\Lambda_S, \mathbb{R}) + \sum_{\substack{z\in\partial\mathbb{S} \\ \xi(z)=0}} \iota(z,\xi) + 2\sum_{\substack{z\in\mathrm{int}(\mathbb{S}) \\ \xi(z)=0}} \iota(z,\xi). \end{aligned}$$

Here the second equality follows from the additivity of the relative Maslov index for paths [30]. It also uses the fact that, for $z \in \mathbb{R} + \mathbf{i}$ with $\xi(z) = 0$, the relative Maslov index of the pair $(\Lambda_z|_{\partial U_z \cap (\mathbb{R}+\mathbf{i})}, \mathbb{R}) = (\Lambda_S|_{\partial U_z \cap (\mathbb{R}+\mathbf{i})}, \mathbb{R})$ appears with a plus sign when using the orientation of $\mathbb{R} + \mathbf{i}$ and thus compensates for the intervals in the relative Maslov index $-\mu(\Lambda_S, \mathbb{R})$ that are not contained in the boundary of $\mathbb{S} \setminus U$. Moreover, for $z \in \mathbb{R}$ with $\xi(z) = 0$, the relative Maslov index of the pair $(\Lambda_z|_{\partial U_z \cap \mathbb{R}}, \mathbb{R})$ is zero. The last equation follows from the formulas (75) and (81) for the first two terms and from (83) and (82) for the last two terms. With this understood, the linear index formula (76) follows from equation (79). This proves Lemma 12.3. □

Lemma 12.4. *The operator D_S is injective whenever* $\mathrm{index}(D_S) \le 0$ *and is surjective whenever* $\mathrm{index}(D_S) \ge 0$.

Proof. If $\xi \in \mathcal{W}$ is a nonzero element in the kernel of D_S then ξ satisfies the hypotheses of Lemma 12.3. Moreover, every zero of ξ has positive index by the argument in the proof of Theorem C.1.10 in [20, pages 561/562]. Hence the index of D_S is positive by the linear index formula in Lemma 12.3. This shows that D_S is injective whenever $\mathrm{index}(D_S) \le 0$. If D_S has nonnegative index then the formal adjoint operator $\eta \mapsto -\partial_s\eta + \mathbf{i}\partial_t\eta + S^T\eta$ has nonpositive index and is therefore injective by what we just proved. Since its kernel is the L^2-orthogonal complement of the image of D_S it follows that D_S is surjective. This proves Lemma 12.4. □

Proof of Theorem 12.2. The index formula (69) follows from the linear index formula (76) in Lemma 12.3 with $\xi := \Phi^{-1}\partial_s v$. The index formula shows that D_v has positive index for every nonconstant (α, β)-holomorphic strip $v : \mathbb{S} \to \Sigma$. Hence D_v is onto by Lemma 12.4. This proves Theorem 12.2. □

Proof of Theorem 12.1. The proof has four steps.

Step 1. *The map* (65) *is well defined.*

Let $u, u' : \mathbb{D} \to \Sigma$ be equivalent smooth (α, β)-lunes from x to y. Then there is an orientation preserving diffeomorphism $\varphi : \mathbb{D} \to \mathbb{D}$ such that $\varphi(\pm1) = \pm1$ and $u' := u \circ \varphi$. Consider the holomorphic strips $v := u \circ \varphi_u \circ g : \mathbb{S} \to \Sigma$ and

$$\begin{aligned} v' &:= u' \circ \varphi_{u'} \circ g \\ &= u \circ \varphi \circ \varphi_{u\circ\varphi} \circ g \\ &= v \circ g^{-1} \circ \varphi_u^{-1} \circ \varphi \circ \varphi_{u\circ\varphi} \circ g. \end{aligned}$$

By the definition of φ_u we have $(u \circ \varphi_u)^* J = \mathbf{i}$ and $(u \circ \varphi \circ \varphi_{u\circ\varphi})^* J = \mathbf{i}$. (See equation (63).) Hence the composition $\varphi_u^{-1} \circ \varphi \circ \varphi_{u\circ\varphi} : \mathbb{D} \setminus \{\pm1\} \to \mathbb{D} \setminus \{\pm1\}$ is holomorphic and so is the composition $g^{-1} \circ \varphi_u^{-1} \circ \varphi \circ \varphi_{u\circ\varphi} \circ g : \mathbb{S} \to \mathbb{S}$. Hence this composition is given by a time shift and this proves Step 1.

Step 2. *The map* (65) *is injective.*

Let $u, u' : \mathbb{D} \to \Sigma$ be smooth (α, β)-lunes from x to y and define $v := u \circ \varphi_u \circ g$ and $v' := u' \circ \varphi_{u'} \circ g$. (See equations (63) and (64).) Assume that $v' = v \circ \tau$ for a translation $\tau : \mathbb{S} \to \mathbb{S}$. Then $u' = u \circ \varphi$, where $\varphi : \mathbb{D} \to \mathbb{D}$ is given by $\varphi|_{\mathbb{D}\setminus\{\pm1\}} = \varphi_u \circ g \circ \tau \circ g^{-1} \circ \varphi_{u'}^{-1}$ and $\varphi(\pm1) = \pm1$. Since u and u' are immersions, it follows that φ is a diffeomorphism of $\mathbb{D}$. This proves Step 2.

Step 3. *Every holomorphic (α, β)-strip $v : \mathbb{S} \to \Sigma$ from x to y with Viterbo–Maslov index one is an immersion and satisfies $\nu_x(v) = 0$ and $\nu_y(v) = 1$.*

This follows immediately from the index formula (69) in Theorem 12.2.

Step 4. *The map (65) is surjective.*

Let $v : \mathbb{S} \to \Sigma$ be a holomorphic (α, β)-strip from x to y with Viterbo–Maslov index one. By Step 3, v is an immersion and satisfies $\nu_x(v) = 0$ and $\nu_y(v) = 1$. Hence it follows from (67) and (68) that

$$\begin{aligned} \psi_y(v(s + \mathbf{i}t)) &= c_y e^{-(\pi-\theta_y)(s+\mathbf{i}t)} + O(e^{-(\pi-\theta_y+\delta)s}), & s > T, \\ \psi_x(v(s + \mathbf{i}t)) &= c_x e^{\theta_x(s+\mathbf{i}t)} + O(e^{(\theta_x+\delta)s}), & s < -T, \end{aligned} \tag{84}$$

for T sufficiently large. This implies that the composition

$$u' := v \circ g^{-1} : \mathbb{D} \setminus \{\pm 1\} \to \Sigma$$

(g as in equation (64)) is an immersion and extends continuously to $\mathbb{D}$ by $u'(-1) := x$ and $u'(1) := y$. Moreover, locally near $z = -1$, the image of u' covers only one of the four quadrants into which Σ is divided by α and β and the same holds near $z = 1$.

We must prove that there exists a homeomorphism $\varphi : \mathbb{D} \to \mathbb{D}$ such that

(a) $\varphi(\pm 1) = \pm 1$ and $\varphi(0) = 0$,

(b) φ restricts to an orientation preserving diffeomorphism of $\mathbb{D} \setminus \{\pm 1\}$,

(c) the map $u := u' \circ \varphi^{-1} : \mathbb{D} \to \Sigma$ is a smooth lune.

Once φ has been found it follows from (c) that $u \circ \varphi = v \circ g^{-1} : \mathbb{D} \setminus \{\pm 1\} \to \Sigma$ is holomorphic and hence $\varphi^* u^* J = \mathbf{i}$. Hence it follows from (a) and (b) that $\varphi = \varphi_u$ (see equation (63)) and this implies that the equivalence class $[v] \in \mathcal{M}^{\mathrm{Floer}}(x, y; J)$ belongs to the image of our map (65) as claimed.

To construct φ, choose a smooth function $\lambda : \mathbb{R} \to (1/2, \infty)$ such that

$$\lambda(s) = \begin{cases} \pi/2\theta_x, & \text{for } s \leq -2, \\ 1, & \text{for } s \geq -1, \end{cases}, \qquad \lambda(s) + \lambda'(s)s > 0. \tag{85}$$

(For example define $\lambda(s) := \pi/\theta_x - 1 + (\pi/\theta_x - 2)/s$ for $-2 \leq s \leq -1$ to obtain a piecewise smooth function and approximate by a smooth function.) Then the map $s \mapsto \lambda(s)s$ is a diffeomorphism of $\mathbb{R}$. Consider the sets

$$\mathbb{K}_\lambda := \left\{ e^{s+\mathbf{i}t} \,\middle|\, 0 \leq t \leq \frac{\pi}{2\lambda(s)} \right\}, \qquad \mathbb{K} := \left\{ e^{s+\mathbf{i}t} \,\middle|\, 0 \leq t \leq \frac{\pi}{2} \right\}.$$

Denote their closures by $\overline{\mathbb{K}}_\lambda := \mathbb{K}_\lambda \cup \{0\}$ and $\overline{\mathbb{K}} := \mathbb{K} \cup \{0\}$ and define the homeomorphism $\rho_\lambda : \overline{\mathbb{K}}_\lambda \to \overline{\mathbb{K}}$ by $\rho_\lambda(0) := 0$ and

$$\rho_\lambda(e^{s+\mathbf{i}t}) := e^{(s+\mathbf{i}t)\lambda(s)}, \qquad e^{s+\mathbf{i}t} \in \mathbb{K}_\lambda.$$

It restricts to a diffeomorphism from $\mathbb{K}_\lambda$ to $\mathbb{K}$ and it satisfies $\rho_\lambda(\zeta) = \zeta^{\pi/2\theta_x}$ for $|\zeta| \le e^{-2}$ and $\rho_\lambda(\zeta) = \zeta$ for $|\zeta| \ge e^{-1}$. Consider the map

$$w_\lambda : \overline{\mathbb{K}}_\lambda \to \Sigma, \qquad w_\lambda(\zeta) := u'\left(\frac{\rho_\lambda(\zeta) - 1}{\rho_\lambda(\zeta) + 1}\right). \tag{86}$$

We claim that $w_\lambda : \overline{\mathbb{K}}_\lambda \to \Sigma$ is a C^1 immersion. If $\zeta := e^{s+\mathbf{i}t} \in \mathbb{K}_\lambda$ with $|\zeta| = e^s \le e^{-2}$ then by (64)

$$\frac{\rho_\lambda(e^{s+\mathbf{i}t}) - 1}{\rho_\lambda(e^{s+\mathbf{i}t}) + 1} = \frac{e^{(s+\mathbf{i}t)\pi/2\theta_x} - 1}{e^{(s+\mathbf{i}t)\pi/2\theta_x} + 1} = g\left(\theta_x^{-1}(s + \mathbf{i}t)\right).$$

Insert this as an argument in u' and use the formula $u' \circ g = v$ to obtain

$$w_\lambda(e^{s+\mathbf{i}t}) = u'\left(\frac{\rho_\lambda(e^{s+\mathbf{i}t}) - 1}{\rho_\lambda(e^{s+\mathbf{i}t}) + 1}\right) = v\left(\theta_x^{-1}(s + \mathbf{i}t)\right), \qquad s \le -2. \tag{87}$$

Thus

$$\begin{aligned} \psi_x\left(w_\lambda(\zeta)\right) &= \psi_x\left(v\left(\theta_x^{-1}(s + \mathbf{i}t)\right)\right) \\ &= c_x e^{s+\mathbf{i}t} + O\big(e^{(1+\delta/\theta_x)s}\big) \\ &= c_x\zeta + O\big(|\zeta|^{1+\delta/\theta_x}\big) \end{aligned}$$

for $\zeta = e^{s+\mathbf{i}t} \in \mathbb{K}_\lambda$ sufficiently small. (Here the second equation follows from (84).) Hence the map $w_\lambda : \overline{\mathbb{K}}_\lambda \to \Sigma$ in (86) is complex differentiable at the origin and $d(\psi_x \circ w_\lambda)(0) = c_x$. Next we prove that the derivative of w_λ is continuous. To see this, recall that $\psi_x \circ w_\lambda$ and $\psi_x \circ v$ are holomorphic wherever defined and denote their complex derivatives by $d(\psi_x \circ w_\lambda)$ and $d(\psi_x \circ v)$. Differentiating equation (87) gives

$$e^{s+\mathbf{i}t} d(\psi_x \circ w_\lambda)(e^{s+\mathbf{i}t}) = \theta_x^{-1} d(\psi_x \circ v)\left(\theta_x^{-1}(s + \mathbf{i}t)\right).$$

By Corollary D.2 (i), $e^{-\theta_x(s+\mathbf{i}t)} d(\psi_x \circ v)(s + \mathbf{i}t)$ converges uniformly to $\theta_x c_x$ as s tends to $-\infty$. Hence $d(\psi_x \circ w_\lambda)(\zeta) = \theta_x^{-1} e^{-(s+\mathbf{i}t)} d(\psi_x \circ v)(\theta_x^{-1}(s + \mathbf{i}t))$ converges to $c_x = d(\psi_x \circ w_\lambda)(0)$ as $\zeta = e^{s+\mathbf{i}t} \in \mathbb{K}_\lambda$ tends to zero. Hence w_λ is continuously differentiable near the origin and is a C^1 immersion as claimed.

Now let $\sigma_\lambda : \overline{\mathbb{K}}_\lambda \to \mathbb{D} \setminus \{1\}$ be any diffeomorphism that satisfies

$$\sigma_\lambda(\zeta) = \frac{\zeta - 1}{\zeta + 1} \qquad \text{for} \quad |\zeta| \geq 1.$$

Define $\varphi'' : \mathbb{D} \to \mathbb{D}$ and $u'' : \mathbb{D} \to \Sigma$ by $\varphi''(1) := 1$, $u''(1) := y$, and

$$\varphi''(z) := \frac{\rho_\lambda(\sigma_\lambda^{-1}(z)) - 1}{\rho_\lambda(\sigma_\lambda^{-1}(z)) + 1}, \qquad u''(z) := u' \circ \varphi''(z) = w_\lambda \circ \sigma_\lambda^{-1}(z)$$

for $z \in \mathbb{D} \setminus \{1\}$. Then $\varphi'' = \mathrm{id}$ on $\{z \in \mathbb{D} \,|\, \mathrm{Re}\, z \geq 0\}$, the map $\varphi'' : \mathbb{D} \to \mathbb{D}$ satisfies (a) and (b), and $u''|_{\mathbb{D}\setminus\{1\}}$ is an orientation preserving C^1 immersion. A similar construction near y yields an orientation preserving C^1 immersion $u''' = u' \circ \varphi''' : \mathbb{D} \to \Sigma$ where $\varphi''' : \mathbb{D} \to \mathbb{D}$ satisfies (a) and (b). Now approximate u''' in the C^1-topology by a smooth lune $u = u' \circ \varphi : \mathbb{D} \to \Sigma$ to obtain the required map φ. This proves Step 4 and Theorem 12.1. □

Proof of Theorem 9.3. By Theorem 12.2, the linearized operator D_v in Floer theory is surjective for every (α, β)-holomorphic strip v. Hence there is a boundary operator on the $\mathbb{Z}_2$ vector space $\mathrm{CF}(\alpha, \beta)$ as defined by Floer [10, 11] in terms of the mod two count of (α, β)-holomorphic strips. By Theorem 12.1 this boundary operator agrees with the combinatorial one defined in terms of the mod two count of (α, β)-lunes. Hence the combinatorial Floer homology of the pair (α, β) agrees with the analytic Floer homology defined by Floer. This proves Theorem 9.3. □

Remark 12.5 (Hearts and Diamonds). We have seen that the combinatorial boundary operator ∂ on $\mathrm{CF}(\alpha, \beta)$ agrees with Floer's boundary operator by Theorem 12.1. Thus we have two proofs that $\partial^2 = 0$: the combinatorial proof using broken hearts and Floer's proof using his gluing construction. He showed (in much greater generality) that two (α, β)-holomorphic strips of index one (one from x to y and one from y to z) can be glued together to give rise to a 1-parameter family of (α, β)-holomorphic strips (modulo time shift) of index two from x to z. This one parameter family can be continued until it ends at another pair of (α, β)-holomorphic strips of index one (one from x to some intersection point y' and one from y' to z). These one parameter families are in one-to-one correspondence to (α, β)-hearts from x to z. This can be seen geometrically as follows. Each glued (α, β)-holomorphic strip from x to z has a critical point on the β-boundary near y for a broken heart of type (a). The 1-manifold is parametrized by the position of the critical value. There is precisely one (α,β)-holomorphic strip in this moduli space without critical point and an angle between π and 2π at z. The critical value then moves onto the α-boundary and tends towards y' at the other end of the moduli space giving a broken heart of type (c) (See Figure 23).

Here is an explicit formula for the gluing construction in the two dimensional setting. Let $h = (u, y, v)$ be a broken (α,β)-heart of type (a) or (b) from x to z. (Types (c) and (d) are analogous with α and β interchanged.) Denote the left and right upper quadrants by $Q_L := (-\infty, 0) + \mathbf{i}(0,\infty)$ and $Q_R := (0,\infty) + \mathbf{i}(0,\infty)$. Define diffeomorphisms $\psi_L : Q_L \to \mathbb{D}\setminus\partial\mathbb{D}$ and $\psi_R : Q_R \to \mathbb{D}\setminus\partial\mathbb{D}$ by

$$\psi_L(\zeta) := \frac{1+\zeta}{1-\zeta}, \qquad \psi_R(\zeta) := \frac{\zeta-1}{\zeta+1}.$$

The extensions of these maps to Möbius transformations of the Riemann sphere are inverses of each other. Define the map $w : Q_L \cup Q_R \to \Sigma$ by

$$w(\zeta) := \begin{cases} u(\psi_L(\zeta)), & \text{for } \zeta \in Q_L, \\ v(\psi_R(\zeta)), & \text{for } \zeta \in Q_R. \end{cases}$$

The maps $u \circ \psi_L$ and $v \circ \psi_R$ send suitable intervals on the imaginary axis starting at the origin to the same arc on β. Modify u and v so that w extends to a smooth map on the slit upper half plane $\mathbb{H}\setminus\mathbf{i}[1,\infty)$, still denoted by w. Here $\mathbb{H}\subset\mathbb{C}$ is the closed upper half plane. Define $\varphi_\varepsilon : \mathbb{D}\to\mathbb{H}$ by

$$\varphi_\varepsilon(z) := \frac{2\varepsilon z}{1-z^2}, \qquad z\in\mathbb{D}, \qquad 0<\varepsilon<1.$$

This map sends the open set $\mathrm{int}(\mathbb{D})\cap Q_L$ diffeomorphically onto Q_L and it sends $\mathrm{int}(\mathbb{D})\cap Q_R$ diffeomorphically onto Q_R. It also sends the interval $\mathbf{i}[0,1]$ diffeomorphically onto $\mathbf{i}[0,\varepsilon]$. The composition $w\circ\varphi_\varepsilon : \mathrm{int}(\mathbb{D})\to\Sigma$ extends to a smooth map on $\mathbb{D}$ denoted by $w_\varepsilon : \mathbb{D}\to\Sigma$. An explicit formula for w_ε is

$$w_\varepsilon(z) = \begin{cases} u\left(\frac{1-z^2+2\varepsilon z}{1-z^2-2\varepsilon z}\right), & \text{if } z\in\mathbb{D} \text{ and } \mathrm{Re}\, z\le 0, \\ v\left(\frac{-1+z^2+2\varepsilon z}{1+z^2+2\varepsilon z}\right), & \text{if } z\in\mathbb{D} \text{ and } \mathrm{Re}\, z\ge 0. \end{cases}$$

The derivative of this map at every point $z\ne\mathbf{i}$ is an orientation preserving isomorphism. Its only critical value is the point

$$c_\varepsilon := u\left(\frac{1+\mathbf{i}\varepsilon}{1-\mathbf{i}\varepsilon}\right) = v\left(\frac{\mathbf{i}\varepsilon-1}{\mathbf{i}\varepsilon+1}\right)\in\beta.$$

Note that c_ε tends to $y = u(1) = v(-1)$ as ε tends to zero. The composition of w_ε with a suitable ε-dependent diffeomorphism $\mathbb{S} \to \mathbb{D} \setminus \{\pm 1\}$ gives the required one-parameter family of glued holomorphic strips.

13 Further Developments

There are many directions in which the theory developed in the present memoir can be extended. Some of these directions and related work in the literature are discussed below.

Floer Homology

If one drops the hypothesis that the loops α and β are not contractible and not isotopic to each other there are three possibilities. In some cases the Floer homology groups are still well defined and invariant under (Hamiltonian) isotopy, in other cases invariance under isotopy breaks down, and there are examples with $\partial \circ \partial \neq 0$, so Floer homology is not even defined. All these phenomena have their counterparts in combinatorial Floer homology.

A case in point is that of two transverse embedded circles $\alpha, \beta \subset \mathbb{C}$ in the complex plane. In this case the boundary operator

$$\partial : \mathrm{CF}(\alpha, \beta) \to \mathrm{CF}(\alpha, \beta)$$

of Section 9 still satisfies $\partial \circ \partial = 0$. However, $\ker \partial = \operatorname{im} \partial$ and so the (combinatorial) Floer homology groups vanish. This must be true because (combinatorial) Floer homology is still invariant under isotopy and the loops can be disjoined by a translation.

A second case is that of two transverse embedded loops in the sphere $\Sigma = S^2$. Here the Floer homology groups are nonzero when the loops intersect and vanish otherwise. An interesting special case is that of two equators. (Following Khanevsky we call an embedded circle $\alpha \subset S^2$ an **equator** when the two halves of $S^2 \setminus \alpha$ have the same area.) In this case the combinatorial Floer homology groups do not vanish, but are only invariant under Hamiltonian isotopy. This is an example of the monotone case for Lagrangian Floer theory (see Oh [21]), and the theory developed by Biran–Cornea applies [5]. For an interesting study of diameters (analogues of equators for discs) see Khanevsky [17].

A similar case is that of two noncontractible transverse embedded loops

$$\alpha, \beta \subset \Sigma$$

that are isotopic to each other. Fix an area form on Σ. If β is Hamiltonian isotopic to α then the combinatorial Floer homology groups do not vanish and are invariant under Hamiltonian isotopy, as in the case of two equators on S^2. If β is non-Hamiltonian isotopic to α (for example a distinct parallel copy), then the Floer homology groups are no longer invariant under Hamiltonian isotopy, as in the case of two embedded circles in S^2 that are not equators. However, if we take account of the areas of the lunes by introducing combinatorial Floer homology with coefficients in an appropriate Novikov ring, the Floer homology groups will be invariant under Hamiltonian isotopy. When the Floer homology groups vanish it is interesting to give a combinatorial description of the relevant torsion invariants (see Hutchings–Lee [15, 16]). This involves an interaction between lunes and annuli.

A different situation occurs when α is not contractible and β is contractible. In this case

$$\partial \circ \partial = \mathrm{id}$$

and one can prove this directly in the combinatorial setting. For example, if α and β intersect in precisely two points x and y then there is precisely one lune from x to y and precisely one lune from y to x. They are embedded and their union is the disc encircled by β rather than a heart as in Section 10. In the analytical setting this disc bubbles off in the moduli space of index two holomorphic strips from x to itself. This is a simple example of the obstruction theory developed in great generality by Fukaya–Oh–Ohta–Ono [13].

Moduli Spaces

Another direction is to give a combinatorial description of all holomorphic strips, not just those of index one. The expected result is that they are uniquely determined, up to translation, by their (α, β)-trace

$$\Lambda = (x, y, \mathrm{w})$$

with $\mathrm{w} \geq 0$, the positions of the critical values, suitable monodromy data, and the angles at infinity. (See Remark 12.5 for a discussion of the Viterbo–Maslov index two case.) This can be viewed as a natural generalization of Riemann–Hurwitz theory. For inspiration see the work of Okounkov and Pandharipande on the Gromov–Witten theory of surfaces [22, 23, 24, 25].

The Donaldson Triangle Product

Another step in the program, already discussed in [6], is the combinatorial description of the product structures

$$\mathrm{HF}(\alpha, \beta) \otimes \mathrm{HF}(\beta, \gamma) \to \mathrm{HF}(\alpha, \gamma)$$

for triples of noncontractible, pairwise nonisotopic, and pairwise transverse embedded loops in a closed oriented 2-manifold Σ. The combinatorial setup involves the study of immersed triangles in Σ. When the triangle count is infinite, for example on the 2-torus, the definition of the product requires Floer homology with coefficients in Novikov rings. The proof that the resulting map

$$\mathrm{CF}(\alpha, \beta) \otimes \mathrm{CF}(\beta, \gamma) \to \mathrm{CF}(\alpha, \gamma)$$

on the chain level is a chain homomorphism is based on similar arguments as in Section 10. The proof that the product on homology is invariant under isotopy is based on similar arguments as in Section 11. A new ingredient is the phenomenon that γ can pass over an intersection point of α and β in an isotopy. In this case the number of intersection points does not change but it is necessary to understand how the product map changes on the chain level. The proof of associativity requires the study of immersed rectangles and uses similar arguments as in Section 10.

In the case of the 2-torus the study of triangles gives rise to Theta-functions as noted by Kontsevich [18]. This is an interesting, and comparatively easy, special case of homological mirror symmetry.

The Fukaya Category

A natural extension of the previous discussion is to give a combinatorial description of the Fukaya category [13]. A directed version of this category was described by Seidel [34]. In dimension two the directed Fukaya category is associated to a finite ordered collection

$$\alpha_1, \alpha_2, \ldots, \alpha_n \subset \Sigma$$

of noncontractible, pairwise nonisotopic, and pairwise transverse embedded loops in Σ. Interesting examples of such tuples arise from vanishing cycles of Lefschetz fibrations over the disc with regular fiber Σ (see [34]).

The Fukaya category, on the combinatorial level, involves the study of immersed polygons. Some of the results in the present memoir (such as the combinatorial techniques in Sections 10 and 11, the surjectivity of the Fredholm operator, and the formula for the Viterbo–Maslov index in Section 12) extend naturally to this setting. On the other hand the algebraic structures are considerably more intricate for A^∞ categories. The combinatorial approach has been used to compute the derived Fukaya category of a surface by Abouzaid [1], and to establish homological mirror symmetry for punctured spheres by Abouzaid–Auroux–Efimov–Katzarkov–Orlov [2] and for a genus two surface by Seidel [35].

Appendices

A The Space of Paths

We assume throughout that Σ is a connected oriented smooth 2-manifold without boundary and $\alpha, \beta \subset \Sigma$ are two embedded loops. Let

$$\Omega_{\alpha,\beta} := \{x \in C^\infty([0,1],\Sigma) \,|\, x(0) \in \alpha,\ x(1) \in \beta\}$$

denote the space of paths connecting α to β.

Proposition A.1. *Assume that α and β are not contractible and that α is not isotopic to β. Then each component of $\Omega_{\alpha,\beta}$ is simply connected and hence $H^1(\Omega_{\alpha,\beta};\mathbb{R}) = 0$.*

The proof was explained to us by David Epstein [9]. It is based on the following three lemmas. We identify $S^1 \cong \mathbb{R}/\mathbb{Z}$.

Lemma A.2. *Let $\gamma : S^1 \to \Sigma$ be a noncontractible loop and denote by*

$$\pi : \widetilde{\Sigma} \to \Sigma$$

the covering generated by γ. Then $\widetilde{\Sigma}$ is diffeomorphic to the cylinder.

Proof. By hypothesis, Σ is oriented and has a nontrivial fundamental group. By the uniformization theorem, choose a metric of constant curvature. Then the universal cover of Σ is isometric to either $\mathbb{R}^2$ with the flat metric or to the upper half space $\mathbb{H}^2$ with the hyperbolic metric. The 2-manifold $\widetilde{\Sigma}$ is a quotient of the universal cover of Σ by the subgroup of the group of covering transformations generated by a single element (a translation in the case of $\mathbb{R}^2$ and a hyperbolic element of $\mathrm{PSL}(2,\mathbb{R})$ in the case of $\mathbb{H}^2$). Since γ is not contractible, this element is not the identity. Hence $\widetilde{\Sigma}$ is diffeomorphic to the cylinder. □

Lemma A.3. *Let $\gamma : S^1 \to \Sigma$ be a noncontractible loop and, for $k \in \mathbb{Z}$, define $\gamma^k : S^1 \to \Sigma$ by*

$$\gamma^k(s) := \gamma(ks).$$

Then γ^k is contractible if and only if $k = 0$.

Proof. Let $\pi : \widetilde{\Sigma} \to \Sigma$ be as in Lemma A.2. Then, for $k \neq 0$, the loop $\gamma^k : S^1 \to \Sigma$ lifts to a noncontractible loop in $\widetilde{\Sigma}$. $\square$

Lemma A.4. *Let $\gamma_0, \gamma_1 : S^1 \to \Sigma$ be noncontractible embedded loops and suppose that k_0, k_1 are nonzero integers such that $\gamma_0^{k_0}$ is homotopic to $\gamma_1^{k_1}$. Then either γ_1 is homotopic to γ_0 and $k_1 = k_0$ or γ_1 is homotopic to ${\gamma_0}^{-1}$ and $k_1 = -k_0$.*

Proof. Let $\pi : \widetilde{\Sigma} \to \Sigma$ be the covering generated by γ_0. Then ${\gamma_0}^{k_0}$ lifts to a closed curve in $\widetilde{\Sigma}$ and is homotopic to ${\gamma_1}^{k_1}$. Hence ${\gamma_1}^{k_1}$ lifts to a closed immersed curve in $\widetilde{\Sigma}$. Hence there exists a nonzero integer j_1 such that ${\gamma_1}^{j_1}$ lifts to an embedding $S^1 \to \widetilde{\Sigma}$. Any embedded curve in the cylinder is either contractible or is homotopic to a generator. If the lift of ${\gamma_1}^{j_1}$ were contractible it would follow that ${\gamma_0}^{k_0}$ is contractible, hence, by Lemma A.3, $k_0 = 0$ in contradiction to our hypothesis. Hence the lift of ${\gamma_1}^{j_1}$ to $\widetilde{\Sigma}$ is not contractible. With an appropriate sign of j_1 it follows that the lift of ${\gamma_1}^{j_1}$ is homotopic to the lift of γ_0. Interchanging the roles of γ_0 and γ_1, we find that there exist nonzero integers j_0, j_1 such that

$$\gamma_0 \sim {\gamma_1}^{j_1}, \qquad \gamma_1 \sim {\gamma_0}^{j_0}$$

in $\widetilde{\Sigma}$. Hence γ_0 is homotopic to ${\gamma_0}^{j_0 j_1}$ in the free loop space of $\widetilde{\Sigma}$. Since the homotopy lifts to the cylinder $\widetilde{\Sigma}$ and the fundamental group of $\widetilde{\Sigma}$ is abelian, it follows that

$$j_0 j_1 = 1.$$

If $j_0 = j_1 = 1$ then γ_1 is homotopic to γ_0, hence $\gamma_0^{k_1}$ is homotopic to ${\gamma_0}^{k_0}$, hence ${\gamma_0}^{k_0 - k_1}$ is contractible, and hence $k_0 - k_1 = 0$, by Lemma A.3. If $j_0 = j_1 = -1$ then γ_1 is homotopic to ${\gamma_0}^{-1}$, hence $\gamma_0^{-k_1}$ is homotopic to ${\gamma_0}^{k_0}$, hence ${\gamma_0}^{k_0 + k_1}$ is contractible, and hence $k_0 + k_1 = 0$, by Lemma A.3. This proves Lemma A.4. $\square$

Proof of Proposition A.1. Orient α and β and and choose orientation preserving diffeomorphisms

$$\gamma_0 : S^1 \to \alpha, \qquad \gamma_1 : S^1 \to \beta.$$

A closed loop in $\Omega_{\alpha,\beta}$ gives rise to a map $u : S^1 \times [0,1] \to \Sigma$ such that

$$u(S^1 \times \{0\}) \subset \alpha, \qquad u(S^1 \times \{1\}) \subset \beta.$$

Let k_0 denote the degree of $u(\cdot, 0) : S^1 \to \alpha$ and k_1 denote the degree of $u(\cdot, 1) : S^1 \to \beta$. Since the homotopy class of a map $S^1 \to \alpha$ or a map $S^1 \to \beta$ is determined by the degree we may assume, without loss of generality, that

$$u(s, 0) = \gamma_0(k_0 s), \qquad u(s, 1) = \gamma_1(k_1 s).$$

If one of the integers k_0, k_1 vanishes, so does the other, by Lemma A.3. If they are both nonzero then γ_1 is homotopic to either γ_0 or γ_0^{-1}, by Lemma A.4. Hence γ_1 is isotopic to either γ_0 or γ_0^{-1}, by [8, Theorem 4.1]. Hence α is isotopic to β, in contradiction to our hypothesis. This shows that

$$k_0 = k_1 = 0.$$

With this established it follows that the map $u : S^1 \times [0, 1] \to \Sigma$ factors through a map $v : S^2 \to \Sigma$ that maps the south pole to α and the north pole to β. Since $\pi_2(\Sigma) = 0$ it follows that v is homotopic, via maps with fixed north and south pole, to one of its meridians. This proves Proposition A.1. □

B Diffeomorphisms of the Half Disc

Proposition B.1. *The group of orientation preserving diffeomorphisms $\varphi : \mathbb{D} \to \mathbb{D}$ that satisfy $\varphi(1) = 1$ and $\varphi(-1) = -1$ is connected.*

Proof. Choose φ as in the proposition. We prove in five steps that φ is isotopic to the identity.

Step 1. *We may assume that $d\varphi(-1) = d\varphi(1) = \mathbb{1}$.*

The differential of φ at -1 has the form

$$d\varphi(-1) = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}.$$

Let $X : \mathbb{D} \to \mathbb{R}^2$ be a vector field on $\mathbb{D}$ that is tangent to the boundary, is supported in an ε-neighborhood of -1, and satisfies

$$dX(-1) = \begin{pmatrix} \log a & 0 \\ 0 & \log b \end{pmatrix}.$$

Denote by $\psi_t : \mathbb{D} \to \mathbb{D}$ the flow of X. Then $d\psi_1(-1) = d\varphi(-1)$. Replace φ by $\varphi \circ {\psi_1}^{-1}$.

Step 2. *We may assume that φ is equal to the identity map near ± 1.*

Choose local coordinates near -1 that identify a neighborhood of -1 with a neighborhood of zero in the right upper quadrant Q. This gives rise to a local diffeomorphism $\psi : Q \to Q$ such that $\psi(0) = 0$. Choose a smooth cutoff function $\rho : [0, \infty) \to [0, 1]$ such that

$$\rho(r) = \begin{cases} 1, & \text{for } r \leq 1/2, \\ 0, & \text{for } r \geq 1, \end{cases}$$

For $0 \leq t \leq 1$ define $\psi_t : Q \to Q$ by

$$\psi_t(z) := \psi(z) + t\rho(|z|^2/\varepsilon^2))(z - \psi(z)).$$

Since $d\psi(0) = \mathbb{1}$ this map is a diffeomorphism for every $t \in [0, 1]$ provided that $\varepsilon > 0$ is sufficiently small. Moreover, $\psi_t(z) = \psi(z)$ for $|z| \geq \varepsilon$, $\psi_0 = \psi$, and $\psi_1(z) = z$ for $|z| \leq \varepsilon/2$.

Step 3. *We may assume that φ is equal to the identity map near ± 1 and on $\partial\mathbb{D}$.*

Define $\tau : [0, \pi] \to [0, \pi]$ by

$$\varphi(e^{i\theta}) = e^{i\tau(\theta)}.$$

Let $X_t : \mathbb{D} \to \mathbb{R}^2$ be a vector field that is equal to zero near ± 1 and satisfies

$$X_t(z + t(\varphi(z) - z)) = \varphi(z) - z$$

for $z \in \mathbb{D} \cap \mathbb{R}$ and

$$X_t(z) = i(\tau(\theta) - \theta)z, \qquad z = e^{i(\theta + t(\tau(\theta) - \theta))}.$$

Let $\psi_t : \mathbb{D} \to \mathbb{D}$ be the isotopy generated by X_t via $\partial_t \psi_t = X_t \circ \psi_t$ and $\psi_0 = \mathrm{id}$. Then ψ_1 agrees with ψ on $\partial\mathbb{D}$ and is equal to the identity near ± 1. Replace φ by $\varphi \circ {\psi_1}^{-1}$.

Step 4. *We may assume that φ is equal to the identity map near $\partial\mathbb{D}$.*

Write

$$\varphi(x + iy) = u(x, y) + iv(x, y).$$

Then

$$u(x, 0) = x, \qquad \partial_y v(x, 0) = a(x).$$

Choose a cutoff function ρ equal to one near zero and equal to zero near one. Define

$$\varphi_t(x,y) := u_t(x,y) + v_t(x,y)$$

where

$$u_t(x,y) := u(x,y) + t\rho(y/\varepsilon)(x - u(x,y))$$

and

$$v_t(x,y) := v(x,y) + t\rho(y/\varepsilon)(a(x)y - v(x,y)).$$

If $\varepsilon > 0$ is sufficiently small then φ_t is a diffeomorphism for every $t \in [0,1]$. Moreover, $\varphi_0 = \varphi$ and φ_1 satisfies

$$\varphi_1(x+iy) = x + ia(x)y$$

for $y \geq 0$ sufficiently small. Now choose a smooth family of vector fields $X_t : \mathbb{D} \to \mathbb{D}$ that vanish on the boundary and near ± 1 and satisfy

$$X_t(x + i(y + t(a(x)y - y))) = i(a(x)y - y)$$

near the real axis. Then the isotopy ψ_t generated by X_t satisfies $\psi_t(x+iy) = x + iy + it(a(x)y - y)$ for y sufficiently small. Hence ψ_1 agrees with φ_1 near the real axis. Hence $\varphi \circ {\psi_1}^{-1}$ has the required form near $\mathbb{D} \cap \mathbb{R}$. A similar isotopy near $\mathbb{D} \cap S^1$ proves Step 4.

Step 5. *We prove the proposition.*

Choose a continuous map $f : \mathbb{D} \to S^2 = \mathbb{C} \cup \{\infty\}$ such that $f(\partial\mathbb{D}) = \{0\}$ and f restricts to a diffeomorphism from $\mathbb{D} \setminus \partial\mathbb{D}$ to $S^2 \setminus \{0\}$. Define $\psi : S^2 \to S^2$ by

$$f \circ \psi = \varphi \circ f.$$

Then ψ is a diffeomorphism, equal to the identity near the origin. By a well known Theorem of Smale [37] (see also [7] and [14]) ψ is isotopic to the identity. Compose with a path in SO(3) which starts and ends at the identity to obtain an isotopy $\psi_t : S^2 \to S^2$ such that $\psi_t(0) = 0$. Let

$$\Psi_t := d\psi_t(0), \qquad U_t := \Psi_t({\Psi_t}^T \Psi_t)^{-1/2}.$$

Then $U_t \in \mathrm{SO}(2)$ and $U_0 = U_1 = \mathbb{1}$. Replacing ψ_t by ${U_t}^{-1}\psi_t$ we may assume that $U_t = \mathbb{1}$ and hence Ψ_t is positive definite for every t. Hence there exists a smooth path $[0,1] \to \mathbb{R}^{2\times 2} : t \mapsto A_t$ such that

$$e^{A_t} = \Psi_t, \qquad A_0 = A_1 = 0.$$

Choose a smooth family of compactly supported vector fields $X_t : \mathbb{C} \to \mathbb{C}$ such that

$$dX_t(0) = A_t, \qquad X_0 = X_1 = 0.$$

For every fixed t let $\chi_t : S^2 \to S^2$ be the time-1 map of the flow of X_t. Then

$$\chi_t(0) = 0, \qquad d\chi_t(0) = \Psi_t, \qquad \chi_0 = \chi_1 = \mathrm{id}.$$

Hence the diffeomorphisms

$$\psi_t' := \psi_t \circ {\chi_t}^{-1}$$

form an isotopy from $\psi_0' = \mathrm{id}$ to $\psi_1' = \psi$ such that $\psi_t'(0) = 0$ and $d\psi_t'(0) = \mathbb{1}$ for every t. Now let $\rho : \mathbb{R} \to [0,1]$ be a smooth cutoff function that is equal to one near zero and equal to zero near one. Define

$$\psi_t''(z) := \psi_t'(z) + \rho(|z|/\varepsilon)(z - \psi_t'(z)).$$

For $\varepsilon > 0$ sufficiently small this is an isotopy from $\psi_0'' = \mathrm{id}$ to $\psi_1'' = \psi$ such that $\psi_t = \mathrm{id}$ near zero for every t. The required isotopy $\varphi_t : \mathbb{D} \to \mathbb{D}$ is now given by $f \circ \psi_t = \varphi_t \circ f$. This proves Proposition B.1. $\square$

C Homological Algebra

Let P be a finite set and $\nu : P \times P \to \mathbb{Z}$ be a function that satisfies

$$\sum_{q \in P} \nu(r,q)\nu(q,p) = 0 \tag{88}$$

for all $p, r \in P$. Any such function determines a chain complex $\partial : C \to C$, where $C = C(P)$ and $\partial = \partial^\nu$ are defined by

$$C := \bigoplus_{p \in P} \mathbb{Z}p, \qquad \partial q := \sum_{p \in P} \nu(q,p)p$$

for $q \in P$. Throughout we fix two elements $\bar{p}, \bar{q} \in P$ such that $\nu(\bar{q}, \bar{p}) = 1$. Consider the set

$$P' := P \setminus \{\bar{p}, \bar{q}\} \tag{89}$$

and the function $\nu' : P' \times P' \to \mathbb{Z}$ defined by

$$\nu'(q,p) := \nu(q,p) - \nu(q,\bar{p})\nu(\bar{q},p) \tag{90}$$

for $p, q \in P$ and denote $C' := C(P')$ and $\partial' := \partial^{\nu'}$. The following lemma is due to Floer [11].

Lemma C.1 (**Floer**)**.** *The endomorphism $\partial' : C' \to C'$ is a chain complex and its homology $H(C', \partial')$ is isomorphic to $H(C, \partial)$.*

Proof. The proof consists of four steps.

Step 1. $\partial' \circ \partial' = 0$.

Let $r \in P'$. Then $\partial' \circ \partial' r = \sum_{p \in P'} \mu'(r,p)p$ where $\mu'(r,p) \in \mathbb{Z}$ is given by

$$\begin{aligned}
\mu'(r,p) &= \sum_{q \in P'} \nu'(r,q)\nu'(q,p) \\
&= \sum_{q \in P} \left(\nu(r,q) - \nu(r,\bar{p})\nu(\bar{q},q)\right)\left(\nu(q,p) - \nu(q,\bar{p})\nu(\bar{q},p)\right) \\
&= 0
\end{aligned}$$

for $p \in P'$. Here the first equation follows from the fact that $\nu(\bar{q},\bar{p}) = 1$ and the last equation follows from the fact that $\partial \circ \partial = 0$.

Step 2. *The operator $\Phi : C' \to C$ defined by*

$$\Phi q := q - \nu(q,\bar{p})\bar{q} \tag{91}$$

for $q \in P'$ is a chain map, i.e.

$$\Phi \circ \partial' = \partial \circ \Phi.$$

For $q \in P'$ we have

$$\begin{aligned}
\Phi\partial' q &= \sum_{p \in P'} \nu'(q,p)\Phi p \\
&= \sum_{p \in P'} \left(\nu(q,p) - \nu(q,\bar{p})\nu(\bar{q},p)\right)\left(p - \nu(p,\bar{p})\bar{q}\right) \\
&= \sum_{p \in P} \left(\nu(q,p) - \nu(q,\bar{p})\nu(\bar{q},p)\right)\left(p - \nu(p,\bar{p})\bar{q}\right) \\
&= \sum_{p \in P} \left(\nu(q,p) - \nu(q,\bar{p})\nu(\bar{q},p)\right) p \\
&= \partial q - \nu(q,\bar{p})\partial\bar{q} \\
&= \partial\Phi q.
\end{aligned}$$

Step 3. *The operator $\Psi : C \to C'$ defined by $\Psi q = q$ for $q \in P'$ and*

$$\Psi\bar{q} := 0, \qquad \Psi\bar{p} := -\sum_{p\in P'} \nu(\bar{q}, p)p \tag{92}$$

is a chain map, i.e.

$$\partial' \circ \Psi = \Psi \circ \partial.$$

For $q \in P'$ we have

$$\begin{aligned}
\Psi\partial q &= \sum_{p\in P} \nu(q,p)\Psi p \\
&= \sum_{p\in P'} \nu(q,p)p + \nu(q,\bar{p})\Psi\bar{p} \\
&= \sum_{p\in P'} \left(\nu(q,p) - \nu(q,\bar{p})\nu(\bar{q},p)\right) p \\
&= \partial' q.
\end{aligned}$$

Moreover,

$$\Psi\partial\bar{q} = \sum_{p\in P} \nu(\bar{q},p)\Psi p = \sum_{p\in P'} \nu(\bar{q},p)p + \Psi\bar{p} = 0 = \partial'\Psi\bar{q},$$

and

$$\begin{aligned}
\partial'\Psi\bar{p} &= -\sum_{q\in P'} \nu(\bar{q},q)\partial' q \\
&= -\sum_{q\in P'}\sum_{p\in P'} \nu(\bar{q},q)\left(\nu(q,p) - \nu(q,\bar{p})\nu(\bar{q},p)\right) p \\
&= \sum_{p\in P'} \left(\nu(\bar{p},p) - \nu(\bar{p},\bar{p})\nu(\bar{q},p)\right) p \\
&= \sum_{p\in P'} \nu(\bar{p},p)p + \nu(\bar{p},\bar{p})\Psi\bar{p} \\
&= \sum_{p\in P} \nu(\bar{p},p)\Psi p \\
&= \Psi\partial\bar{p}.
\end{aligned}$$

Step 4. *The operator $\Psi \circ \Phi : C' \to C'$ is equal to the identity and*

$$\mathrm{id} - \Phi \circ \Psi = \partial \circ T + T \circ \partial,$$

where $T : C \to C$ is defined by $T\bar{p} = \bar{q}$ and $Tq = 0$ for $q \in P \setminus \{\bar{p}\}$.

For $q \in P'$ we have

$$\Phi\Psi q = \Phi q = q - \nu(q, \bar{p})\bar{q}$$

and hence

$$q - \Phi\Psi q = \nu(q, \bar{p})\bar{q} = \nu(q, \bar{p})T\bar{p} = T\partial q = T\partial q + \partial T q.$$

Moreover,

$$\bar{q} - \Phi\Psi\bar{q} = \bar{q} = \nu(\bar{q}, \bar{p})T\bar{p} = T\partial\bar{q} = T\partial\bar{q} + \partial T\bar{q}$$

and

$$\begin{aligned}
\bar{p} - \Phi\Psi\bar{p} &= \bar{p} + \sum_{p \in P'} \nu(\bar{q}, p)\Phi p \\
&= \bar{p} + \sum_{p \in P'} \nu(\bar{q}, p)p - \sum_{p \in P'} \nu(\bar{q}, p)\nu(p, \bar{p})\bar{q} \\
&= \bar{p} + \sum_{p \in P'} \nu(\bar{q}, p)p + \nu(\bar{q}, \bar{q})\bar{q} + \nu(\bar{p}, \bar{p})\bar{q} \\
&= \partial\bar{q} + \nu(\bar{p}, \bar{p})\bar{q} \\
&= \partial T\bar{p} + T\partial\bar{p}.
\end{aligned}$$

This proves Lemma C.1. □

Now let $(P, \preceq)$ be a finite poset. An ordered pair $(p, q) \in P \times P$ is called **adjacent** if $p \preceq q$, $p \neq q$, and

$$p \preceq r \preceq q \qquad \Longrightarrow \qquad r \in \{p, q\}.$$

Fix an adjacent pair $(\bar{p}, \bar{q}) \in P \times P$ and consider the relation $\preceq'$ on

$$P' := P \setminus \{\bar{p}, \bar{q}\}$$

defined by

$$p \preceq' q \qquad \Longleftrightarrow \qquad \left\{ \begin{array}{ll} \text{either} & p \preceq q, \\ \text{or} & \bar{p} \preceq q \text{ and } p \preceq \bar{q}. \end{array} \right. \tag{93}$$

Lemma C.2. $(P', \preceq')$ *is a poset.*

Proof. We prove that the relation $\preceq'$ is transitive. Let $p, q, r \in P'$ such that

$$p \preceq' q, \qquad q \preceq' r.$$

There are four cases.
Case 1: $p \preceq q$ *and* $q \preceq r$. Then $p \preceq r$ and hence $p \preceq' r$.
Case 2: $p \not\preceq q$ *and* $q \preceq r$. Then

$$\bar{p} \preceq q \preceq r, \qquad p \preceq \bar{q},$$

and hence $p \preceq' r$.
Case 3: $p \preceq q$ *and* $q \not\preceq r$. The argument is as in the Case 2.
Case 4: $p \not\preceq q$ *and* $q \not\preceq r$. Then

$$p \preceq \bar{q}, \qquad \bar{p} \preceq r,$$

and hence $p \preceq' r$.

Next we prove that the relation $\preceq'$ is anti-symmetric. Hence assume that $p, q \in P'$ such that $p \preceq' q$ and $q \preceq' p$. We claim that $p \preceq q$ and $q \preceq p$. Assume otherwise that $p \not\preceq q$. Then $\bar{p} \preceq q$ and $p \preceq \bar{q}$. Since $q \preceq' p$, it follows that $\bar{p} \preceq p \preceq \bar{q}$ and $\bar{p} \preceq q \preceq \bar{q}$, and hence $\{p, q\} \subset \{\bar{p}, \bar{q}\}$, a contradiction. Thus we have shown that $p \preceq q$. Similarly, $q \preceq p$ and hence $p = q$. This proves Lemma C.2 $\square$

A function $\mu : P \to \mathbb{Z}$ is called an **index function** for $(P, \preceq)$ if

$$p \preceq q \qquad \Longrightarrow \qquad \mu(p) < \mu(q). \tag{94}$$

Let μ be an index function for P. A function

$$\nu : P \times P \to \mathbb{Z}$$

is called a **connection matrix** for $(P, \preceq, \mu)$ if it satisfies (88) and

$$\nu(q, p) \neq 0 \qquad \Longrightarrow \qquad \mu(q) - \mu(p) = 1, \;\; p \preceq q \tag{95}$$

for $p, q \in P$.

Lemma C.3. *If $\mu : P \to \mathbb{Z}$ is an index function for $(P, \preceq)$ then $\mu' := \mu|_{P'}$ is an index function for $(P', \preceq')$. Moreover, if ν is a connection matrix for $(P, \preceq, \mu)$ and $\nu(\bar{q}, \bar{p}) = 1$ then ν' is a connection matrix for $(P', \preceq', \mu')$.*

Proof. We prove that μ' is an index function for $(P', \preceq')$. Let $p', q' \in P'$ such that $p \preceq' q$. If $p \preceq q$ then $\mu(p) < \mu(q)$, since μ is an index function for $(P, \preceq)$. If $p \not\preceq q$ then $p \preceq \bar{q}$ and $\bar{p} \preceq q$, and hence

$$\mu(p) < \mu(\bar{q}) = \mu(\bar{p}) + 1 \leq \mu(q).$$

Hence μ' satisfies (94), as claimed. Next we prove that ν' is a connection matrix for $(P', \preceq', \mu')$. By Lemma C.1, ν' satisfies (88). We prove that it satisfies (95). Let $p, q \in P'$ such that $\nu'(q, p) \neq 0$. If $\nu(q, p) \neq 0$ then, since ν is a connection matrix for $(P, \preceq, \mu)$, we have $\mu(q) - \mu(p) = 1$ and $p \preceq' q$. If $\nu(q, p) = 0$ then in follows from the definition of ν' that $\nu(q, \bar{p}) \neq 0$ and $\nu(\bar{q}, p) \neq 0$. Hence

$$\mu(q) - \mu(\bar{p}) = 1, \qquad \mu(\bar{q}) - \mu(p) = 1, \qquad \mu(\bar{q}) - \mu(\bar{p}) = 1,$$

and hence

$$\bar{p} \preceq q, \qquad p \preceq \bar{q}.$$

It follows again that $\mu(q) - \mu(p) = 1$ and $p \preceq' q$. Hence ν' satisfies (95), as claimed. This proves Lemma C.3. □

D Asymptotic behavior of holomorphic strips

This appendix deals with the asymptotic behaviour of pseudoholomorphic strips in symplectic manifolds that satisfy Lagrangian boundary conditions. More precisely, let (M, ω) be a symplectic manifold and

$$L_0, L_1 \subset M$$

be closed (not necessarily compact) Lagrangian submanifolds that intersect transversally. Fix a t-dependent family of ω-tame almost complex structures J_t on M. We consider smooth maps $v : \mathbb{S} \to M$ that satisfy the boundary value problem

$$\partial_s v + J_t(v)\partial_t v = 0, \qquad v(\mathbb{R}) \subset L_0, \qquad v(\mathbb{R} + \mathbf{i}) \subset L_1. \tag{96}$$

Theorem D.1. *Let $v : \mathbb{S} \to M$ be a solution of (96). Assume v has finite energy*

$$E(v) := \int_{\mathbb{S}} v^*\omega < \infty$$

and that the image of v has compact closure. Then the following hold.

(i) *There exist intersection points $x, y \in L_0 \cap L_1$ such that*

$$x = \lim_{s\to-\infty} v(s,t), \qquad y = \lim_{s\to+\infty} v(s,t), \tag{97}$$

where the convergence is uniform in t. Moreover, $\partial_s v$ decays exponentially in the C^∞ topology, i.e. there are positive constants δ and $c_1, c_2, c_3, \ldots$ such that $\|\partial_s u\|_{C^k([s-1,s+1]\times[0,1])} \le c_k e^{-\delta|s|}$ for all s and k.

(ii) *Assume v is nonconstant. Then there exist positive real numbers λ_x, λ_y and smooth paths $\eta_x : [0,1] \to T_xM$ and $\eta_y : [0,1] \to T_yM$ satisfying*

$$\begin{aligned} &J_t(x)\partial_t\eta_x(t) + \lambda_x\eta_x(t) = 0, \qquad \eta_x(0) \in T_xL_0, \qquad \eta_x(1) \in T_xL_1, \\ &J_t(y)\partial_t\eta_y(t) - \lambda_y\eta_y(t) = 0, \qquad \eta_y(0) \in T_yL_0, \qquad \eta_y(1) \in T_yL_1, \end{aligned} \tag{98}$$

and

$$\eta_x(t) = \lim_{s\to-\infty} e^{-\lambda_x s}\partial_s v(s,t), \qquad \eta_y(t) = \lim_{s\to+\infty} e^{\lambda_y s}\partial_s v(s,t), \tag{99}$$

where the convergence is uniform in t.

(iii) *Assume v is nonconstant and let $\lambda_x, \lambda_y > 0$ and $\eta_x(t) \in T_xM$ and $\eta_y(t) \in T_yM$ be as in (ii). Then there exists a constant $\delta > 0$ such that*

$$\begin{aligned} &v(s,t) = \exp_x\left(\frac{1}{\lambda_x}e^{\lambda_x s}\eta_x(t) + R_x(s,t)\right), \\ &\lim_{s\to-\infty} e^{-(\lambda_x+\delta)s}\sup_t|R_x(s,t)| = 0, \end{aligned} \tag{100}$$

and

$$\begin{aligned} &v(s,t) = \exp_y\left(-\frac{1}{\lambda_y}e^{-\lambda_y s}\eta_y(t) + R_y(s,t)\right), \\ &\lim_{s\to+\infty} e^{(\lambda_y+\delta)s}\sup_t|R_y(s,t)| = 0. \end{aligned} \tag{101}$$

As a warmup the reader is encouraged to verify the signs in (ii) and (iii) in the linear setting where M is a symplectic vector space, L_0, L_1 are transverse Lagrangian subspaces, $J_t = J$ is a constant ω-tame complex structure, and $v(s,t) = \lambda_x^{-1}e^{\lambda_x s}\eta_x(t)$, respectively $v(s,t) = -\lambda_y^{-1}e^{-\lambda_y s}\eta_y(t)$.

Proof of Theorem D.1. Assertion (i) is standard (see e.g. [32, Theorem A]). Assertions (ii) and (iii) are proved in [32, Theorem B] for ω-compatible almost complex structures J_t. The ω-tame case is treated in [36]. $\square$

The next corollary summarizes the consequences of Theorem D.1 in the special case of dimension two. Assume Σ and $\alpha, \beta \subset \Sigma$ satisfy (H) and fix a complex structure J on Σ. For $x \in \alpha \cap \beta$ denote by $\theta_x \in (0, \pi)$ the angle from $T_x\alpha$ to $T_x\beta$ with respect to our complex structure J so that

$$T_x\beta = (\cos(\theta_x) + \sin(\theta_x)J)\, T_x\alpha, \qquad 0 < \theta_x < \pi.$$

(See equation (66).) Fix two intersection points $x, y \in \alpha \cap \beta$ and two holomorphic coordinate charts $\psi_x : U_x \to \mathbb{C}$ and $\psi_y : U_y \to \mathbb{C}$ defined on neighborhoods $U_x \subset \Sigma$ of x and $U_y \subset \Sigma$ of y such that $\psi_x(x) = 0$ and $\psi_y(y) = 0$.

Corollary D.2. *Let $v : \mathbb{S} \to \Sigma$ be a nonconstant holomorphic (α, β)-strip from x to y. Thus v is has finite energy and satisfies the boundary conditions $v(\mathbb{R}) \subset \alpha$ and $v(\mathbb{R} + \mathbf{i}) \subset \beta$ and the endpoint conditions*

$$\lim_{s\to-\infty} v(s + \mathbf{i}t) = x, \qquad \lim_{s\to+\infty} v(s + \mathbf{i}t) = y.$$

(See equation (62).) Then there exist nonzero complex numbers

$$c_x \in T_0\psi_x(U_x \cap \alpha), \qquad c_y \in T_0\psi_y(U_y \cap \alpha)$$

and integers $\nu_x \geq 0$ and $\nu_y \geq 1$ such that the following holds.

(i) *Define*

$$\lambda_x := \nu_x\pi + \theta_x, \qquad \lambda_y := \nu_y\pi - \theta_y. \tag{102}$$

Then

$$\begin{aligned} \lambda_x c_x &= \lim_{s\to-\infty} e^{-\lambda_x(s+\mathbf{i}t)} d(\psi_x \circ v)(s + \mathbf{i}t), \\ -\lambda_y c_y &= \lim_{s\to+\infty} e^{\lambda_y(s+\mathbf{i}t)} d(\psi_y \circ v)(s + \mathbf{i}t), \end{aligned} \tag{103}$$

where the convergence is uniform in t.

(ii) *Let $\lambda_x, \lambda_y > 0$ be given by (102). There exists a constant $\delta > 0$ such that*

$$\psi_x(v(s + \mathbf{i}t)) = c_x e^{\lambda_x(s+\mathbf{i}t)} + O(e^{(\lambda_x+\delta)s}) \tag{104}$$

for $s < 0$ close to $-\infty$ and

$$\psi_y(v(s + \mathbf{i}t)) = c_y e^{-\lambda_y(s+\mathbf{i}t)} + O(e^{-(\lambda_y+\delta)s}) \tag{105}$$

for $s > 0$ close to $+\infty$.

Proof. Define $\eta_x : [0,1] \to T_x\Sigma$ and $\eta_y : [0,1] \to T_y\Sigma$ by

$$d\psi_x(x)\eta_x(t) := \lambda_x c_x e^{\mathbf{i}\lambda_x t}, \qquad d\psi_y(y)\eta_y(t) := -\lambda_y c_y e^{-\mathbf{i}\lambda_y t}.$$

These functions satisfy the conditions in (98) if and only if $c_x \in T_0\psi_x(U_x \cap \alpha)$, $c_y \in T_0\psi_y(U_y \cap \alpha)$, and λ_x, λ_y are given by (102) with integers $\nu_x \geq 0$ and $\nu_y \geq 1$. Moreover, the limit condition (103) is equivalent to (99). Hence assertion (i) in Corollary D.2 follows from (ii) in Theorem D.1. With this understood, assertion (ii) in Corollary D.2 follows immediately from (iii) in Theorem D.1. This proves the corollary. (For assertion (ii) see also [32, Theorem C].) $\square$

References

[1] Mohammed Abouzaid, On the Fukaya Categories of Higher Genus Surfaces. *Advances in Mathematics* **217** (2008), 1192–1235. http://arxiv.org/pdf/math/0606598.pdf (Cited on pages 7, 105.)

[2] Mohammed Abouzaid, Denis Auroux, Alexander I. Efimov, Ludmil Katzarkov, Dmitri Orlov, Homological mirror symmetry for punctured spheres. Preprint, March 2011. http://arxiv.org/pdf/1103.4322.pdf (Cited on page 105.)

[3] Lars V. Ahlfors, *Complex Analysis*, 3rd edition, McGraw-Hill, 1979. (Cited on page 46.)

[4] Samuel J. Blank, Extending immersions of the circle, Dissertation, Brandise University, Waltham, Mass., 1967. (Cited on page 6.)

[5] Paul Biran, Octav Cornea, A Lagrangian quantum homology. *New Perspectives and Challenges in Symplectic Field Theory*, edited by Miguel Abreu, Francois Lalonde, Leonid Polterovich. CRM-AMS, Proceedings and Lecture Notes, Vol. **49**, 2009, pp. 1–44. (Cited on page 102.)

[6] Vin De Silva, *Products in the symplectic Floer homology of Lagrangian intersections*, PhD thesis, Oxford, 1999. (Cited on pages 7, 4, 104.)

[7] Clifford Earle, James Eells, A fibre bundle approach to Teichmüller theory, *J. Differential Geometry*, **3** (1969) 19–43. (Cited on page 110.)

[8] David Epstein, Curves on 2-manifolds and isotopies, *Acta Math.* **115** (1966), 83–107. (Cited on pages 30, 50, 108.)

[9] David Epstein, private communication, 6 April 2000. (Cited on pages 66, 106.)

[10] Andreas Floer, The unregularized gradient flow of the symplectic action, *Comm. Pure Appl. Math.* **41** (1988), 775–813. (Cited on pages 4, 100.)

[11] Andreas Floer, Morse theory for Lagrangian intersections, *J. Diff. Geom.* **28** (1988), 513–547. (Cited on page 4, 100, 111.)

[12] George K. Francis, Spherical curves that bound immersed discs, *Proceedings of the A.M.S.* **41** (1973), 87–93. (Cited on page 6.)

[13] Kenji Fukaya, Yong-Geun Oh, Hiroshi Ohta, Kaoru Ono, *Lagrangian Intersection Floer Theory: Anomaly and Obstruction*. AMS/IP Studies in Advanced Mathematics, vol. **46**, American Mathematical Society, Providence, RI, 2009. MR2310372 (2008d:53033) (Cited on page 103, 104.)

[14] Ralf Gautschi, Joel Robbin, Dietmar Salamon, Heegaard splittings and Morse-Smale flows. *International Journal of Mathematics and Mathematical Sciences* **2003**, 3539–3572. (Cited on page 110.)

[15] Michael Hutchings, Y. Lee, Circle valued Morse theory and Reidemeister torsion. *Geometry and Topology* **38** (1999), 369–396. (Cited on page 103.)

[16] Michael Hutchings, Y. Lee, Circle valued Morse theory, Reidemeister torsion, and Seiberg–Witten invariants of 3-manifolds. *Topology* **38** (1999), 861–888. (Cited on page 103.)

[17] Mikhael Khanevsky, Hofer's metric on the space of diameters, *Journal of Topology and Analysis* **1** (2009), 407–416. (Cited on page 102.)

[18] Maxim Kontsevich, Homological algebra of mirror symmetry. ICM Lecture, Zürich 1994. (Cited on page 104.)

[19] John W. Milnor, *Topology from the Differentiable Viewpoint*. The University Press of Virginia, 1969. (Cited on page 11.)

[20] Dusa McDuff, Dietmar Salamon, *J-holomorphic Curves and Symplectic Topology*, AMS Colloquium Publications, Volume **52**, 2004. (Cited on page 97.)

[21] Yong-Geun Oh, Floer cohomology of Lagrangian intersections and pseudoholomorphic discs I, II. *Comm Pure Appl. Math.*, **46** (1993), 949–94, 995–1012. III. Arnold–Givental conjecture. In *The Floer Memorial Volume*, edited by Hofer, Taubes, Weinstein, and Zehnder, Birkhäuser, Basel, 1995, pp. 555–574. (Cited on pages 4, 102.)

[22] Andrei Okounkov, Rahul Pandharipande, Gromov–Witten theory, Hurwitz numbers, and matrix models. Proceedings of Algebraic geometry - Seattle 2005, *Proc. Sympos. Pure Math.* **80**, (2005), 324-414. (Cited on page 103.)

[23] Andrei Okounkov, Rahul Pandharipande, Gromov–Witten theory, Hurwitz numbers, and completed cycles. *Annals of Mathematics* **163** (2006), 517–560. (Cited on page 103.)

[24] Andrei Okounkov, Rahul Pandharipande, The equivariant Gromov–Witten theory of $\mathbb{P}^1$. *Annals of Mathematics* **163** (2006), 561–605. (Cited on page 103.)

[25] Andrei Okounkov, Rahul Pandharipande, Virasoro constraints for link curves. *Inventiones Mathematicae* **163** (2006), 47–108. (Cited on page 103.)

[26] Peter Ozsvath, Zoltan Szabo, Holomorphic disks and topological invariants for closed three-manifolds. *Annals of Mathematics* **159** (2004), 1027–1158.

[27] Peter Ozsvath, Zoltan Szabo, Holomorphic disks and three-manifold invariants: properties and applications. *Annals of Mathematics* **159** (2004), 1059–1245.

[28] Valentin Poenaru, Extensions des immersions en codimension 1 (d'apres Samuel Blank), Sem. Bourbaki, 1967/68, Exposé 342. (Cited on page 6.)

[29] Marcin Pozniak, Floer homology, Novikov rings, and clean intersections, PhD thesis, University of Warwick, 1994. (Cited on page 71.)

[30] Joel Robbin, Dietmar Salamon, The Maslov index for paths, *Topology* **32** (1993), 827–844. (Cited on pages 14, 14, 14, 42, 63, 90, 96.)

[31] Joel Robbin, Dietmar Salamon, The spectral flow and the Maslov index *Bulletin of the L.M.S.* **27** (1995), 1–33. (Cited on pages 92, 95.)

[32] Joel Robbin, Dietmar Salamon, Asymptotic behaviour of holomorphic strips, *Annales de l'Institute Poincaré - Analyse Nonlinéaire* **18** (2001), 573–612. (Cited on pages 89, 91, 91, 91, 93, 118, 119.)

[33] Paul Seidel, Graded Lagrangian submanifolds, *Bull. Soc. math. France* **128** (2000), 103–149. (Cited on page 73.)

[34] Paul Seidel, *Fukaya categories and Picard-Lefschetz theory*. Zurich Lectures in Advanced Mathematics, European Mathematical Society (EMS), Zürich, 2008. MR2441780 (Cited on page 92, 104, 104.)

[35] Paul Seidel, Homological mirror symmetry for a genus two curve. Preprint 2008, last revised August 2011. http://arxiv.org/pdf/0812.1171.pdf (Cited on page 105.)

[36] Tatjana Simcevic, A Hardy space approach to Floer gluing. PhD thesis, ETH Zürich, in preparation. (Cited on page 118.)

[37] Steve Smale, Diffeomorphisms of the 2-sphere, *Proc. Amer. Math. Society*, **10** (1959) 621–626. (Cited on page 110.)

[38] C.J. Titus, The combinatorial topology of analytic functions on the boundary of a disc, *Acta Math.* **106** (1961), 45–64. (Cited on page 6.)

[39] Claude Viterbo, Intersections de sous-variétés Lagrangiennes, fonctionelles d'action et indice des systèmes Hamiltoniens, *Bull. Soc. Math. France* **115** (1987), 361–390. (Cited on pages 14, 14, 42, 42, 63, 90.)

Index