

Journal de l'École polytechnique

Mathématiques

Kai CIELIEBAK, Tobias EKHOLM, Janko LATSCHEV, & Lenhard NG

Knot contact homology, string topology, and the cord algebra

Tome 4 (2017), p. 661-780.

http://jep.cedram.org/item?id=JEP_2017__4__661_0

© Les auteurs, 2017.

Certains droits réservés.



Cet article est mis à disposition selon les termes de la licence
CREATIVE COMMONS ATTRIBUTION – PAS DE MODIFICATION 3.0 FRANCE.
<http://creativecommons.org/licenses/by-nd/3.0/fr/>

L'accès aux articles de la revue « Journal de l'École polytechnique — Mathématiques » (<http://jep.cedram.org/>), implique l'accord avec les conditions générales d'utilisation (<http://jep.cedram.org/legal/>).

Publié avec le soutien
du Centre National de la Recherche Scientifique

cedram

Article mis en ligne dans le cadre du
Centre de diffusion des revues académiques de mathématiques
<http://www.cedram.org/>

KNOT CONTACT HOMOLOGY, STRING TOPOLOGY, AND THE CORD ALGEBRA

BY KAI CIELIEBAK, TOBIAS EKHOLOM, JANKO LATSCHEV
& LENHARD NG

ABSTRACT. — The conormal Lagrangian L_K of a knot K in \mathbb{R}^3 is the submanifold of the cotangent bundle $T^*\mathbb{R}^3$ consisting of covectors along K that annihilate tangent vectors to K . By intersecting with the unit cotangent bundle $S^*\mathbb{R}^3$, one obtains the unit conormal ν_K , and the Legendrian contact homology of ν_K is a knot invariant of K , known as knot contact homology. We define a version of string topology for strings in $\mathbb{R}^3 \setminus L_K$ and prove that this is isomorphic in degree 0 to knot contact homology. The string topology perspective gives a topological derivation of the cord algebra (also isomorphic to degree 0 knot contact homology) and relates it to the knot group. Together with the isomorphism this gives a new proof that knot contact homology detects the unknot. Our techniques involve a detailed analysis of certain moduli spaces of holomorphic disks in $T^*\mathbb{R}^3$ with boundary on $\mathbb{R}^3 \setminus L_K$.

RÉSUMÉ (Homologie de contact pour les nœuds, topologie des cordes et algèbre des cordes)

Le fibré conormal lagrangien L_K d'un nœud K dans \mathbb{R}^3 est la sous-variété du fibré cotangent $T^*\mathbb{R}^3$ formée des covecteurs le long de K qui annulent les vecteurs tangents à K . En l'intersectant avec le fibré cotangent unitaire $S^*\mathbb{R}^3$, on obtient le fibré conormal unitaire ν_K , dont l'homologie de contact legendrienne est un invariant du nœud K , appelé homologie de contact pour les nœuds. Nous définissons une version de la topologie des cordes pour des cordes dans $\mathbb{R}^3 \setminus L_K$ et montrons qu'elle est isomorphe en degré 0 à l'homologie de contact pour les nœuds. La topologie des cordes permet une approche topologique de l'algèbre des cordes (qui est aussi isomorphe à l'homologie de contact pour les nœuds en degré 0) et la relie au groupe du nœud. Ceci donne, joint à cet isomorphisme, une nouvelle démonstration du fait que l'homologie de contact pour les nœuds détecte le nœud trivial. Nos techniques font intervenir une analyse détaillée de certains espaces de modules de disques holomorphes dans $T^*\mathbb{R}^3$ avec bord dans $\mathbb{R}^3 \setminus L_K$.

CONTENTS

1. Introduction.....	662
2. String homology in degree zero.....	665
3. Roadmap to the proof of Theorem 1.2.....	681
4. Holomorphic functions near corners.....	682

MATHEMATICAL SUBJECT CLASSIFICATION (2010). — 53D42, 55P50, 57R17, 57M27.

KEYWORDS. — Holomorphic curve, string topology, conormal bundle, knot invariant, Lagrangian submanifold, Legendrian submanifold.

The work of KC was supported by DFG grants CI 45/2-1 and CI 45/5-1. The work of TE was supported by the Knut and Alice Wallenberg Foundation and by the Swedish Research Council. The work of JL was supported by DFG grant LA 2448/2-1. The work of LN was supported by NSF grant DMS-1406371 and a grant from the Simons Foundation (# 341289 to Lenhard Ng).

5. String homology in arbitrary degree.....	687
6. The chain map from Legendrian contact homology to string homology.....	699
7. Proof of the isomorphism in degree zero.....	719
8. Properties of holomorphic disks.....	738
9. Transversely cut out solutions and orientations.....	750
10. Compactification of moduli spaces and gluing.....	762
References.....	779

1. INTRODUCTION

To a smooth n -manifold Q we can naturally associate a symplectic manifold and a contact manifold: its cotangent bundle T^*Q with the canonical symplectic structure $\omega = dp \wedge dq$, and its unit cotangent bundle (with respect to any Riemannian metric) $S^*Q \subset T^*Q$ with its canonical contact structure $\xi = \ker(pdq)$. Moreover, a k -dimensional submanifold $K \subset Q$ naturally gives rise to a Lagrangian and a Legendrian submanifold in T^*Q resp. S^*Q : its conormal bundle

$$L_K = \{(q, p) \in T^*Q \mid q \in K, p|_{T_q K} = 0\}$$

and its unit conormal bundle $\Lambda_K = L_K \cap S^*Q$. Symplectic field theory (SFT [20]) provides a general framework for associating algebraic invariants to a pair (M, Λ) consisting of a contact manifold and a Legendrian submanifold; when applied to (S^*Q, Λ_K) , these invariants will be diffeotopy invariants of the manifold pair (Q, K) . The study of the resulting invariants was first suggested by Y. Eliashberg.

In this paper we concentrate on the case where K is a framed oriented knot in $Q = \mathbb{R}^3$. Moreover, we consider only the simplest SFT invariant: *Legendrian contact homology*. For $Q = \mathbb{R}^3$, S^*Q is contactomorphic to the 1-jet space $J^1(S^2)$, for which Legendrian contact homology has been rigorously defined in [15]. The Legendrian contact homology of the pair $(S^*\mathbb{R}^3, \Lambda_K)$ is called the *knot contact homology* of K . We will denote it $H^{\text{contact}}(K)$.

In its most general form (see [11, 33]), knot contact homology is the homology of a differential graded algebra over the group ring $Z[H_2(S^*\mathbb{R}^3, \Lambda_K)] = Z[\lambda^{-1}, \mu^{-1}, U^{-1}]$, where the images of λ, μ under the connecting homomorphism generate $H_1(\Lambda_K) = H_1(T^2)$ and U generates $H_2(S^*\mathbb{R}^3)$. The isomorphism class of $H^{\text{contact}}(K)$ as a $Z[\lambda^{-1}, \mu^{-1}, U^{-1}]$ -algebra is then an isotopy invariant of the framed oriented knot K .

The topological content of knot contact homology has been much studied in recent years; see for instance [1] for a conjectured relation, which we will not discuss here, to colored HOMFLY-PT polynomials and topological strings. One part of knot contact homology that has an established topological interpretation is its $U = 1$ specialization. In [31, 32], the third author constructed a knot invariant called the *cord algebra* $\text{Cord}(K)$, whose definition we will review in Section 2.2. The combined results of [31, 32, 12] then prove that the cord algebra is isomorphic as a $Z[\lambda^{-1}, \mu^{-1}]$ -algebra

to the $U = 1$ specialization of degree 0 knot contact homology. We will assume throughout this paper that we have set $U = 1$;⁽¹⁾ then the result is:

THEOREM 1.1 ([31, 32, 12]). — $H_0^{\text{contact}}(K) = \text{Cord}(K)$.

It has been noticed by many people that the definition of the cord algebra bears a striking resemblance to certain operations in string topology [4, 36]. Indeed, Basu, McGibbon, Sullivan, and Sullivan used this observation in [2] to construct a theory called “transverse string topology” associated to any codimension 2 knot $K \subset Q$, and proved that it determines the $U = \lambda = 1$ specialization of the cord algebra.

In this paper, we present a different approach to knot contact homology and the cord algebra via string topology. Motivated by the general picture sketched by the first and third authors in [6], we use string topology operations to define the *string homology* $H^{\text{string}}(K)$ of K . Then the main result of this paper is:

THEOREM 1.2. — *For any framed oriented knot $K \subset \mathbb{R}^3$, we have an isomorphism between $U = 1$ knot contact homology and string homology in degree 0,*

$$H_0^{\text{contact}}(K) = H_0^{\text{string}}(K),$$

defined by a count of punctured holomorphic disks in $T \mathbb{R}^3$ with Lagrangian boundary condition $L_K \subset \mathbb{R}^3$.

On the other hand, degree 0 string homology is easily related to the cord algebra:

PROPOSITION 1.3. — *For any framed oriented knot $K \subset \mathbb{R}^3$, we have an isomorphism*

$$H_0^{\text{string}}(K) = \text{Cord}(K).$$

As a corollary we obtain a new geometric proof of Theorem 1.1. In fact, we even prove a slight refinement of the usual formulation of Theorem 1.1, as we relate certain noncommutative versions of the two sides where the coefficients λ, μ do not commute with everything; see Section 2.2 for the version of $\text{Cord}(K)$ and Section 6.2 for the definition of $H_0^{\text{contact}}(K)$ that we use.

Our proof is considerably more direct than the original proof of Theorem 1.1, which was rather circuitous and went as follows. The third author constructed in [30, 32] a combinatorial differential graded algebra associated to a braid whose closure is K , and then proved in [31, 32] that the degree 0 homology of this combinatorial complex is isomorphic to $\text{Cord}(K)$ via a mapping class group argument. The second and third authors, in joint work with Etnyre and Sullivan [12], then proved that the combinatorial complex is equal to the differential graded algebra for knot contact homology, using an analysis of degenerations of holomorphic disks to Morse flow trees.

Besides providing a cleaner proof of Theorem 1.1, the string topology formulation also gives a geometric explanation for the somewhat mystifying skein relations that

⁽¹⁾However, we note that it is an interesting open problem to find a similar topological interpretation of the full degree 0 knot contact homology as a $\mathbb{Z}[\lambda^{-1}, \mu^{-1}, U^{-1}]$ -algebra.

define the cord algebra. Moreover, string homology can be directly related to the group ring $Z\pi$ of the fundamental group $\pi = \pi_1(\mathbb{R}^3 \setminus K)$ of the knot complement:

PROPOSITION 1.4 (see Proposition 2.21). — *For a framed oriented knot $K \subset \mathbb{R}^3$, $H_0^{\text{string}}(K) = H_0^{\text{contact}}(K)$ is isomorphic to the subring of $Z\pi$ generated by λ^{-1} , μ^{-1} , and $\text{im}(1 - \mu)$, where λ, μ are the elements of π representing the longitude and meridian of K , and $1 - \mu$ denotes the map $Z\pi \rightarrow Z\pi$ given by left multiplication by $1 - \mu$.*

As an easy consequence of Proposition 1.4, we recover the following result from [32]:

COROLLARY 1.5 (see Section 2.4). — *Knot contact homology detects the unknot: if $H_0^{\text{contact}}(K) = H_0^{\text{contact}}(U)$ where K is a framed oriented knot in \mathbb{R}^3 and U is the unknot with any framing, then $K = U$ as framed oriented knots.*

The original proof of Corollary 1.5 in [32] uses the result that the A -polynomial detects the unknot [8], which in turn relies on results from gauge theory [28]. By contrast, our proof of Corollary 1.5 uses no technology beyond the Loop Theorem (more precisely, the consequence of the Loop Theorem that the longitude is null-homotopic in $\mathbb{R}^3 \setminus K$ if and only if K is unknotted).

Organization of the paper. — In Section 2 we define degree 0 string homology and prove Proposition 1.3, Proposition 1.4 and Corollary 1.5. The remainder of the paper is occupied by the proof of Theorem 1.2, beginning with an outline in Section 3. After a digression in Section 4 on the local behavior of holomorphic functions near corners, which serves as a model for the behavior of broken strings at switches, we define string homology in arbitrary degrees in Section 5.

The main work in proving Theorem 1.2 is an explicit description of the moduli spaces of holomorphic disks in $T\mathbb{R}^3$ with boundary on $L_K \subset \mathbb{R}^3$ and punctures asymptotic to Reeb chords. In Section 6 we state the main results about these moduli spaces and show how they give rise to a chain map from Legendrian contact homology to string homology (in arbitrary degrees). Moreover, we show that this chain map respects a natural length filtration. In Section 7 we construct a length decreasing chain homotopy and prove Theorem 1.2.

The technical results about moduli spaces of holomorphic disks and their compactifications as manifolds with corners are proved in the remaining Sections 8, 9 and 10.

Extensions. — The constructions in this paper have several possible extensions. Firstly, the definition of string homology and the construction of a homomorphism from Legendrian contact homology to string homology in degree zero work the same way for a knot K in an arbitrary 3-manifold Q instead of \mathbb{R}^3 (the corresponding sections are actually written in this more general setting), and more generally for a codimension 2 submanifold K of an arbitrary manifold Q .⁽²⁾ The fact that the

⁽²⁾In the presence of contractible closed geodesics in Q , this will require augmentations by holomorphic planes in TQ , see e.g. [6].

ambient manifold is \mathbb{R}^3 is only used to obtain a certain finiteness result in the proof that this map is an isomorphism (see Remark 7.9). If this result can be generalized, then Theorem 1.2 will hold for arbitrary codimension 2 submanifolds $K \subset Q$.

Secondly, for knots in 3-manifolds, the homomorphism from Legendrian contact homology to string homology is actually constructed in arbitrary degrees. Proving that it is an isomorphism in arbitrary degrees will require analyzing codimension three phenomena in the space of strings with ends on the knot, in addition to the codimension one and two phenomena described in this paper.

Acknowledgments. — We thank Chris Cornwell, Tye Lidman, and especially Yasha Eliashberg for stimulating conversations. This project started when the authors met at the Workshop “SFT 2” in Leipzig in August 2006, and the final technical details were cleaned up when we met during the special program on “Symplectic geometry and topology” at the Mittag-Leffler institute in Djursholm in the fall of 2015. We would like to thank the sponsors of these programs for the opportunities to meet, as well as for the inspiring working conditions during these events. Finally, we thank the referee for suggesting numerous improvements.

2. STRING HOMOLOGY IN DEGREE ZERO

In this section, we introduce the degree 0 string homology $H_0^{\text{string}}(K)$. The discussion of string homology here is only a first approximation to the more precise approach in Section 5, but is much less technical and suffices for the comparison to the cord algebra. We then give several formulations of the cord algebra $\text{Cord}(K)$ and use these to prove that $H_0^{\text{string}}(K) = \text{Cord}(K)$ and that string homology detects the unknot. Throughout this section, K denotes an oriented framed knot in some oriented 3-manifold Q .

2.1. A STRING TOPOLOGY CONSTRUCTION. — Here we define $H_0^{\text{string}}(K)$ for an oriented knot $K \subset Q$. Let N be a tubular neighborhood of K . For this definition we do not need a framing for the knot K ; later, when we identify $H_0^{\text{string}}(K)$ with the cord algebra, it will be convenient to fix a framing, which will in turn fix an identification of N with $S^1 \times D^2$.

Any tangent vector v to Q at a point on K has a tangential component parallel to K and a normal component lying in the disk fiber; write v^{normal} for the normal component of v . Fix a base point $x_0 \in \partial N$ and a unit tangent vector $v_0 \in T_{x_0} \partial N$.

DEFINITION 2.1. — A *broken (closed) string with 2ℓ switches on K* is a tuple $s = (a_1, \dots, a_{2\ell+1}; s_1, \dots, s_{2\ell+1})$ consisting of real numbers $0 = a_0 < a_1 < \dots < a_{2\ell+1}$ and C^1 maps

$$s_{2i+1} : [a_{2i}, a_{2i+1}] \rightarrow N, \quad s_{2i} : [a_{2i-1}, a_{2i}] \rightarrow Q$$

satisfying the following conditions:

- (i) $s_0(0) = s_{2\ell+1}(a_{2\ell+1}) = x_0$ and $\dot{s}_0(0) = \dot{s}_{2\ell+1}(a_{2\ell+1}) = v_0$;
- (ii) for $j = 1, \dots, 2\ell$, $s_j(a_j) = s_{j+1}(a_j) \in K$;

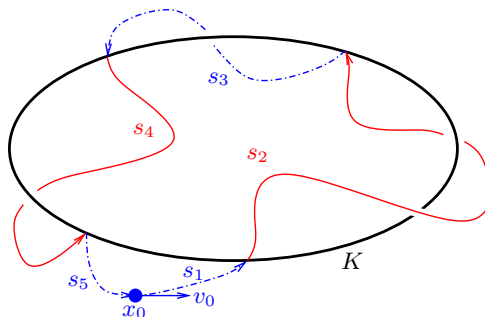


FIGURE 2.1. A broken closed string with 4 switches. Here, as in subsequent figures, we draw the knot K in black, Q -strings (s_2, s_4) in red, and N -strings (s_1, s_3, s_5) in blue (dashed for clarity to distinguish from the red Q -strings).

(iii) for $i = 1, \dots, \ell$,

$$\begin{aligned} (\dot{s}_{2i}(a_{2i}))^{\text{normal}} &= (\dot{s}_{2i+1}(a_{2i}))^{\text{normal}} \\ (\dot{s}_{2i-1}(a_{2i-1}))^{\text{normal}} &= (\dot{s}_{2i}(a_{2i-1}))^{\text{normal}}. \end{aligned}$$

We will refer to the s_{2i} and s_{2i+1} as Q -strings and N -strings, respectively. Denote by Σ^ℓ the set of broken strings with 2ℓ switches.

The last condition, involving normal components of the tangent vectors to the ends of the Q - and N -strings, models the boundary behavior of holomorphic disks in this context (see Sections 4.1 and 5.1 for more on this point). A typical picture of a broken string is shown in Figure 2.1.

We call a broken string $s = (s_1, \dots, s_{2\ell+1})$ *generic* if none of the derivatives $\dot{s}_i(a_{i-1}), \dot{s}_i(a_i)$ is tangent to K and no s_i intersects K away from its end points. We call a smooth 1-parameter family of broken strings $s^\lambda = (s_1^\lambda, \dots, s_{2\ell+1}^\lambda), \lambda \in [0, 1]$, *generic* if s^0 and s^1 are generic strings, none of the derivatives $\dot{s}_i^\lambda(a_{i-1}^\lambda), \dot{s}_i^\lambda(a_i^\lambda)$ is tangent to K , and for each i the family s_i^λ intersects K transversally in the interior. The boundary of this family is given by

$$\partial f_{s^\lambda} g := s^1 - s^0.$$

We define *string coproducts* δ_Q and δ_N as follows, cf. Section 5.3. Fix a family of bump functions (which we will call *spikes*) $s_\nu : [0, 1] \rightarrow D^2$ for $\nu \in D^2$ such that $s_\nu^{-1}(0) = \{0, 1\}$, $\dot{s}_\nu(0) = \nu$ and $\dot{s}_\nu(1) = -\nu$; for each ν , s_ν lies in the line joining 0 to ν . For a generic 1-parameter family of broken strings $f_{s^\lambda} g$ denote by λ^j, b^j the finitely many values for which $s_{2i}^{\lambda^j}(b^j) \in K$ for some $i = i(j)$. For each j , let $s^j = s_{\nu^j}(\quad b^j) : [b^j, b^j + 1] \rightarrow N$ be a shift of the spike associated to the normal derivative $\nu^j := (\dot{\sigma}_{2i}^{\lambda^j}(b_j))^{\text{normal}}$, with constant value $s_{2i}^{\lambda^j}(b^j)$ along K ; interpret this as an N -string in the normal disk to K at the point $s_{2i}^{\lambda^j}(b^j)$, traveling along the line

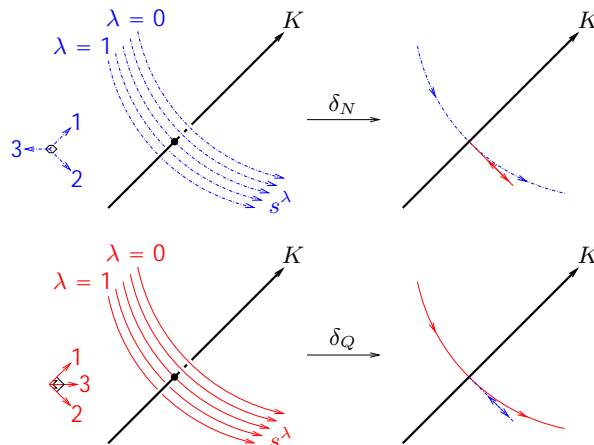


FIGURE 2.2. The definition of δ_N and δ_Q . The two configurations shown have sign $\varepsilon = 1$. If the orientation of the 1-parameter family s^λ is switched, i.e., the $\lambda = 0$ and $\lambda = 1$ ends are interchanged, then δ_N and δ_Q are still as shown, but with sign $\varepsilon = -1$. The coordinate axes denote orientations chosen on N (top) and Q (bottom).

joining 0 to $\nu^j \supset D^2$. Now set

$$\delta_Q f s^\lambda g := \sum_j \varepsilon^j \left(s_1^{\lambda^j}, \dots, s_{2i}^{\lambda^j} j_{[a_{2i-1}, b^j]}, S^j, \widehat{s}_{2i}^{\lambda^j} j_{[b^j, a_{2i}]}, \dots, \widehat{s}_{2\ell+1}^{\lambda^j} \right),$$

where the hat means shift by 1 in the argument, and $\varepsilon^j = \pm 1$ are signs defined as in Figure 2.2.⁽³⁾ Loosely speaking, δ_Q inserts an N -spike at all points where some Q -string meets K , in such a way that (iii) still holds. The operation δ_N is defined analogously, inserting a Q -spike where an N -string meets K (and defining ν^j without the minus sign).

Denote by $C_0(\Sigma^\ell)$ and $C_1(\Sigma^\ell)$ the free \mathbb{Z} -modules generated by generic broken strings and generic 1-parameter families of broken strings with 2ℓ switches, respectively, and set

$$C_i(\Sigma) := \bigoplus_{\ell=0}^1 C_i(\Sigma^\ell), \quad i = 0, 1.$$

Concatenation of broken strings at the base point gives $C_0(\Sigma)$ the structure of a (noncommutative but strictly associative) algebra over \mathbb{Z} . The operations defined

⁽³⁾Regarding the signs: from our considerations of orientation bundles in Section 9, we can assign the same sign (which we have chosen to be $\varepsilon = 1$) to both configurations shown in Figure 2.2, provided we choose orientations on Q and N appropriately. More precisely, at a point p on K , if (v_1, v_2, v_3) is a positively oriented frame in Q where v_1 is tangent to K and v_2, v_3 are normal to K , then we need $(v_1, Jv_2, -Jv_3)$ to be a positively oriented frame in N , where J is the almost complex structure that rotates normal directions in Q to normal directions in N . As a result, if we give Q any orientation and view N as the subset of Q given by a tubular neighborhood of K , then we assign the *opposite* orientation to N .

above yield linear maps

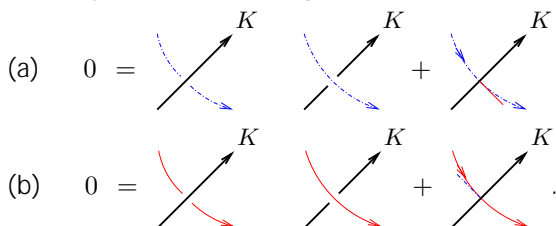
$$\partial : C_1(\Sigma^\ell) \rightarrow C_0(\Sigma^\ell) \rightarrow C_0(\Sigma), \quad \delta_N, \delta_Q : C_1(\Sigma^\ell) \rightarrow C_0(\Sigma^{\ell+1}) \rightarrow C_0(\Sigma).$$

Define the degree zero string homology of K as

$$H_0^{\text{string}}(K) = H_0(\Sigma) := C_0(\Sigma)/\text{im}(\partial + \delta_N + \delta_Q).$$

Since $\partial + \delta_N + \delta_Q$ commutes with multiplication by elements in $C_0(\Sigma)$, its image is a two-sided ideal in $C_0(\Sigma)$. Hence degree zero string homology inherits the structure of an algebra over \mathbb{Z} . By definition, $H_0^{\text{string}}(K)$ is an isotopy invariant of the oriented knot K (the framing was used only for convenience but is not really needed for the construction, cf. Remark 2.3 below).

Considering 1-parameter families consisting of generic strings (on which δ_N and δ_Q vanish), we see that for the computation of $H_0^{\text{string}}(K)$ we may replace the algebra $C_0(\Sigma)$ by its quotient under homotopy of generic strings. On the other hand, if $f_s^{\lambda g}$ is a generic 1-parameter family of strings that consists of generic strings except for an N -string (resp. a Q -string) that passes through K exactly once, then δ_N (resp. δ_Q) contributes a term to $(\partial + \delta_N + \delta_Q)$, and setting $(\partial + \delta_N + \delta_Q)(f_s^{\lambda g}) = 0$ in these two cases yields the following “skein relations”:



Since any generic 1-parameter family of broken closed strings can be divided into 1-parameter families each of which crosses K at most once, we have proved the following result.

PROPOSITION 2.2. — *Let B be the quotient of $C_0(\Sigma)$ by homotopy of generic broken strings and let $J \subset B$ be the two-sided ideal generated by the skein relations (a) and (b). Then*

$$H_0^{\text{string}}(K) = B/J .$$

REMARK 2.3. — Degree zero string homology H_0^{string} (as well as its higher degree version defined later) is an invariant of an oriented knot $K \rightarrow Q$. Reversing the orientation of K has the result of changing the signs of δ_N and δ_Q but not of ∂ and gives rise to isomorphic H_0^{string} . More precisely, if \overleftarrow{K} is K with the opposite orientation, the map $C_0(\Sigma) \rightarrow C_0(\Sigma)$ given by multiplication by $(-1)^\ell$ on the summand $C_0(\Sigma^\ell)$ intertwines the differentials $\partial + \delta_N + \delta_Q$ for K and \overleftarrow{K} and induces an isomorphism $H_0^{\text{string}}(K) \cong H_0^{\text{string}}(\overleftarrow{K})$. Similarly, mirroring does not change H_0^{string} up to isomorphism: if \overline{K} is the mirror of K , then the mirror (reflection) map induces a map $C_0(\Sigma) \rightarrow C_0(\overline{\Sigma})$, and composing with the above map $C_0(\Sigma) \rightarrow C_0(\Sigma)$ gives a chain isomorphism $C_0(\Sigma) \rightarrow C_0(\overline{\Sigma})$.

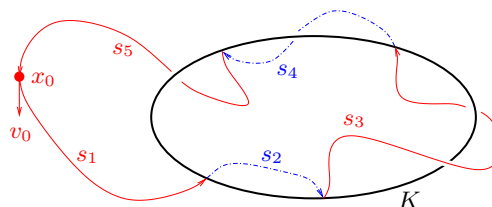


FIGURE 2.3. In the alternate definition that produces modified string homology, a broken closed string with 4 switches. As usual, Q -strings are in red, N -strings in (dashed) blue.

In Sections 2.2 through 2.4, we will “improve” H_0^{string} from an abstract ring to one that canonically contains the ring $\mathbb{Z}[\lambda^{-1}, \mu^{-1}]$. This requires a choice of framing of K (though for $Q = \mathbb{R}^3$, there is a canonical choice given by the Seifert framing). In the improved setting, H_0^{string} changes under orientation reversal of K by replacing (λ, μ) by (λ^{-1}, μ^{-1}) ; under framing change by $f \in \mathbb{Z}$ by replacing (λ, μ) by $(\lambda\mu^f, \mu)$; and under mirroring by replacing (λ, μ) by (λ, μ^{-1}) . In particular, the improved H_0^{string} is very sensitive to framing change and mirroring. For a related discussion, see [32, §4.1].

A modified version of string homology. — The choice of the base point in N rather than Q in the definition of string homology $H_0^{\text{string}}(K)$ is dictated by the relation to Legendrian contact homology. However, from the perspective of string topology we could equally well pick the base point in Q , as we describe next.

Choose a base point $x_0 \in Q \cap K$ and a tangent vector $v_0 \in T_{x_0}Q$. Modify the definition of a broken string with 2ℓ switches to $s = (a_0, \dots, a_{2\ell+1}; s_0, \dots, s_{2\ell})$, where now

$$s_{2i} : [a_{2i}, a_{2i+1}] \dashrightarrow Q, \quad s_{2i-1} : [a_{2i-1}, a_{2i}] \dashrightarrow N,$$

and we require that $s_0(a_0) = s_{2\ell}(a_{2\ell+1}) = x_0$, $\dot{s}_0(a_0) = \dot{s}_{2\ell}(a_{2\ell+1}) = v_0$ and conditions (ii) and (iii) of Definition 2.1 hold. See Figure 2.3.

Let $\widehat{C}_0(\Sigma)$ denote the ring generated as a \mathbb{Z} -module by generic broken strings with base point $x_0 \in Q$. (As usual, the product operation on $\widehat{C}_0(\Sigma)$ is given by string concatenation.) We can define string coproducts δ_N, δ_Q as before, and then define the degree 0 *modified string homology* of K as

$$\widehat{H}_0^{\text{string}}(K) = \widehat{C}_0(\Sigma) / \text{im}(\partial + \delta_N + \delta_Q).$$

We have the following analogue of Proposition 2.2.

PROPOSITION 2.4. — *Let \widehat{B} be the quotient of $\widehat{C}_0(\Sigma)$ by homotopy of generic broken strings and let $\widehat{J} \subset \widehat{B}$ be the two-sided ideal generated by the skein relations (a) and (b). Then*

$$\widehat{H}_0^{\text{string}}(K) = \widehat{B} / \widehat{J}.$$

There is one key difference between $\widehat{H}_0^{\text{string}}$ and H_0^{string} . Since any element in $\pi_1(Q \cap K, x_0)$ can be viewed as a pure Q -string, we have a canonical map

$Z\pi_1(Q \curvearrowright K, x_0) \cong \widehat{H}_0^{\text{string}}(K)$. In fact, we will see in Proposition 2.17 that this is a ring isomorphism. The same is not the case for $H_0^{\text{string}}(K)$.

2.2. THE CORD ALGEBRA. — The definition of $H_0^{\text{string}}(K)$ in Section 2.1 is very similar to the definition of the cord algebra of a knot [31, 32, 34]. Here we review the cord algebra, or more precisely, present a noncommutative refinement of it, in which the “coefficients” λ, μ do not commute with the “cords”.

Let $K \curvearrowright Q$ be an oriented knot equipped with a framing, and let K^θ be a parallel copy of K with respect to this framing. Choose a base point x_0 on K and a corresponding base point x_0^θ on K^θ (in fact only the base point on K^θ will be needed).

DEFINITION 2.5. — A (framed) cord of K is a continuous map $\gamma : [0, 1] \rightarrow Q$ such that $\gamma([0, 1]) \setminus K = \emptyset$ and $\gamma(0), \gamma(1) \in K^\theta \curvearrowright f \cdot g$. Two framed cords are homotopic if they are homotopic through framed cords.

We now construct a noncommutative unital ring A as follows: as a ring, A is freely generated by homotopy classes of cords and four extra generators λ^{-1}, μ^{-1} , modulo the relations

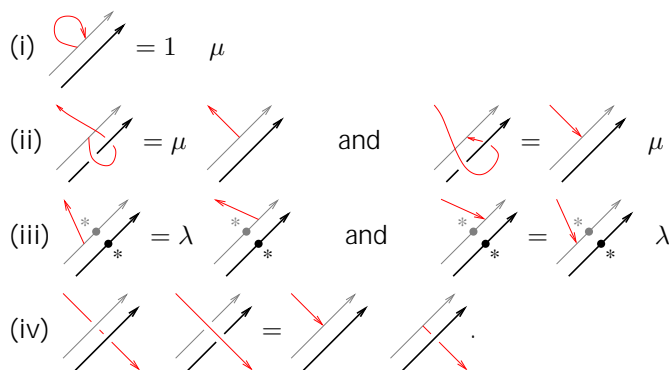
$$\lambda \lambda^{-1} = \lambda^{-1} \lambda = \mu \mu^{-1} = \mu^{-1} \mu = 1, \quad \lambda \mu = \mu \lambda.$$

Thus A is generated as a \mathbb{Z} -module by (noncommutative) words in homotopy classes of cords and powers of λ and μ (and the powers of λ and μ commute with each other, but not with any cords).

DEFINITION 2.6. — The cord algebra of K is the quotient ring

$$\text{Cord}(K) = A / I,$$

where I is the two-sided ideal of A generated by the following “skein relations”:



Here K is depicted in black and K^θ parallel to K in gray, and cords are drawn in red.

REMARK 2.7. — The skein relations in Definition 2.6 depict cords in space that agree outside of the drawn region (except in (iv), where either of the two cords on the left hand side of the equation splits into the two on the right). Thus (ii) states that

appending a meridian to the beginning or end of a cord multiplies that cord by μ on the left or right, and (iv) is equivalent to:

REMARK 2.8. — Our stipulation that λ, μ not commute with cords necessitates a different normalization of the cord algebra of $K \sqcup Q$ from previous definitions [32, 34]. In the definition from [34] ([32] is the same except for a change of variables), λ, μ commute with cords, and the parallel copy K^∂ is not used. Instead, cords are defined to be paths that begin and end on K with no interior point lying on K , and the skein relations are suitably adjusted, with the key relation, the equivalent of (iv), being:

Let $\text{Cord}^\partial(K)$ denote the resulting version of cord algebra.

If we take the quotient of the cord algebra $\text{Cord}(K)$ from Definition 2.6 where λ, μ commute with everything, then the result is a $\mathbb{Z}[\lambda^{-1}, \mu^{-1}]$ -algebra isomorphic to $\text{Cord}^\partial(K)$, as long as we take the Seifert framing ($\text{lk}(K, K^\partial) = 0$). The isomorphism is given as follows: given a framed cord γ , extend γ to an oriented closed loop $\tilde{\gamma}$ in $Q \sqcup K$ by joining the endpoints of γ along K^∂ in a way that does not pass through the base point, and map γ to $\mu^{-\text{lk}(\tilde{\gamma}, K)} \gamma$. This is a well-defined map on $\text{Cord}(K)$ and sends the relations for $\text{Cord}(K)$ to the relations for $\text{Cord}^\partial(K)$. See also the proof of Theorem 2.10 in [32].

We now show that the cord algebra is exactly equal to degree 0 string homology. This follows from the observation that the Q -strings in a generic broken closed string are each a framed cord of K , once we push the endpoints of the Q -string off of K ; and thus a broken closed string can be thought of as a product of framed cords.

PROPOSITION 2.9. — *Let $K \sqcup Q$ be a framed oriented knot. Then we have a ring isomorphism*

$$\text{Cord}(K) = H_0^{\text{string}}(K).$$

Proof. — Choose a normal vector field v along K defining the framing and let K^∂ be the pushoff of K in the direction of v , placed so that K^∂ lies on the boundary of the tubular neighborhood N of K . Fix a base point $p \notin K$, and let p^∂ be the corresponding point on K^∂ , so that $v(p)$ is mapped to p^∂ under the diffeomorphism between the normal bundle to K and N . Identify p^∂ with $x_0 \in \partial N$ from Definition 2.1 (the definition of broken closed string). We can homotope any cord of K so that it begins and ends at p^∂ , by pushing the endpoints of the cord along K^∂ , away from p , until they reach p^∂ .

Every generator of $\text{Cord}(K)$ as a \mathbb{Z} -module has the form $\alpha_1 x_1 \alpha_2 x_2 \dots x_\ell \alpha_{\ell+1}$, where $\ell > 0$, x_1, \dots, x_ℓ are cords of K , and $\alpha_1, \dots, \alpha_{\ell+1}$ are each of the form $\lambda^a \mu^b$ for $a, b \in \mathbb{Z}$. We can associate a broken closed string with 2ℓ switches as follows.

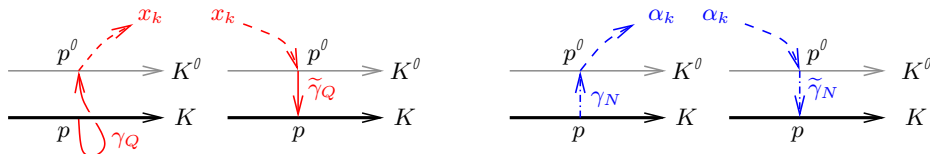


FIGURE 2.4. Turning an element of the cord algebra into a broken closed string.

Assume that each cord x_1, \dots, x_ℓ begins and ends at p^θ . Fix paths $\gamma_Q, \tilde{\gamma}_Q$ in Q from p, p^θ to p^θ, p respectively, and paths $\gamma_N, \tilde{\gamma}_N$ in N from p, p^θ to p^θ, p respectively, as shown in Figure 2.4: these are chosen so that the derivative of $\gamma_Q, \tilde{\gamma}_Q, \gamma_N, \tilde{\gamma}_N$ at p is $v(p), v(p), v(p), v(p)$, respectively. For $k = 1, \dots, \ell$, let \bar{x}_k be the Q -string with endpoints at p given by the concatenation $\gamma_Q x_k \tilde{\gamma}_Q$ (more precisely, smoothen this string at p^θ). Similarly, for $k = 1, \dots, \ell + 1$, identify $\alpha_k \in \pi_1(\partial N) = \pi_1(T^2)$ with a loop in ∂N with basepoint p^θ representing this class; then define $\bar{\alpha}_k$ to be the N -string $\gamma_N \alpha_k \tilde{\gamma}_N$ for $k = 1, \dots, \ell$, $\bar{\alpha}_0 = \gamma_N \alpha_1 \tilde{\gamma}_N$ for $k = 0$, and $\bar{\alpha}_{\ell+1} = \gamma_N \alpha_{\ell+1} \tilde{\gamma}_N$ for $k = \ell + 1$. (If $\ell = 0$, then $\bar{\alpha}_1 = \alpha_1$.) Then the concatenation

$$\bar{\alpha}_1 \bar{x}_1 \bar{\alpha}_2 \bar{x}_2 \dots \bar{x}_\ell \bar{\alpha}_{\ell+1}$$

is a broken closed string with 2ℓ switches.

Extend this map from generators of $\text{Cord}(K)$ to broken closed strings to a map on $\text{Cord}(K)$ by \mathbb{Z} -linearity. We claim that this induces the desired isomorphism $\phi : \text{Cord}(K) \rightarrow H_0^{\text{string}}(K)$. Recall that $\text{Cord}(K)$ is defined by skein relations (i), (ii), (iii), (iv) from Definition 2.6, while $H_0^{\text{string}}(K)$ is defined by skein relations (a), (b) from Proposition 2.2.

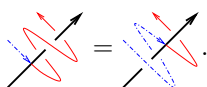
To check that ϕ is well-defined, we need for the skein relations (i), (ii), (iii), (iv) to be preserved by ϕ . Indeed, (i) maps under ϕ to



which holds in $H_0^{\text{string}}(K)$ since both sides are equal to blue loop on line : the left hand side by rotating the end of the red Q -string and the beginning of the blue N -string around K at their common endpoint, the right hand side by skein relation (a). Skein relation (iv) maps under ϕ to



which holds by (b). Finally, (ii) and (iii) map to homotopies of broken closed strings: for instance, the left hand relation in (ii) maps to



To show that ϕ is an isomorphism, we simply describe the inverse map from broken closed strings to the cord algebra. Given any broken closed string, homotope it so that the switches all lie at p , and so that the tangent vector to the endpoint of all strings ending at p is $v(p)$; then the result is in the image of ϕ by construction. There is more than one way to homotope a broken closed string into this form, but any such form gives the same element of the cord algebra: moving the switches along K to p in a different way gives the same result by (iii), while moving the tangent vectors to $v(p)$ in a different way gives the same result by (ii). The two skein relations (a) and (b) are satisfied in the cord algebra because of (i) and (iv).

As mentioned in the Introduction, when $Q = \mathbb{R}^3$, it is an immediate consequence of Theorem 1.2 and Proposition 2.9 that the cord algebra is isomorphic to degree 0 knot contact homology:

$$H_0^{\text{contact}}(K) = H_0^{\text{string}}(K) = \text{Cord}(K).$$

This recovers a result from the literature (see Theorem 1.1), modulo one important point. Recall (or see Section 6.2) that $H_0^{\text{contact}}(K)$ is the degree 0 homology of a differential graded algebra (A, ∂) . In much of the literature on knot contact homology, cf. [11, 32, 33], this DGA is an algebra over the coefficient ring $\mathbb{Z}[\lambda^{-1}, \mu^{-1}]$ (or $\mathbb{Z}[\lambda^{-1}, \mu^{-1}, U^{-1}]$, but in this paper we set $U = 1$): A is generated by a finite collection of noncommuting generators (Reeb chords) along with powers of λ, μ that commute with Reeb chords. By contrast, in this paper (A, ∂) is the *fully noncommutative* DGA in which the coefficients λ, μ commute with each other but not with the Reeb chords; see [12, 34].

The isomorphism $\text{Cord}(K) = H_0^{\text{contact}}(K)$ in Theorem 1.1 is stated in the existing literature as an isomorphism of $\mathbb{Z}[\lambda^{-1}, \mu^{-1}]$ -algebras, i.e., the coefficients λ, μ commute with everything for both $H_0^{\text{contact}}(K)$ and $\text{Cord}(K)$. However, an inspection of the proof of Theorem 1.1 from [12, 31, 32] shows that it can be lifted to the fully noncommutative setting, in which λ, μ do not commute with Reeb chords (for $H_0^{\text{contact}}(K)$) or cords (for $\text{Cord}(K)$). We omit the details here, and simply note that our results give a direct proof of Theorem 1.1 in the fully noncommutative setting.

REMARK 2.10. — Besides being more natural from the viewpoint of string homology, the stipulation that λ, μ do not commute with cords (in the cord algebra) or Reeb chords (in the DGA) is essential for our construction, in Section 2.4 below, of a map from degree 0 homology to the group ring of π , the fundamental group of the knot complement. This in turn is what allows us to (re)prove that knot contact homology detects the unknot, among other things. If we pass to the quotient where λ, μ commute with everything, then there is no well-defined map to $\mathbb{Z}\pi$.

REMARK 2.11. — As already mentioned in the introduction, in [2] Basu, McGibbon, Sullivan and Sullivan have given a string topology description of a version of the cord algebra for a codimension 2 submanifold $K \subset Q$ of some ambient manifold Q , proving a theorem which formally looks quite similar to Proposition 2.9. In the language we

use here, the main difference in their work is the absence of N -strings, so that for knots $K \subset \mathbb{R}^3$ the version of $H_0^{\text{string}}(K)$ they define only recovers the specialization at $\lambda = 1$ of (the commutative version of) $\text{Cord}(K)$.

2.3. HOMOTOPY FORMULATION OF THE CORD ALGEBRA. — We now reformulate the cord algebra in terms of fundamental groups, more precisely the knot group and its peripheral subgroup, along the lines of the Appendix to [31]. In light of Proposition 2.9, we will henceforth denote the cord algebra as $H_0^{\text{string}}(K)$.

We first introduce some notation. Let K be an oriented knot in an oriented 3-manifold Q (in fact we only need an orientation and coorientation of K). Let N be a tubular neighborhood of K ; as suggested by the notation, we will identify this neighborhood with the conormal bundle $N \cong T^*Q$ via the tubular neighborhood theorem. We write

$$\begin{aligned}\pi &= \pi_1(Q \setminus K) \\ \hat{\pi} &= \pi_1(\partial N);\end{aligned}$$

note that the inclusion $\partial N \hookrightarrow N$ induces a map $\hat{\pi} \rightarrow \pi$, typically an injection. Let $Z\pi$, $Z\hat{\pi}$ denote the group rings of $\pi, \hat{\pi}$. We fix a framing on K ; this, along with the orientation and coorientation of K , allows us to specify two elements μ, λ for $\hat{\pi}$ corresponding to the meridian and longitude, and to write

$$Z\hat{\pi} = Z[\lambda^{-1}, \mu^{-1}].$$

The group ring $Z\pi$ and the cord algebra $H_0^{\text{string}}(K)$ both have natural maps from $Z[\lambda^{-1}, \mu^{-1}]$ (which are injective unless K is the unknot). This motivates the following definition, where “NC” stands for “noncommutative”.

DEFINITION 2.12. — Let R be a ring. An R -NC-algebra is a ring S equipped with a ring homomorphism $R \rightarrow S$. Two R -NC-algebras S_1, S_2 are *isomorphic* if there is a ring isomorphism $S_1 \cong S_2$ that commutes with the maps $R \rightarrow S_1, R \rightarrow S_2$.

Note that when R is commutative, the notion of an R -NC-algebra differs from the usual notion of an R -algebra; for example, an R -algebra S requires $s_1(rs_2) = rs_1s_2$ for $r \in R$ and $s_1, s_2 \in S$, while an R -NC-algebra does not. (One can quotient an R -NC-algebra by commutators involving elements of R to obtain an R -algebra.) If R and S are both commutative, however, then the notions agree. Also note that any R -NC-algebra is automatically an R -bimodule, where R acts on the left and on the right by multiplication.

By the construction of the cord algebra $\text{Cord}(K)$ from Section 2.2, $H_0^{\text{string}}(K)$ is a $Z\hat{\pi}$ -NC-algebra. We now give an alternative definition of $H_0^{\text{string}}(K)$ that uses π and $\hat{\pi}$ in place of cords.

A *broken word* in $\pi, \hat{\pi}$ is a nonempty word in elements of π and $\hat{\pi}$ whose letters alternate between elements in π and $\hat{\pi}$. For clarity, we use Roman letters for elements in π and Greek for $\hat{\pi}$, and enclose elements in $\pi, \hat{\pi}$ by square and curly brackets, respectively. Thus examples of broken words are $f\alpha g$, $[x]$, $[x]f\alpha g$, and $f\alpha_1 g[x_1]f\alpha_2 g[x_2]f\alpha_3 g$.

Consider the \mathbb{Z} -module freely generated by broken words in $\pi, \hat{\pi}$, divided by the following *string relations*:

- (i) ${}_1[x\alpha_1]f\alpha_2g_2 = {}_1[x]f\alpha_1\alpha_2g_2$
- (ii) ${}_1f\alpha_1g[\alpha_2x]_2 = {}_1f\alpha_1\alpha_2g[x]_2$
- (iii) $({}_1[x_1x_2]_2) ({}_1[x_1\mu x_2]_2) = {}_1[x_1]f1g[x_2]_2$
- (iv) $({}_1f\alpha_1\alpha_2g_2) ({}_1f\alpha_1\mu\alpha_2g_2) = {}_1f\alpha_1g[1]f\alpha_2g_2$.

Here ${}_1$ is understood to represent the same (possibly empty) subword each time it appears, as is $_2$. We denote the resulting quotient by $S(\pi, \hat{\pi})$.

The \mathbb{Z} -module $S(\pi, \hat{\pi})$ splits into a direct sum corresponding to the four possible beginnings and endings for broken words:

$$S(\pi, \hat{\pi}) = S^{\hat{\pi}\hat{\pi}}(\pi, \hat{\pi}) \oplus S^{\hat{\pi}\pi}(\pi, \hat{\pi}) \oplus S^{\pi\hat{\pi}}(\pi, \hat{\pi}) \oplus S^{\pi\pi}(\pi, \hat{\pi}),$$

where the superscripts denote which of π and $\hat{\pi}$ contain the first and last letters in the broken word. Thus $S^{\hat{\pi}\hat{\pi}}(\pi, \hat{\pi})$ is generated by broken words beginning and ending with curly brackets (elements of $\hat{\pi}$)— $f\alpha g$, $f\alpha_1g[x]f\alpha_2g$, etc.—while $S^{\pi\pi}(\pi, \hat{\pi})$ is generated by $[x]$, $[x]f\alpha g[y]$, etc. We think of these broken words as broken strings with base point on $N \setminus Q$ beginning and ending with N -strings (for $S^{\hat{\pi}\hat{\pi}}(\pi, \hat{\pi})$) or Q -strings (for $S^{\pi\pi}(\pi, \hat{\pi})$). The other two summands $S^{\hat{\pi}\pi}(\pi, \hat{\pi})$, $S^{\pi\hat{\pi}}(\pi, \hat{\pi})$ can similarly be interpreted in terms of broken strings, but we will not consider them further.

On $S^{\hat{\pi}\hat{\pi}}(\pi, \hat{\pi})$ and $S^{\pi\pi}(\pi, \hat{\pi})$, we can define multiplications by

$$({}_1f\alpha_1g)({}_1f\alpha_2g_2) = {}_1f\alpha_1\alpha_2g_2$$

and

$$({}_1[x_1])([x_2]_2) = {}_1[x_1x_2]_2,$$

respectively. These turn $S^{\hat{\pi}\hat{\pi}}(\pi, \hat{\pi})$ and $S^{\pi\pi}(\pi, \hat{\pi})$ into rings. Note for future reference that $S^{\hat{\pi}\hat{\pi}}(\pi, \hat{\pi})$ is generated as a ring by $f\alpha g$ and $f1g[x]f1g$ for $\alpha \in \hat{\pi}$ and $x \in \pi$.

PROPOSITION 2.13. — *$S^{\hat{\pi}\hat{\pi}}(\pi, \hat{\pi})$ is a $\mathbb{Z}\hat{\pi}$ -NC-algebra, while $S^{\pi\pi}(\pi, \hat{\pi})$ is a $\mathbb{Z}\pi$ -NC-algebra and hence a $\mathbb{Z}\hat{\pi}$ -NC-algebra as well. Both $S^{\hat{\pi}\hat{\pi}}(\pi, \hat{\pi})$ and $S^{\pi\pi}(\pi, \hat{\pi})$ are knot invariants as NC-algebras.*

Proof. — We only need to specify the ring homomorphisms $\mathbb{Z}\hat{\pi} \rightarrow S^{\hat{\pi}\hat{\pi}}(\pi, \hat{\pi})$ and $\mathbb{Z}\pi \rightarrow S^{\pi\pi}(\pi, \hat{\pi})$; these are given by $\alpha \mapsto f\alpha g$ and $x \mapsto [x]$, respectively.

REMARK 2.14. — View $\mathbb{Z}\pi$ as a $\mathbb{Z}\hat{\pi}$ -bimodule via the map $\hat{\pi} \rightarrow \pi$. Then $S^{\hat{\pi}\hat{\pi}}(\pi, \hat{\pi})$ and $S^{\pi\pi}(\pi, \hat{\pi})$ can alternatively be defined as follows. Let A, \hat{A} be defined by

$$A = \mathbb{Z}\hat{\pi} \oplus \mathbb{Z}\pi \oplus (\mathbb{Z}\pi \oplus_{\mathbb{Z}\hat{\pi}} \mathbb{Z}\pi) \oplus (\mathbb{Z}\pi \oplus_{\mathbb{Z}\hat{\pi}} \mathbb{Z}\pi \oplus_{\mathbb{Z}\hat{\pi}} \mathbb{Z}\pi)$$

$$\hat{A} = \mathbb{Z}\pi \oplus (\mathbb{Z}\pi \oplus_{\mathbb{Z}\hat{\pi}} \mathbb{Z}\pi) \oplus (\mathbb{Z}\pi \oplus_{\mathbb{Z}\hat{\pi}} \mathbb{Z}\pi \oplus_{\mathbb{Z}\hat{\pi}} \mathbb{Z}\pi) \quad .$$

Each of A, \hat{A} has a multiplication operation given by concatenation (e.g. $a(b \cdot c) = a \cdot b \cdot c$); multiplying by an element of $\mathbb{Z}\hat{\pi} \rightarrow A$ uses the $\mathbb{Z}\hat{\pi}$ -bimodule structure on $\mathbb{Z}\pi$. There are two-sided ideals $I \subset A, \hat{I} \subset \hat{A}$ generated by

$$x_1x_2 - x_1\mu x_2 - x_1 - x_2$$

$$1_{\hat{\pi}} - (1 - \mu)_{\pi}$$

where $x_1, x_2 \in \pi$, $x_1 x_2, x_1 \mu x_2$ are viewed as elements in $Z\pi$, and $1_{\widehat{\pi}}$ denotes the element $1 \in Z\widehat{\pi}$ while $(1 - \mu)_{\pi}$ denotes the element $1 - \mu \in Z\pi$. Then

$$\begin{aligned} S^{\widehat{\pi}\widehat{\pi}}(\pi, \widehat{\pi}) &= A / I \\ S^{\pi\pi}(\pi, \widehat{\pi}) &= \widehat{A} / \widehat{I}. \end{aligned}$$

We conclude this subsection by noting that $S^{\widehat{\pi}\widehat{\pi}}(\pi, \widehat{\pi})$ is precisely the cord algebra of K .

PROPOSITION 2.15. — *We have the following isomorphism of $Z\widehat{\pi}$ -NC-algebras:*

$$H_0^{\text{string}}(K) = S^{\widehat{\pi}\widehat{\pi}}(\pi, \widehat{\pi}).$$

Proof. — We use the cord-algebra formulation of $H_0^{\text{string}}(K) = \text{Cord}(K)$ from Definition 2.6. Let K^{θ} be the parallel copy of K , and choose a base point p^{θ} for $\pi = \pi_1(Q \sqcup K)$ with $p^{\theta} \in K^{\theta} \cap f \cdot g$. Given a cord γ of K , define $\widetilde{\gamma} \in \pi$ as in Remark 2.8: extend γ to a closed loop $\widetilde{\gamma}$ in $Q \sqcup K$ with endpoints at p^{θ} by connecting the endpoints of γ to p^{θ} along $K^{\theta} \cap f \cdot g$. Then the isomorphism $\phi : \text{Cord}(K) \xrightarrow{\sim} S^{\widehat{\pi}\widehat{\pi}}(\pi, \widehat{\pi})$ is the ring homomorphism defined by:

$$\begin{aligned} \phi(\gamma) &= f_1 g[\widetilde{\gamma}] f_1 g \\ \phi(\alpha) &= f \alpha g, \end{aligned}$$

for γ any cord of K and α any element of $\text{Cord}(K)$ of the form $\lambda^a \mu^b$.

The skein relations in $\text{Cord}(K)$ from Definition 2.6 are mapped by ϕ to:

- (i) $f_1 g[1] f_1 g = f_1 g - f \mu g$
- (ii) $f_1 g[\mu \widetilde{\gamma}] f_1 g = f \mu g[\widetilde{\gamma}] f_1 g$ and $f_1 g[\widetilde{\gamma} \mu] f_1 g = f_1 g[\widetilde{\gamma}] f \mu g$
- (iii) $f_1 g[\lambda \widetilde{\gamma}] f_1 g = f \lambda g[\widetilde{\gamma}] f_1 g$ and $f_1 g[\widetilde{\gamma} \lambda] f_1 g = f_1 g[\widetilde{\gamma}] f \lambda g$
- (iv) $f_1 g[\widetilde{\gamma}_1 \widetilde{\gamma}_2] f_1 g - f_1 g[\widetilde{\gamma}_1 \mu \widetilde{\gamma}_2] f_1 g = f_1 g[\widetilde{\gamma}_1] f_1 g[\widetilde{\gamma}_2] f_1 g$.

In $S^{\widehat{\pi}\widehat{\pi}}(\pi, \widehat{\pi})$, these follow from string relations (iv), (i) and (ii), (i) and (ii), and (iii), respectively.

Thus ϕ is well-defined. It is straightforward to check that ϕ is an isomorphism (indeed, the string relations are constructed so that this is the case), with inverse ϕ^{-1} defined by

$$\begin{aligned} \phi^{-1}(f \alpha g) &= \alpha \\ \phi^{-1}(f_1 g[\widetilde{\gamma}] f_1 g) &= \widetilde{\gamma}, \end{aligned}$$

for $\alpha \in \widehat{\pi}$ and $\widetilde{\gamma} \in \pi$: note that a closed loop at $p^{\theta} \in K^{\theta} \cap f \cdot g$ representing $\widetilde{\gamma}$ is also by definition a cord of K .

REMARK 2.16. — Similarly, one can show that $\widehat{H}_0^{\text{string}}(K) = S^{\pi\pi}(\pi, \widehat{\pi})$ as $Z\pi$ -NC-algebras. In the same vein, there is also a cord formulation for modified string homology $\widehat{H}_0^{\text{string}}(K)$ (as introduced at the end of Section 2.1), along the lines of Definition 2.6: this is $\widehat{A} / \widehat{I}$, where \widehat{A} is the non-unital algebra generated by nonempty products of cords (the difference from A being that \widehat{A} does not contain words of the form $\lambda^a \mu^b$, which have no cords), and \widehat{I} is the ideal of \widehat{A} generated by skein relations (ii) through (iv) from Definition 2.6, without (i).

2.4. THE CORD ALGEBRA AND GROUP RINGS. — Having defined the cord algebra in terms of homotopy groups, we can now give an even more explicit interpretation not involving broken words, in terms of the group ring $Z\pi$. Notation is as in Section 2.3: in particular, $K \subset Q$ is a framed oriented knot with tubular neighborhood N , $\pi = \pi_1(Q \setminus K)$, and $\hat{\pi} = \pi_1(\partial N)$. When $Q = \mathbb{R}^3$, we assume for simplicity that the framing on K is the Seifert framing.

Before addressing the cord algebra $H_0^{\text{string}}(K) = S^{\hat{\pi}\hat{\pi}}(\pi, \hat{\pi})$ itself, we first note that the modified version $\hat{H}_0^{\text{string}}(K) = S^{\pi\pi}(\pi, \hat{\pi})$ is precisely $Z\pi$.

PROPOSITION 2.17. — For a knot $K \subset Q$, we have an isomorphism as $Z\pi$ -NC-algebras

$$S^{\pi\pi}(\pi, \hat{\pi}) = Z\pi.$$

Proof. — The map $Z\pi \rightarrow S^{\pi\pi}(\pi, \hat{\pi})$, $x \mapsto [x]$, has inverse ϕ given by

$$\begin{aligned} \phi([x_1]f\alpha_1g[x_2]f\alpha_2g \dots [x_{n-1}]f\alpha_{n-1}g[x_n]) \\ = x_1(1 - \mu)\alpha_1x_2(1 - \mu)\alpha_2 \dots x_{n-1}(1 - \mu)\alpha_{n-1}x_n; \end{aligned}$$

note that ϕ is well-defined (just check the string relations) and preserves ring structure.

The corresponding description of the cord algebra $S^{\hat{\pi}\hat{\pi}}(\pi, \hat{\pi})$ is a bit more involved, and we give two interpretations.

PROPOSITION 2.18. — For a knot $K \subset Q$, we have a Z -module isomorphism

$$S^{\hat{\pi}\hat{\pi}}(\pi, \hat{\pi}) = Z[\lambda^{-1}] \otimes Z\pi.$$

For any $\alpha \in Z[\lambda^{-1}]$, the left and right actions of α on $S^{\hat{\pi}\hat{\pi}}(\pi, \hat{\pi})$ induced from the $\hat{\pi}$ -NC-algebra structure on $S^{\hat{\pi}\hat{\pi}}(\pi, \hat{\pi})$ coincide under this isomorphism with the actions of α on the factors of $Z[\lambda^{-1}] \otimes Z\pi$ by left and right multiplication.

Proof. — The isomorphism $Z[\lambda^{-1}] \otimes Z\pi \rightarrow S^{\hat{\pi}\hat{\pi}}(\pi, \hat{\pi})$ sends $(\alpha, 0)$ to $f\alpha g$ and $(0, x)$ to $f\lambda g[x]f\lambda g$. Note that this map commutes with left and right multiplication by powers of λ ; for example, $f\lambda^k\alpha g = \lambda^k f\alpha g$ and $f\lambda g[\lambda^k x]f\lambda g = f\lambda^k g[x]f\lambda g = \lambda^k f\lambda g[x]f\lambda g$.

To see that the map is a bijection, note that the generators of $S^{\hat{\pi}\hat{\pi}}(\pi, \hat{\pi})$ can be separated into “trivial” broken words of the form $f\alpha g$ and “nontrivial” broken words of length at least 3. Using the string relations, we can write any trivial broken word uniquely as a sum of some $f\lambda^a g$ and some nontrivial broken words:

$$f\lambda^a \mu^b g = f\lambda^a g + \sum_{i=0}^{b-1} f\lambda^a g[\mu^i]f\lambda g$$

if $b > 0$, and similarly for $b < 0$. On the other hand, any nontrivial broken word in $S^{\hat{\pi}\hat{\pi}}(\pi, \hat{\pi})$ can be written uniquely as a Z -linear combination of words of the form $f\lambda g[x]f\lambda g$, $x \in \pi$: just use the map ϕ from the proof of Proposition 2.17 to reduce any nontrivial broken word to broken words of length 3, and then apply the identity $f\alpha_1 g[x]f\alpha_2 g = f\lambda g[\alpha_1 x \alpha_2]f\lambda g$.

PROPOSITION 2.19. — *For knots $K \subset \mathbb{R}^3$, string homology $H_0^{\text{string}}(K) = S^{\widehat{\pi}\widehat{\pi}}(\pi, \widehat{\pi})$, and thus knot contact homology, detects the unknot U . More precisely, left multiplication by $\lambda - 1$ on $S^{\widehat{\pi}\widehat{\pi}}(\pi, \widehat{\pi})$ has nontrivial kernel if and only if K is unknotted.*

Proof. — First, if $K = U$, then $\lambda = 1$ in π , and so

$$(\lambda - 1)f_1g[1]f_1g = f_1\lambda g[1]f_1g - f_1g[1]f_1g = f_1g[\lambda]f_1g - f_1g[1]f_1g = 0$$

in $H_0^{\text{string}}(U)$, while $f_1g[1]f_1g \neq 0$ by the proof of Proposition 2.18.

Next assume that $K \neq U$, and consider the effect of multiplication by $\lambda - 1$ on $Z[\lambda^{-1}] \cong Z\pi$. Clearly this map is injective on the $Z[\lambda^{-1}]$ summand; we claim that it is injective on the $Z\pi$ summand as well. Indeed, suppose that some nontrivial sum $\sum a_i[x_i] \in Z\pi$ is unchanged by multiplication by λ . Then $[\lambda^k x_1]$ must appear in this sum for all k , whence the sum is infinite since $\widehat{\pi}$ injects into π by the Loop Theorem.

REMARK 2.20. — It was first shown in [32] that the cord algebra detects the unknot. That proof uses a relationship between the cord algebra and the A -polynomial, along with the fact that the A -polynomial detects the unknot [8], which in turn relies on gauge-theoretic results of Kronheimer and Mrowka [28]. As noted previously, by contrast, the above proof that string homology detects the unknot uses only the Loop Theorem. Either proof shows that knot contact homology detects the unknot. However, we emphasize that for our argument, unlike the argument in [32], it is crucial that we use the fully noncommutative version of knot contact homology.

We can recover the multiplicative structure on $S^{\widehat{\pi}\widehat{\pi}}(\pi, \widehat{\pi})$ under the isomorphism of Proposition 2.18 as follows. On $Z[\lambda^{-1}] \cong Z\pi$, define a multiplication operation \cdot by

$$(\lambda^{k_1}, x_1) \cdot (\lambda^{k_2}, x_2) = (\lambda^{k_1+k_2}, \lambda^{k_1}x_2 + x_1\lambda^{k_2} + x_1x_2 - x_1\mu x_2).$$

It is easy to check that \cdot is associative, and that the isomorphism

$$S^{\widehat{\pi}\widehat{\pi}}(\pi, \widehat{\pi}) \cong (Z[\lambda^{-1}] \cong Z\pi, \cdot)$$

now becomes an isomorphism of $Z\widehat{\pi}$ -NC-algebras, where $Z[\lambda^{-1}] \cong Z\pi$ is viewed as a $Z\widehat{\pi}$ -NC-algebra via the map $Z[\widehat{\pi}] \rightarrow Z[\lambda^{-1}] \cong Z\pi$ sending λ to $(\lambda, 0)$ and μ to $(1, 0) - (0, 1)$.

We now turn to another formulation of string homology in terms of the group ring $Z\pi$. This formulation is a bit cleaner than the one in Proposition 2.18, as the multiplication operation is easier to describe.

PROPOSITION 2.21. — *For a knot $K \subset Q$, let R denote the subring of $Z\pi$ generated by $\widehat{\pi}$ and $\text{im}(1 - \mu)$, where $1 - \mu$ denotes the map $Z\pi \rightarrow Z\pi$ given by left multiplication by $1 - \mu$. There is a ring homomorphism*

$$\psi : H_0^{\text{string}}(K) \rightarrow R$$

determined by $\psi(f\alpha g) = \alpha$ and $\psi(f_1g[x]f_1g) = x - \mu x$.

If $\widehat{\pi} \rightarrow \pi$ is an injection (in particular, if $K \subset \mathbb{R}^3$ is nontrivial), then ψ is an isomorphism of $Z\widehat{\pi}$ -NC-algebras.

Proof. — It is easy to check that ψ respects all of the string relations defining $S^{\widehat{\pi}}(\pi, \widehat{\pi}) = H_0^{\text{string}}(K)$: the key relation $[x_1x_2] - [x_1\mu x_2] - [x_1]f_1g[x_2]$ is sent to $(1 - \mu)x_1x_2 - (1 - \mu)x_1\mu x_2 - (1 - \mu)x_1(1 - \mu)x_2 = 0$. Thus ψ is well-defined as a map $H_0^{\text{string}}(K) \rightarrow \mathbb{R}$. This map acts as the identity on $\widehat{\pi}$ and thus is a $Z\widehat{\pi}$ -NC-algebra map.

Since ψ is surjective by construction, it remains only to show that ψ is injective when $\widehat{\pi} \rightarrow \pi$ is injective. Suppose that

$$(2.1) \quad 0 = \psi \left(\sum_i a_i f \lambda^i g + \sum_j b_j f_1 g[x_j] f_1 g \right) = \sum_i a_i \lambda^i + \sum_j b_j (1 - \mu)x_j$$

for some $a_i, b_j \in \mathbb{Z}$ and $x_j \in \pi$. We claim that $b_j = 0$ for all j , whence $a_i = 0$ for all i since $\widehat{\pi}$ injects into π for K nontrivial. Assume without loss of generality that the framing on K is the 0-framing (changing framing simply replaces λ by $\lambda\mu^k$ for some k). Then the linking number with K gives a homomorphism $\text{lk}: \pi \rightarrow \mathbb{Z}$ satisfying $\text{lk}(\lambda) = 0$ and $\text{lk}(\mu) = 1$. If $\sum_j b_j x_j$ is not a trivial sum, then let x_ℓ be the contributor to this sum of maximal linking number. The term $b_\ell \mu x_\ell$ in $\sum_j b_j (1 - \mu)x_j$ cannot be canceled by any other term in that sum; thus for (2.1) to hold, x_ℓ must have linking number -1 . But a similar argument shows that the contributor to $\sum_j b_j x_j$ of minimal linking number must have linking number 0, contradiction. We conclude that $\sum_j b_j x_j$ must be a trivial sum, as claimed.

REMARK 2.22. — To be clear, as a knot invariant derived from knot contact homology, the cord algebra $H_0^{\text{string}}(K)$ (for $K \neq U$) is the ring $\mathbb{R} \langle \lambda^{-1}, \mu^{-1} \rangle$ along with the map $Z\widehat{\pi} = \mathbb{Z}[\lambda^{-1}, \mu^{-1}] \rightarrow \mathbb{R}$. Proposition 2.21 implies that the $\mathbb{Z}[\lambda^{-1}, \mu^{-1}]$ -NC-algebra structure on $Z\pi = \widehat{H}_0^{\text{string}}(K)$ completely determines $H_0^{\text{string}}(K)$. We do not know if $H_0^{\text{string}}(K)$ determines $\widehat{H}_0^{\text{string}}(K)$ as well, nor whether $H_0^{\text{string}}(K)$ is a complete knot invariant.⁽⁴⁾

On the other hand, $\widehat{H}_0^{\text{string}}(K)$ as a ring is a complete knot invariant for prime knots in \mathbb{R}^3 up to mirroring, as we can see as follows. By Proposition 2.17, $\widehat{H}_0^{\text{string}}(K) = \mathbb{Z}[\pi]$, and for prime knots K , Gordon and Luecke [24] show that $\pi = \pi_1(\mathbb{R}^3 \setminus K)$ determines K up to mirroring. On the other hand, π is a left-orderable group, and the ring isomorphism type of $\mathbb{Z}[G]$ when G is left-orderable is determined by the group isomorphism type of G [29]. We thank Tye Lidman for pointing this out to us.

We conclude this section with two examples.

EXAMPLE 2.23. — When K is the unknot U , then

$$H_0^{\text{string}}(U) = \mathbb{Z}[\lambda^{-1}, \mu^{-1}] / ((\lambda^{-1} - 1)(\mu^{-1} - 1)),$$

⁽⁴⁾Added in revision: it has now been proven by Shende [35], and then reproven in [17], that the Legendrian isotopy type of π_K completely determines the knot K . The proof in [17] relies on the present paper and shows that an enhanced version of knot contact homology (or of $H_0^{\text{string}}(K)$) determines K . The question of whether $H_0^{\text{string}}(K)$ is a complete invariant remains open.

while $\widehat{H}_0^{\text{string}}(U) = \mathbb{Z}[\mu^{-1}]$. The ring homomorphism ψ from Proposition 2.21, which is not injective, is given by $\psi(\lambda) = 1$, $\psi(\mu) = \mu$. The isomorphism from Proposition 2.18 is the (inverse of the) map

$$\begin{aligned} \mathbb{Z}[\lambda^{-1}] &\cong \mathbb{Z}[\mu^{-1}] \cong \mathbb{Z}[\lambda^{-1}, \mu^{-1}] / ((\lambda^{-1})(\mu^{-1})) \\ &\cong (\alpha, \beta) \cong \alpha + (\mu^{-1})\beta. \end{aligned}$$

As noticed by Lidman, this computation of $H_0^{\text{string}}(U)$ along with Proposition 2.21 gives an alternative (and shorter) proof that knot contact homology detects the unknot (Proposition 2.19), and more generally that this continues to hold even if the knot is not assumed to be Seifert framed (Corollary 1.5).

Proof of Corollary 1.5. — Suppose that $H_0^{\text{string}}(K) = H_0^{\text{string}}(U)$ where K is a framed oriented knot and U is the unknot with some framing. By changing the framing of both, we can assume that K has its Seifert framing. If K is knotted, then $\mathbb{Z}\pi$ has no zero divisors since π is left-orderable, and thus $H_0^{\text{string}}(K) \cong \mathbb{Z}\pi$ also has no zero divisors by Proposition 2.21. On the other hand,

$$H_0^{\text{string}}(U) = \mathbb{Z}[\lambda^{-1}, \mu^{-1}] / ((\lambda\mu^f - 1)(\mu^{-1}))$$

for some $f \in \mathbb{Z}$. Thus K must be the unknot and must further have the same framing as U .

In [23], Gordon and Lidman extend this line of argument (i.e., applying Proposition 2.21) to prove that knot contact homology detects torus knots as well as cabling and compositeness.

EXAMPLE 2.24. — When K is the right-handed trefoil T , a slightly more elaborate version of the calculation of the cord algebra from [32] (see also [34]) gives the following expression for $H_0^{\text{string}}(T)$: it is generated by λ^{-1} , μ^{-1} , and one more generator x , along with the relations:

$$\begin{aligned} \lambda\mu &= \mu\lambda \\ \lambda\mu^6 x &= x\lambda\mu^6 \\ 1 + \mu + x - \lambda\mu^5 x\mu^{-3} x\mu^{-1} &= 0 \\ 1 - \mu - \lambda\mu^4 x\mu^{-2} - \lambda\mu^5 x\mu^{-2} x\mu^{-1} &= 0. \end{aligned}$$

On the other hand, $\widehat{H}_0^{\text{string}}(T) = \mathbb{Z}\pi$ is the ring generated by μ^{-1} and a^{-1} modulo the relation $\mu a \mu = a \mu a$; the longitudinal class is $\lambda = a \mu a^{-1} \mu a \mu^{-3}$. The explicit map from $H_0^{\text{string}}(T)$ to $\mathbb{Z}\pi$ is given by:

$$\begin{aligned} \mu &\mapsto \mu \\ \lambda &\mapsto \lambda = a \mu a^{-1} \mu a \mu^{-3} \\ x &\mapsto (1 - \mu) a \mu^{-1} a^{-1}. \end{aligned}$$

It can be checked that this map preserves the relations in $H_0^{\text{string}}(T)$.

3. ROADMAP TO THE PROOF OF THEOREM 1.2

The remainder of this paper is devoted to the proof of Theorem 1.2. To avoid getting lost in the details, we give here a roadmap to the proof and explain the technical issues to be addressed along the way.

The proof follows the scheme that is described for a different situation in [6] and consists of 3 steps. Let A be the free $\mathbb{Z}\widehat{\pi}$ -NC-algebra generated by Reeb chords and $\partial : A \rightarrow A$ the boundary operator for Legendrian contact homology. For a Reeb chord a and an integer $\ell > 0$ denote by $M_\ell(a)$ the moduli space of J -holomorphic disks in T^*Q with one positive puncture asymptotic to a and boundary on $Q \cup L_K$ with 2ℓ corners at which it switches between L_K and Q .

Step 1. — Show that $M_\ell(a)$ can be compactified to a manifold with corners $\overline{M}_\ell(a)$ and that the generating functions $\phi(a) := \sum_{\ell=0}^1 \overline{M}_\ell(a)$ (extended as algebra maps to A) satisfy the relation

$$\partial\phi = \phi\partial - \delta\phi,$$

where $\delta\overline{M}_\ell(a)$ is the subset of elements in $\overline{M}_\ell(a)$ that intersect K at the interior of some boundary string.

Step 2. — Construct a chain complex $(C(\Sigma), \partial + \delta)$ of suitable chains of broken strings such that ϕ induces a chain map

$$\Phi : (A, \partial) \rightarrow (C(\Sigma), \partial + \delta),$$

and the homology $H_0(\Sigma, \partial + \delta)$ agrees with the string homology $H_0^{\text{string}}(K)$ as defined in Section 2.1.

Step 3. — Prove that Φ induces an isomorphism on homology in degree zero.

Step 1 occupies Sections 8 to 10. It involves detailed descriptions of

- the behavior of holomorphic disks at corner points;
- compactifications of moduli spaces of holomorphic disks;
- transversality and gluing of moduli spaces.

In Step 2 (Sections 4 to 6) we encounter the following problem: The direct approach to setting up the complex $(C(\Sigma), \partial + \delta)$ would involve chains in spaces of broken strings with varying number of switches. These spaces could probably be given smooth structures using the polyfold theory by Hofer, Wysocki and Zehnder [27]. Here we choose a different approach, keeping the number of switches fixed and inserting small “spikes” in the definition of the string operation $\delta = \delta_Q + \delta_N$. Since this involves non-canonical choices, one does not expect identities such as $\partial\delta + \delta\partial = 0$ to hold strictly but only up to homotopy, thus leading to an \mathcal{A} -structure as described by Sullivan in [37]. We avoid \mathcal{A} -structures by carefully defining δ via induction over the dimension of chains such that all identities hold strictly on the chain level.

Step 3 (Section 7) follows the scheme described in [6]. This involves

- a length estimate for the boundary of holomorphic disks, which implies that Φ respects the filtrations of A and $C(\Sigma)$ by the actions of Reeb chords and the total lengths of Q -strings, respectively.
- construction of a length-decreasing chain homotopy deforming $C(\Sigma)$ to chains $C(\Sigma_{\text{lin}})$ of broken strings all of whose Q -strings are *linear straight line segments* (at this point we specialize to $Q = \mathbb{R}^3$);
- Morse-theoretical arguments on the space Σ_{lin} to prove that Φ induces an isomorphism on degree zero homology.

4. HOLOMORPHIC FUNCTIONS NEAR CORNERS

In this section, we call a function $f : R \rightarrow \mathbb{C}$ on a subset $R \subset \mathbb{C}$ with piecewise smooth boundary *holomorphic* if it is continuous on R and holomorphic in the interior of R .

4.1. POWER SERIES EXPANSIONS. — Denote by $D \subset \mathbb{C}$ the open unit disk and set

$$\begin{aligned} D^+ &:= \{z \in D \mid \operatorname{Im}(z) > 0\}, \\ Q^+ &:= \{z \in D \mid \operatorname{Re}(z) > 0, \operatorname{Im}(z) > 0\}. \end{aligned}$$

Consider a holomorphic function $f : Q^+ \rightarrow \mathbb{C}$ (in the above sense, i.e., continuous on Q^+ and holomorphic in the interior) with $f(0) = 0$. We distinguish four cases according to their boundary conditions.

Case 1. — f maps \mathbb{R}_+ to \mathbb{R} and $i\mathbb{R}_+$ to $i\mathbb{R}$.

In this case, we extend f to a map $f : D^+ \rightarrow \mathbb{C}$ by the formula

$$f(z) := \overline{f(\bar{z})}, \quad \operatorname{Re}(z) < 0, \operatorname{Im}(z) > 0,$$

and then to a map $f : D \rightarrow \mathbb{C}$ by the formula

$$f(z) := \overline{f(\bar{z})}, \quad \operatorname{Im}(z) < 0.$$

The resulting map f is continuous on D and holomorphic outside the axes $\mathbb{R} \cup i\mathbb{R}$, hence holomorphic on D , and it maps \mathbb{R} to \mathbb{R} and $i\mathbb{R}$ to $i\mathbb{R}$. Thus it has a power series expansion

$$f(z) = \sum_{j=1}^{\infty} a_{2j} z^{2j-1}, \quad a_j \in \mathbb{R}.$$

This shows that each holomorphic function $f : Q^+ \rightarrow \mathbb{C}$ mapping \mathbb{R}_+ to \mathbb{R} and $i\mathbb{R}_+$ to $i\mathbb{R}$ is uniquely the restriction of such a power series. In particular, f has an isolated zero at the origin unless it vanishes identically. Similar discussions apply in the other cases.

Case 2. — f maps $(\mathbb{R}_+, i\mathbb{R}_+)$ to $(i\mathbb{R}, \mathbb{R})$. Then it has a power series expansion

$$f(z) = i \sum_{j=1}^{\infty} a_{2j} z^{2j-1}, \quad a_j \in \mathbb{R}.$$

Case 3. — f maps $(\mathbb{R}_+, i\mathbb{R}_+)$ to (\mathbb{R}, \mathbb{R}) . Then it has a power series expansion

$$f(z) = \sum_{j=1}^{\infty} a_{2j} z^{2j}, \quad a_j \in \mathbb{R}.$$

Case 4. — f maps $(\mathbb{R}_+, i\mathbb{R}_+)$ to $(i\mathbb{R}, i\mathbb{R})$. Then it has a power series expansion

$$f(z) = i \sum_{j=1}^{\infty} a_{2j} z^{2j}, \quad a_j \in \mathbb{R}.$$

REMARK 4.1. — We can summarize the four cases by saying that $f : \mathbb{Q}^+ \rightarrow \mathbb{C}$ is given by a power series

$$f(z) = \sum_{k=1}^{\infty} a_k z^k$$

with either only odd (in Cases 1 and 2) or only even (in Cases 3 and 4) indices k , and with the a_k either all real (in Cases 1 and 3) or all imaginary (in Cases 2 and 4). Such holomorphic functions f will appear as projections onto a normal direction of the holomorphic curves considered in Section 6.3 near switches. Then Case 1 corresponds to a switch from Q to N , Case 2 to a switch from N to Q , Case 3 to a switch from N to N , and Case 4 to a switch from Q to Q .

REMARK 4.2. — It will sometimes be convenient to switch from the positive quadrant to other domains. For example, the map $\psi(z) := \sqrt{z}$ maps the upper half disk D^+ biholomorphically onto \mathbb{Q}^+ . Thus in Case 1 the composition $f \circ \psi$ is a holomorphic function on D^+ which maps \mathbb{R}_+ to \mathbb{R} and $i\mathbb{R}_+$ to $i\mathbb{R}$, and it has an expansion in powers of \sqrt{z} by

$$f \circ \psi(z) = \sum_{j=1}^{\infty} a_{2j} z^{j-1/2}, \quad a_j \in \mathbb{R}.$$

As another example, the map $\phi(s, t) := ie^{-\pi(s+it)/2}$ maps the strip $(0, 1) \times [0, 1]$ biholomorphically onto $\mathbb{Q}^+ \cap \text{int}(\mathbb{D})$. Thus in Case 1 the composition $f \circ \phi$ is a continuous function on $\mathbb{R}_+ \times [0, 1]$ which is holomorphic in the interior and maps $\mathbb{R}_+ \times \{0\}$ to $i\mathbb{R}$ and $\mathbb{R}_+ \times \{1\}$ to \mathbb{R} , and it has a power series expansion

$$f \circ \phi(s, t) = i \sum_{j=1}^{\infty} (-1)^j a_{2j} e^{-(2j-1)\pi(s+it)/2}, \quad a_j \in \mathbb{R}.$$

Similar discussions apply to the other cases.

Let us consider once more the function $f : \mathbb{Q}^+ \rightarrow \mathbb{C}$ of Case 1 mapping $(\mathbb{R}_+, i\mathbb{R}_+)$ to $(\mathbb{R}, i\mathbb{R})$. Its restrictions to $i\mathbb{R}_+$ resp. \mathbb{R}_+ naturally give rise to functions

$$f_- : (-1, 0] \rightarrow \mathbb{R} \quad \text{resp.} \quad f_+ : [0, 1) \rightarrow \mathbb{R}$$

via

$$f_-(t) := (-i)f(-it), \quad t \in [0, 1), \quad f_+(t) := f(t), \quad t \in [0, 1).$$

Here and in the sequel we always use the isomorphism $(-i) = i^{-1} : i\mathbb{R} \rightarrow \mathbb{R}$ to identify $i\mathbb{R}$ with \mathbb{R} in the target. So f_{\pm} are related by $f_{\pm} = r f_{\mp}$, where the reflection $r f$ of a complex valued power series $f(t) = \sum_{k=1}^{\infty} a_k t^k, a_k \in \mathbb{C}$, is defined by

$$r f(t) := (-i) f(i t) = \sum_{k=1}^{\infty} (-i)^{k+1} a_k t^k.$$

(Note that the domain \mathbb{C} and the target \mathbb{C} play different roles here: multiplication by $(-i)$ on the domain comes from opening up the positive quadrant to the upper half plane, while multiplication by $(-i)$ in the target corresponds to the canonical rotation by J from $i\mathbb{R} \rightarrow \mathbb{Q}$ to $\mathbb{R} \rightarrow \mathbb{N}$.)

The effect of r on the power series expansion $f(t) = \sum_{j=1}^{\infty} a_{2j-1} t^{2j-1}$ in Case 1 is as follows:

$$r f(t) = (-i) \sum_{j=1}^{\infty} a_{2j-1} (i t)^{2j-1} = \sum_{j=1}^{\infty} (-1)^j a_{2j-1} t^{2j-1},$$

so the coefficient a_{2j-1} is changed to $(-1)^j a_{2j-1}$. Note that a_1 is changed to $-a_1$, which justifies the name “reflection”.

Now consider f as in Case 2 mapping $(\mathbb{R}_+, i\mathbb{R}_+)$ to $(i\mathbb{R}, \mathbb{R})$. Here the restrictions to $i\mathbb{R}_+$ resp. \mathbb{R}_+ naturally give rise to functions $f_- : (0, 1] \rightarrow \mathbb{R}$ resp. $f_+ : [0, 1) \rightarrow \mathbb{R}$ via

$$f_-(t) := f(i t), t \in (0, 1], \quad f_+(t) := (-i) f(t), t \in [0, 1).$$

So f_{\pm} are related by $f_{\pm} = r f_{\mp}$, and the coefficient a_{2j-1} in the power series expansion of f_+ is changed to $(-1)^{j+1} a_{2j-1}$. In particular, a_1 is unchanged so that f and f_+ fit together to a function $(-1, 1) \rightarrow \mathbb{R}$ of class C^2 (but not C^3).

4.2. WINDING NUMBERS. — Consider a holomorphic function $f : Q^+ \rightarrow \mathbb{C}$ given by a power series $f(z) = \sum_{k=1}^{\infty} a_k z^k$ as in Cases 1–4 of the previous subsection. In each of these cases we define its winding number at 0 as

$$w(f, 0) := \frac{1}{2} \inf \{ \sum_{k=1}^{\infty} k |a_k| \neq 0 \}.$$

Note that the winding number is a half-integer in the first two cases and an integer in the last two cases. Also note that the winding number is given by

$$w(f, 0) = \frac{1}{\pi} \int_{\gamma} f d\theta,$$

where γ is a small arc in Q^+ connecting $(0, 1)$ to $i(0, 1)$. This can be seen, for example, by choosing γ as a small quarter circle $Q^+ \setminus \partial D_{\varepsilon}$; then the symmetry of f with respect to reflections at the coordinate axes implies

$$\frac{1}{\pi} \int_{\gamma} f d\theta = \frac{1}{4\pi} \int_{\partial D_{\varepsilon}} f d\theta = \frac{1}{4\pi} \cdot 2\pi \inf \{ \sum_{k=1}^{\infty} k |a_k| \neq 0 \} = w(f, 0).$$

Next let $r > 1$, denote by D_r the open disk of radius r , by

$$H^+ := \{ z \in \mathbb{C} \mid \operatorname{Im}(z) > 0 \}$$

the upper half plane, and set $D_r^+ := D_r \setminus H^+$. Consider a nonconstant continuous map $f : D_r^+ \rightarrow \mathbb{C}$ which is holomorphic in the interior and maps the interval $(-r, r)$ to $\mathbb{R} \cup i\mathbb{R}$. Suppose that f has no zeroes on the semi-circle $\partial D_1 \setminus H^+$. Then f has finitely many zeroes s_1, \dots, s_k in the interior of D_1^+ as well as finitely many zeroes t_1, \dots, t_ℓ in $(-1, 1)$. (Finiteness holds because the holomorphic function $z \mapsto f(z)^2$ maps \mathbb{R} to \mathbb{R} , and thus can only have finitely many zeroes by the Schwarz reflection principle and unique continuation.) Denote by $w(f, s_i) \in \mathbb{Z}$ resp. $w(f, t_j) \in \frac{1}{2}\mathbb{Z}$ the winding numbers at the zeroes. Thus with the closed angular form $d\theta$ on $\mathbb{C} \setminus \{0\}$,

$$w(f, s_i) := \frac{1}{\pi} \int_{\alpha_i} f \, d\theta, \quad w(f, t_j) := \frac{1}{\pi} \int_{\beta_j} f \, d\theta,$$

where α_i is a small circle around s_i and β_j is a small semi-circle around t_j in D_1^+ , both oriented in the counterclockwise direction. (Thus the $w(f, s_i)$ are even integers and the $w(f, t_j)$ are integers or half-integers). Denote by γ the semi-circle $\partial D_1 \setminus H^+$ oriented in the counterclockwise direction. Then Stokes' theorem yields

$$\frac{1}{\pi} \int_{\gamma} f \, d\theta = \sum_{i=1}^k w(f, s_i) + \sum_{j=1}^{\ell} w(f, t_j).$$

Since all winding numbers are nonnegative, we have shown the following result.

LEMMA 4.3. — *Consider a nonconstant continuous map $f : D_r^+ \rightarrow \mathbb{C}$ which is holomorphic in the interior and maps $(-r, r)$ to $\mathbb{R} \cup i\mathbb{R}$. Suppose that f has no zeroes on the semi-circle $\gamma = \partial D_1 \setminus H^+$ and zeroes at $t_1, \dots, t_m \in (-1, 1)$ (plus possibly further zeroes in D_1^+). Then*

$$\frac{1}{\pi} \int_{\gamma} f \, d\theta > \sum_{j=1}^m w(f, t_j).$$

More generally, for $n > 1$ consider a nonconstant continuous map $f : D_r^+ \rightarrow \mathbb{C}^n$ which is holomorphic in the interior and maps $(-r, r)$ to $\mathbb{R}^n \cup i\mathbb{R}^n$. Suppose that f has no zeroes on the semi-circle $\partial D_1 \setminus H^+$ and zeroes z_1, \dots, z_m in D_1^+ (in the interior or on the boundary). For each direction $v \in S^{n-1} \subset \mathbb{R}^n$ we obtain a holomorphic map $f_v := \pi_v \circ f$, where π_v is the projection onto the complex line spanned by v . Fix a positive volume form Ω on S^{n-1} of total volume 1. Then there exists an open subset $V \subset S^{n-1}$ of measure 1 such that for all $v \in V$, f_v has zeroes precisely at the z_j and their winding numbers are independent of $v \in V$. So we can define

$$w(f, z_j) := \int_V w(f_v, z_j) \Omega(v) = w(f_{v_0}, z_j)$$

for any $v_0 \in V$ and obtain

COROLLARY 4.4. — *Consider a nonconstant continuous map $f : D_r^+ \rightarrow \mathbb{C}^n$ which is holomorphic in the interior and maps $(-r, r)$ to $\mathbb{R}^n \cup i\mathbb{R}^n$. Suppose that f has no zeroes on the semi-circle $\gamma = \partial D_1 \setminus H^+$ and zeroes at $t_1, \dots, t_m \in (-1, 1)$ (plus possibly*

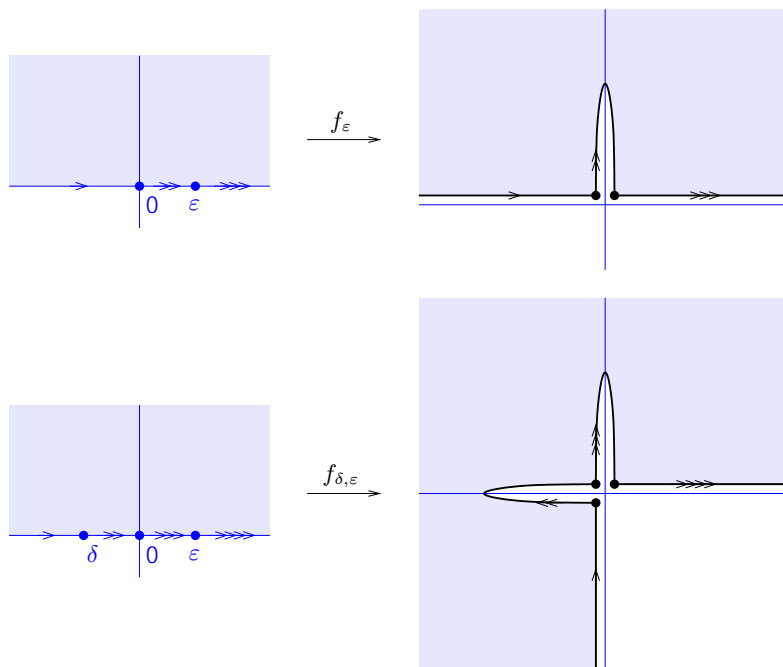


FIGURE 4.1. Spikes in the model families f_ε and $f_{\delta,\varepsilon}$.

further zeroes in D_1^+). Then there exists an open subset $V \subset S^{n-1}$ of measure 1 such that for every $v_0 \in V$,

$$\frac{1}{\pi} \int_\gamma f_{v_0} d\theta = \int_V \left(\frac{1}{\pi} \int_\gamma f_v d\theta \right) \Omega(v) > \sum_{j=1}^m w(f, t_j).$$

4.3. SPIKES. — Consider again the upper half disk $D^+ = \{z \in D \mid \text{Im}(z) > 0\}$ and real points $0 < b_1 < b_2 < \dots < b_\ell < 1$. We are interested in holomorphic functions $f : D^+ \setminus \bigcup_{i=1}^\ell [b_{i-1}, b_i] \rightarrow \mathbb{C}$, continuous on D^+ , mapping the intervals $[b_{i-1}, b_i]$ alternately to \mathbb{R} and $i\mathbb{R}$. We wish to describe models of 1- resp. 2-parameter families in which 2 resp. 3 of the b_i come together. A model for such a 1-parameter family is

$$(4.1) \quad f_\varepsilon(z) := \sqrt{z(z - \varepsilon)}, \quad \varepsilon > 0$$

with zeroes at $0, \varepsilon$. A model for a 2-parameter family is

$$(4.2) \quad f_{\delta,\varepsilon}(z) := \sqrt{z(z + \delta)(z - \varepsilon)}, \quad \varepsilon, \delta > 0$$

with zeroes at $-\delta, 0, \varepsilon$. Here we choose appropriate branches of the square root so that the functions become continuous. The images of these functions are shown in Figure 4.1. They show that f_ε has a “spike” in the direction $i\mathbb{R}_+$ which disappears as $\varepsilon \rightarrow 0$, and $f_{\delta,\varepsilon}$ has two “spikes” in the directions \mathbb{R} resp. $i\mathbb{R}_+$ which disappear as δ resp. ε approaches zero. Based on these models, the notion of a “spike” will be formalized in Section 5.

In the following section, functions with two spikes will appear in the following local model. Consider the 1-parameter family of functions $f_a : \mathbb{Q}^+ \rightarrow \mathbb{C}$,

$$f_a(z) = i(az - z^3), \quad a \geq \mathbb{R}.$$

They map $(\mathbb{R}_+, i\mathbb{R}_+)$ to $(i\mathbb{R}, \mathbb{R})$ and thus correspond to Case 2 in Section 4.1. Via the identifications in that section, f_a induces functions

$$\begin{aligned} f_-(a, t) &:= f_a(-it) = at + t^3, & t \leq 0, \\ f_+(a, t) &:= (i)f_a(t) = at - t^3, & t > 0, \end{aligned}$$

which fit together to a C^2 (though not C^3) function

$$f(a, t) = at - \operatorname{sgn}(t)t^3, \quad t \in \mathbb{R}.$$

In Case 1, one considers the functions $f_a(z) = az + z^3$ mapping $(\mathbb{R}_+, i\mathbb{R}_+)$ to $(\mathbb{R}, i\mathbb{R})$. Here the induced functions

$$\begin{aligned} f_-(a, t) &:= (i)f_a(-it) = at + t^3, & t \leq 0, \\ f_+(a, t) &:= f_a(t) = at + t^3, & t > 0 \end{aligned}$$

do not fit together to a C^1 function, but when we replace f_+ by $-f_+$ they fit together to the function $f(a, t)$ above.

5. STRING HOMOLOGY IN ARBITRARY DEGREE

5.1. **BROKEN STRINGS.** — Let K be a framed oriented knot in some oriented 3-manifold Q . Fix a tubular neighborhood N of K and a diffeomorphism $N \cong S^1 \times D^2$.

Fix an integer $m > 3$ and a base point $x_0 \in \partial N$. We also fix an m -jet of a curve passing through x_0 in N . Using the diffeomorphism $N \cong S^1 \times D^2$, this is equivalent to specifying suitable vectors $v_0^{(k)} \in \mathbb{R}^3$, $1 \leq k \leq m$. The following definition refines the one given in Section 2, which corresponds to the case $m = 1$.

DEFINITION 5.1. — A broken (closed) string with 2ℓ switches on K is a tuple $s = (a_1, \dots, a_{2\ell+1}; s_1, \dots, s_{2\ell+1})$ consisting of real numbers $0 = a_0 < a_1 < \dots < a_{2\ell+1}$ and C^m -maps

$$s_{2i+1} : [a_{2i}, a_{2i+1}] \rightarrow N, \quad s_{2i} : [a_{2i-1}, a_{2i}] \rightarrow Q$$

satisfying the following matching conditions at the end points a_i :

- (i) $s_1(0) = s_{2\ell+1}(a_{2\ell+1}) = x_0$ and $s_1^{(k)}(0) = s_{2\ell+1}^{(k)}(a_{2\ell+1}) = v_0^{(k)}$ for $1 \leq k \leq m$.
- (ii) For $i = 1, \dots, \ell$,

$$s_{2i}(a_{2i}) = s_{2i+1}(a_{2i}) \in K, \quad s_{2i-1}(a_{2i-1}) = s_{2i}(a_{2i-1}) \in K.$$

(iii) Denote by σ_i the D^2 -component of s_i near its end points. Then for $i = 1, \dots, \ell$ and $1 \leq k \leq m/2$ (for the left hand side) resp. $1 \leq k \leq (m+1)/2$ (for the right hand side)

$$\begin{aligned} \sigma_{2i}^{(2k)}(a_{2i}) = \sigma_{2i+1}^{(2k)}(a_{2i}) = 0, & \quad \sigma_{2i}^{(2k-1)}(a_{2i}) = (-1)^k \sigma_{2i+1}^{(2k-1)}(a_{2i}), \\ \sigma_{2i-1}^{(2k)}(a_{2i-1}) = \sigma_{2i}^{(2k)}(a_{2i-1}) = 0, & \quad \sigma_{2i-1}^{(2k-1)}(a_{2i-1}) = (-1)^{k+1} \sigma_{2i}^{(2k-1)}(a_{2i-1}). \end{aligned}$$

We will refer to the s_{2i} and s_{2i+1} as *Q-strings* and *N-strings*, respectively. A typical picture of a broken string is shown in Figure 2.1 on page 666. Conditions (i) and (ii) in Definition 5.1 mean that the s_i fit together to a continuous loop $s : [0, a_{2\ell+1}] \rightarrow Q$ with end points at x_0 (which fit together in C^m).

Condition (iii) is motivated as follows: In Section 8 below, we consider almost complex structures J on TQ which are particularly well adapted to the immersed Lagrangian submanifold $Q \subset L_K \subset TQ$. For such a J , Lemma 8.6 then provides a holomorphic embedding of a neighborhood O of $S^1 \times \{0\} \subset C^2$ onto a neighborhood of $K \subset TQ$ mapping $O \setminus (S^1 \times iR^2)$ to Q and $O \setminus (S^1 \times R^2)$ to L_K . Condition (iii) requires that the normal component σ of s at the switching points a_i behaves like the boundary values of a holomorphic disk with boundary on $Q \subset L_K$ when projected to C^2 in these coordinates near K .

To see this, let us reformulate condition (iii). As in Section 4.1, to a complex valued polynomial $p(t) = \sum_{k=1}^m p_k t^k$, $p_k \in C^2$, we associate its reflection

$$r p(t) = (\bar{i})p(\bar{it}) = \sum_{k=1}^m (\bar{i})^{k+1} p_k t^k.$$

Then two *real valued* polynomials $p(t) = \sum_{k=1}^m p_k t^k$ and $q(t) = \sum_{k=1}^m q_k t^k$, $p_k, q_k \in R^2$, satisfy $r p = q$ if and only if for $1 \leq k \leq m/2$ (on the left hand side) resp. $1 \leq k \leq (m + 1)/2$ (on the right hand side)

$$p_{2k} = q_{2k} = 0 \text{ and } p_{2k-1} = (-1)^k q_{2k-1}.$$

So in terms of the *normal Taylor polynomials* at the switching points

$$T^m \sigma_i(a_{i-1})(t) := \sum_{k=1}^m \frac{\sigma_i^{(k)}(a_{i-1})}{k!} t^k, \quad T^m \sigma_i(a_i)(t) := \sum_{k=1}^m \frac{\sigma_i^{(k)}(a_i)}{k!} t^k,$$

condition (iii) is equivalent to the conditions

$$T^m \sigma_{2i}(a_{2i}) = r T^m \sigma_{2i+1}(a_{2i}), \quad T^m \sigma_{2i-1}(a_{2i-1}) = r T^m \sigma_{2i}(a_{2i-1}).$$

These are precisely the conditions in Section 4.1 describing the boundary behavior of holomorphic disks at a corner going from the imaginary to the real axis (Case 1, corresponding to a switch from Q to N), resp. from the real to the imaginary axis (Case 2, corresponding to a switch from N to Q).

REMARK 5.2

(a) The case $m = 3$ suffices for the purposes of this paper. In fact, for 0- and 1-parametric families of strings we only need the conditions on the first derivatives (the case $m = 1$ considered in Section 2), while for 2-parametric families we also need the conditions on the second and third derivatives). Explicitly, condition (iii) for $m = 3$ reads

$$(5.1) \quad \begin{aligned} \sigma_{2i}^0(a_{2i}) &= \sigma_{2i+1}^0(a_{2i}), & \sigma_{2i-1}^0(a_{2i-1}) &= \sigma_{2i}^0(a_{2i-1}), \\ \sigma_{2i}^{00}(a_{2i}) &= \sigma_{2i+1}^{00}(a_{2i}) = \sigma_{2i-1}^{00}(a_{2i-1}) = \sigma_{2i}^{00}(a_{2i-1}) = 0, \\ \sigma_{2i}^{000}(a_{2i}) &= \sigma_{2i+1}^{000}(a_{2i}), & \sigma_{2i-1}^{000}(a_{2i-1}) &= \sigma_{2i}^{000}(a_{2i-1}). \end{aligned}$$

(b) In Definition 5.1 one could add the condition that all derivatives of the tangent components agree at switches (as it is the case for boundaries of holomorphic disks). However, we will not need such a condition and thus chose not to include it. Similarly, one could have required all the s_j to be C^1 rather than C^m .

We denote by Σ^ℓ the space of broken strings with 2ℓ switches. We make it a Banach manifold by equipping it with the topology of \mathbb{R} on the a_j and the C^m -topology on the s_j . It comes with interior evaluation maps

$$ev_i : (0, 1) \times \Sigma^\ell \rightarrow Q \text{ resp. } N, \quad (t, s) \mapsto s_i((1-t)a_{i-1} + ta_i)$$

and corner evaluation maps

$$T_i : \Sigma^\ell \rightarrow (\mathbb{R}^2)^{b(m+1)/2c}, \quad s \mapsto T^m \sigma_i(a_i) = (\sigma_i^{(2k-1)}(a_i))_{1 \leq k \leq b(m+1)/2c}$$

Moreover, concatenation at the base point x_0 yields a smooth map

$$\Sigma^\ell \times \Sigma^{\ell^0} \rightarrow \Sigma^{\ell+\ell^0}.$$

5.2. GENERIC CHAINS OF BROKEN STRINGS. — Now we define the generators of the string chain complex in degrees $d \geq 0, 1, 2g$. Set $\Delta_0 := \mathbb{R}^d$ and let

$$\Delta_d = \{(\lambda_1, \dots, \lambda_d) \in \mathbb{R}^d \mid \lambda_i > 0, \lambda_1 + \dots + \lambda_d \leq 1\}$$

denote the d -dimensional standard simplex for $d > 1$. It is stratified by the sets where some of the inequalities are equalities. Fix $m > 3$ as in the previous subsection.

DEFINITION 5.3. — A *generic d -chain* in Σ^ℓ is a smooth map $S : \Delta_d \rightarrow \Sigma^\ell$ such that the maps $ev_i \circ S : (0, 1) \times \Delta_d \rightarrow Q$ and $T_i \circ S : \Delta_d \rightarrow (\mathbb{R}^2)^{b(m+1)/2c}$ are jointly transverse to K resp. jet-transverse to 0 (on all strata of Δ_d).

Let us spell out what this means for $m = 3$ in the cases $d = 0, 1, 2$.

$d = 0$. — A generic 0-chain is a broken string $s = (s_1, \dots, s_{2\ell+1})$ such that

- (0a) $\dot{\sigma}_i(a_i) \neq 0$ for all i ;
- (0b) s_i intersects K only at its end points.

$d = 1$. — A generic 1-chain of broken strings is a smooth map

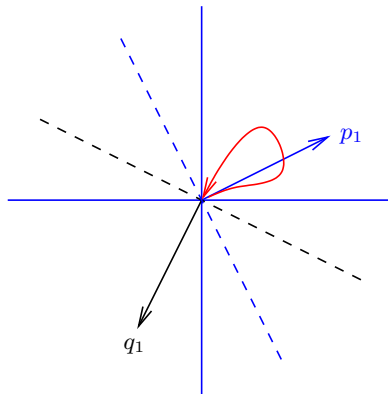
$$S : [0, 1] \rightarrow \Sigma^\ell, \quad \lambda \mapsto s^\lambda = (s_1^\lambda, \dots, s_{2\ell+1}^\lambda)$$

such that

- (1a) s^0 and s^1 are generic strings;
- (1b) $\dot{\sigma}_i^\lambda(a_i^\lambda) \neq 0$ for all i, λ ;
- (1c) for each i the map

$$(0, 1) \times (0, 1) \rightarrow Q \text{ resp. } N, \quad (t, \lambda) \mapsto s_i^\lambda((1-t)a_{i-1}^\lambda + ta_i^\lambda)$$

meets K transversely in finitely many points (t_a, λ_a) . Moreover, distinct such intersections (even for different i) appear at distinct parameter values λ_a .

FIGURE 5.1. A spike with ends (p, q) .

$d = 2$. — A generic 2-chain of broken strings is a smooth map

$$S : \Delta_2 \rightarrow \Sigma^\ell, \quad \lambda \in \Delta_2 \mapsto s^\lambda = (s_1^\lambda, \dots, s_{2\ell+1}^\lambda)$$

such that

- (2a) the s^λ at vertices $\lambda \in \Delta_2$ are generic strings;
- (2b) the restrictions of S to edges of Δ_2 are generic 1-chains;
- (2c) for each i the map

$$(0, 1) \rightarrow \text{int}\Delta_2 \rightarrow Q \text{ resp. } N, \quad (t, \lambda) \mapsto s_i^\lambda((1-t)a_{i-1}^\lambda + ta_i^\lambda)$$

is transverse to K ; moreover, we assume that the projection of the preimage of K to Δ_2 is an immersed submanifold $D_i \subset \Delta_2$ with transverse double points;

- (2d) for all i, j the submanifolds $D_i, D_j \subset \Delta_2$ from
- (2c) meet transversely in finitely many points;
- (2e) for each i the map

$$\text{int}\Delta_2 \rightarrow \mathbb{R}^2, \quad \lambda \in \Delta_2 \mapsto \sigma_i^\lambda(a_i^\lambda)$$

meets 0 transversely in finitely many points satisfying $(\sigma_i^\lambda)^{(3)}(a_i^\lambda) \notin 0$; moreover, these points do not meet the D_j .

We will see in the next subsection that the points in (2e) are limit points of both D_i and D_{i+1} .

5.3. STRING OPERATIONS. — Now we define the relevant operations on generic chains of broken strings. Let ∂ denote the singular boundary operator, thus

$$\partial f s^\lambda g := s^1 - s^0, \quad \partial S := S|_{\partial \Delta_2}$$

for 1- resp. 2-chains. For the definition of string coproducts we need the following

DEFINITION 5.4. — Let $p(t) = \sum_{k=1}^m p_k t^k$ and $q(t) = \sum_{k=1}^m q_k t^k$, $p_k, q_k \in \mathbb{R}^2$, be real polynomials with $\langle p_1, q_1 \rangle < 0$. A *spike* with ends (p, q) is a C^m -function $f : [a, b] \rightarrow D^2$ with the following properties (see Figure 5.1):

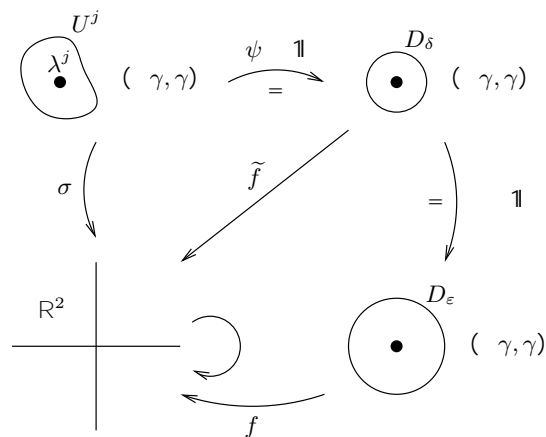


FIGURE 5.2. Construction of the map f .

- (S1) the Taylor polynomials to order m of f at a resp. b agree with p resp. q ;
- (S2) $hf(t), p_1 i > 0$ and $hf(t), q_1 i < 0$ for all $t \geq (a, b)$.

REMARK 5.5. — Note that the spikes with fixed ends (p, q) and fixed or varying $a < b$ form a convex (hence contractible) space.

We choose a family of preferred spikes $s_{p,q} : [0, 1] \rightarrow D^2$ for all (p, q) depending smoothly (with respect to the C^m -topology) on the coefficients of p and q . Now we are ready to define the string coproducts δ_N, δ_Q on generic d -chains for $d \leq 2$.

$d = 0$. — On 0-chains set $\delta_N = \delta_Q = 0$.

$d = 1$. — For a 1-chain $f s^\lambda g_{\lambda \in [0,1]}$ let (λ^j, b^j) be the finitely many values for which $s_{2i}^{\lambda^j}(b^j) \in K$ for some $i = i(j)$. Set

$$\delta_Q f s^\lambda g := \sum_j \varepsilon^j \left(s_1^{\lambda^j}, \dots, s_{2i}^{\lambda^j} J_{[a_{2i-1}, b^j]}, S^j, \widehat{s}_{2i}^{\lambda^j} J_{[b^j, a_{2i}]}, \dots, \widehat{s}_{2\ell+1}^{\lambda^j} \right),$$

where $s^j = s(\cdot, b^j) : [b^j, b^j + 1] \rightarrow N$ is a shift of the preferred spike s with ends $(r T^m \sigma_{2i}^{\lambda^j}(b_j), T^m \sigma_{2i}^{\lambda^j}(b_j))$ in the normal directions, with constant value $s_{2i}^{\lambda^j}(b^j)$ along K . The hat means shift by 1 in the argument, and $\varepsilon^j = \pm 1$ are the signs defined in Figure 2.2. Loosely speaking, δ_Q inserts an N -spike at all points where some Q -string meets K . The operation δ_N is defined analogously, inserting a Q -spike where an N -string meets K . Note that by Definition 5.4 the spikes stay in N and meet K only at their end points.

$d = 2$. — Finally, consider a generic 2-chain $S : \Delta_2 \rightarrow \Sigma^\ell$. Let $\lambda^j \in \text{int} \Delta_2$ be the finitely many points where $\sigma_i^{\lambda^j}(a_i^{\lambda^j}) = 0$ for some $i = i(j)$. For the following construction see Figure 5.2. Let $\delta > 0$ be a number ≤ 1 such that the map $\psi : \lambda \rightarrow \sigma_i^\lambda(a_i^\lambda)$ is a diffeomorphism from a neighborhood U^j of λ^j onto the δ -disk $D_\delta \subset \mathbb{R}^2$ (such δ exists by condition (2e) in Section 5.2). We choose U^j so small that it contains no other λ^i .

Let $\gamma > 0$ be a number ≤ 1 such that $j\sigma^\lambda(t + a_i^\lambda)j \leq 1$ for all $jtj \leq \gamma$. Consider the function $\sigma : U^j \subset (\gamma, \gamma) \times \mathbb{R}^2$ defined by

$$\sigma(\lambda, t) := \begin{cases} \sigma_i^\lambda(t + a_i^\lambda) & : t < 0, \\ \sigma_{i+1}^\lambda(t + a_i^\lambda) & : t > 0 \text{ if } i \text{ is even,} \\ \sigma_{i+1}^\lambda(t + a_i^\lambda) & : t > 0 \text{ if } i \text{ is odd.} \end{cases}$$

According to conditions (5.1), the function $\sigma(\lambda, t)$ is smooth in λ and of class C^2 but not C^3 in t . Define the function

$$\tilde{f} : D_\delta \subset (\gamma, \gamma) \times \mathbb{R}^2, \quad (a, b, t) \mapsto \sigma(\psi^{-1}(a, b), t).$$

By construction we have $\frac{\partial \tilde{f}}{\partial t}(a, b, 0) = (a, b)$ for all (a, b) . Moreover, by condition (2e) in Section 5.2 we have $v^j := (\sigma_i^{\lambda^j})^{(3)}(a_i^{\lambda^j}) \neq 0$. Let Ψ be the rotation of \mathbb{R}^2 which maps v^j onto a vector $(\mu, 0)$ with $\mu > 0$, let $\Phi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ be multiplication by $6/\mu$, and set $\varepsilon := 6\delta/\mu$. Then the map

$$(5.2) \quad f := \Phi \circ \Psi \circ \tilde{f} \circ (\Phi^{-1} \circ \mathbb{1}) : D_\varepsilon \subset (\gamma, \gamma) \times \mathbb{R}^2$$

satisfies

$$\begin{aligned} f(a, b, 0) &= (0, 0), & \frac{\partial f}{\partial t}(a, b, 0) &= (a, b), \\ \frac{\partial^2 f}{\partial t^2}(a, b, 0) &= (0, 0), & \frac{\partial f^3}{\partial t^3}(0, 0, 0) &= (6, 0) \end{aligned}$$

for all (a, b) . Here the map f is C^2 but not C^3 , and the statement about the third derivative $\partial f^3/\partial t^3(0, 0, 0)$ means that it equals $(6, 0)$ from the left and $(-6, 0)$ from the right. Therefore, f has a Taylor expansion (again considered for $t \leq 0$ and $t > 0$ separately)

$$(5.3) \quad f(a, b, t) = \left(at - \operatorname{sgn}(t)t^3, bt \right) + O(|a|jt^3 + |b|jt^3 + jt^4).$$

Here to simplify notation we tacitly assume that the restrictions of f to $t \leq 0$ and $t > 0$ are C^4 rather than C^3 . The following argument carries over to the C^3 case if we replace throughout $O(jt^4)$ by $o(jt^3)$.

Consider first the model case without higher order terms, i.e., the function

$$f^0(a, b, t) = \left(at - \operatorname{sgn}(t)t^3, bt \right).$$

Note that the first component $at - \operatorname{sgn}(t)t^3$ of f^0 is exactly the function that we encountered at the end of Section 4.3. The zero set of f^0 consists of three strata

$$\tilde{t} = 0g \cap \tilde{t}b = 0, a > 0, t = \sqrt[3]{a} > 0g \cap \tilde{t}b = 0, a < 0, t = -\sqrt[3]{a} < 0g.$$

For $a > 0$ and $b = 0$ the function

$$f_a : [0, \sqrt[3]{a}] \rightarrow \mathbb{R}^2, \quad t \mapsto f^0(a, 0, t) = (at - t^3, 0)$$

is a spike with ends satisfying

$$f_a^0(0) = (a, 0), \quad f_a^0(\sqrt[3]{a}) = (2a, 0), \quad f_a^{000}(0) = f_a^{000}(\sqrt[3]{a}) = 6.$$

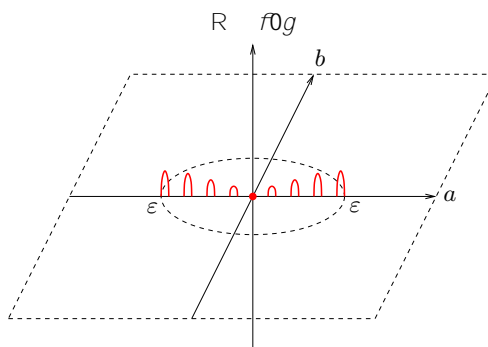


FIGURE 5.3. Two families of spikes vanishing at the origin.

Similarly, for $a < 0$ the function

$$f_a : [-a, 0] \rightarrow \mathbb{R}^2, \quad t \mapsto f^0(a, 0, t) = (at + t^3, 0)$$

is a spike with ends satisfying

$$f_a^0(0) = (a, 0), \quad f_a^0(-a) = (-2a, 0), \quad f_a^{000}(0) = f_a^{000}(-a) = +6.$$

So two families of spikes pointing in the same directions come together from both sides along the a -axis $tb = 0g$ and vanish at $(a, b) = (0, 0)$, see Figure 5.3. The following lemma states that this qualitative picture persists in the presence of higher order terms.

LEMMA 5.6. — *Let $f : D_\varepsilon \subset (\gamma, \gamma) \rightarrow \mathbb{R}^2$ be a function satisfying (5.3). Then for ε and γ sufficiently small there exist smooth functions $\beta(a, t)$ and $\tau(a)$ for $a \in [-\varepsilon, \varepsilon] \cap f0g$ such that with $\beta(a) := \beta(a, \tau(a))$ the zero set of f in $D_\varepsilon \subset (\gamma, \gamma)$ consists of three strata*

$$ft = 0g \cup [tb = \beta(a), a > 0, t = \tau(a) > 0g \cup [tb = \beta(a), a < 0, t = \tau(a) < 0g.$$

The functions β, τ satisfy the estimates

$$\beta(a, t) = O(|ajt^2 + jt^3|), \quad \tau(a)^2 - a = O(a^{3/2}), \quad \beta(a) = O(a^{3/2}).$$

Moreover, the functions

$$\begin{aligned} f_a : [0, \tau(a)] &\rightarrow \mathbb{R}^2, & t \mapsto f(a, \beta(a), t), & a > 0, \\ f_a : [\tau(a), 0] &\rightarrow \mathbb{R}^2, & t \mapsto f(a, \beta(a), t), & a < 0 \end{aligned}$$

are spikes with ends satisfying

$$f_a^0(0) = (a, 0) + O(a^{3/2}), \quad f_a^0(\tau(a)) = (-2a, 0) + O(a^{3/2}).$$

Proof. — We consider the case $a, t > 0$, the case $a, t < 0$ being analogous. Setting the second component in (5.3) to zero and dividing by t yields $b = O(at^2 + bt^2 + t^3)$, which for t sufficiently small can be solved for $b = \beta(a, t)$ satisfying the estimate

$\beta(a, t) = O(at^2 + t^3)$. Inserting this into the first component in (5.3), setting it to zero and dividing by t yields

$$a - t^2 = O(at^2 + \beta(a, t)t^2 + t^3) = O(at^2 + t^3),$$

which for (a, t) sufficiently small can be solved for $t = \tau(a)$ satisfying the estimate $\tau(a)^2 - a = O(a^{3/2})$. Inserting $t = \tau(a)$ in $\beta(a, t)$ we obtain the estimate $\beta(a) = O(a^{3/2})$. This proves the first assertions.

Now consider the function $f_a(t) = f(a, \beta(a), t)$ for $t \geq [0, \tau(a)]$ and $a > 0$. Inserting $\beta(a) = O(a^{3/2})$ we find

$$\begin{aligned} f_a(t) &= (at - t^3, 0) + O(\beta(a)t + at^3 + \beta(a)t^3 + t^4) \\ &= (at - t^3, 0) + O(a^{3/2}t + at^3 + t^4), \end{aligned}$$

and therefore $f_a^0(t) = (a - 3t^2, 0) + O(a^{3/2} + at^2 + t^3)$. This immediately gives $f_a^0(0) = (a, 0) + O(a^{3/2})$ and, using $\tau(a) = O(a^{1/2})$, also $f_a^0(\tau(a)) = (-2a, 0) + O(a^{3/2})$.

It remains to prove that the functions f_a are spikes in the sense of Definition 5.4. Write in components $f = (f^1, f^2)$ and $f_a = (f_a^1, f_a^2)$ and abbreviate $\tau := \tau(a)$. We claim that there exist constants $\delta, D > 0$ independent of a, t such that for all $t \geq [0, \tau]$ we have

$$f_a^1(t) > 2\delta t(\tau^2 - t^2), \quad |f_a^2(t)| \leq Dt(a+t)(\tau - t).$$

For the first estimate, note that

$$\frac{1}{t} f_a^1(t) = a - t^2 + O(a^{3/2} + at^2 + t^3),$$

viewed as a function of t^2 , has transversely cut out zero locus $t = \tau$ and is therefore $> 2\delta(\tau^2 - t^2)$ for some $\delta > 0$. The second estimate holds because

$$\frac{1}{t} f_a^2(t) = O(a^{3/2} + at^2 + t^3)$$

vanishes at $t = \tau$, so $|f_a^2(t)| \leq Dt(a+t)(\tau - t)$ for some constant D . Using these estimates as well as $f_a^0(0) = (a, 0) + O(a^{3/2})$ and $\tau = O(a^{1/2})$ we compute with a generic constant C (independent of a, t):

$$\begin{aligned} \langle hf_a^0(0), f_a(t) \rangle &= (a + O(a^{3/2}))f_a^1(t) + hO(a^{3/2}), f_a^2(t) \rangle \\ &> a\delta t(\tau^2 - t^2) - Ca^{3/2}t(\tau - t)(a+t) \\ &= at(\tau - t) \left(\delta(\tau + t) - Ca^{1/2}(a+t) \right) \\ &> a^{3/2}t(\tau - t)(\delta - C(a+t)), \end{aligned}$$

which is positive for $0 < t < \tau$ and a sufficiently small. An analogous computation, using $f_a^0(\tau) = (-2a, 0) + O(a^{3/2})$, shows $\langle hf_a^0(\tau), f_a(t) \rangle < 0$, so f_a is a spike.

REMARK 5.7. — The spikes from Lemma 5.6 can be connected to the spike of the model function f^0 without higher order terms by rescaling: For $s \in (0, 1]$ set

$$\begin{aligned} f^s(a, b, t) &:= \frac{1}{s^3} f(s^2 a, s^2 b, st) \\ &= (at - \operatorname{sgn}(t)t^3, bt) + sO(jajjtj^3 + jbjjtj^3 + jtj^4) \\ &\stackrel{s \neq 0}{\sim} (at - \operatorname{sgn}(t)t^3, bt). \end{aligned}$$

Thus for $j \in \mathbb{N}$ the corresponding family of spikes $(f_a^s)_{s \in [0, 1]}$ connects f_a to the spike f_a^0 .

Now we return to the points $\lambda^j \in U^j$ and the corresponding maps

$$f : D_\varepsilon \times (\gamma, \gamma) \rightarrow \mathbb{R}^2$$

defined by (5.2). After shrinking $\varepsilon, \gamma > 0$ and replacing U^j by $(\Phi \circ \psi)^{-1}(D_\varepsilon) \cap \Delta_2$ (where ψ, Φ are the maps defined above), we may assume that ε, γ satisfy the smallness requirement in Lemma 5.6 for each j . Define

$$\begin{aligned} M_{\tilde{\delta}_Q} &:= \bigcup_i (\operatorname{ev}_{2i}^{-1} S)^{-1}(K) \cap \bigcup_j (U^j \cap (0, 1)), \\ M_{\tilde{\delta}_N} &:= \bigcup_i (\operatorname{ev}_{2i-1}^{-1} S)^{-1}(K) \cap \bigcup_j (U^j \cap (0, 1)). \end{aligned}$$

By construction, $M_{\tilde{\delta}_Q}$ and $M_{\tilde{\delta}_N}$ are 1-dimensional submanifolds with boundary of $\Delta_2 \cap (0, 1)$. Define $\tilde{\delta}_Q S : M_{\tilde{\delta}_Q} \rightarrow \Sigma^{\ell+1}$ by inserting preferred N -spikes at all points where some Q -string meets K (via the same formula as the one above for δ_Q on 1-chains), and similarly for $\tilde{\delta}_N S$. See Figure 5.4.

Note that the boundary $\partial M_{\tilde{\delta}_Q}$ consists of intersections with $\partial \Delta_2$ and with the boundaries ∂U^j . Thus each j contributes a unique point λ_Q^j to $\partial M_{\tilde{\delta}_Q}$, which corresponds in the above coordinates to $a = +\varepsilon$ if the associated index i is odd and to $a = -\varepsilon$ if i is even. Similarly, each j contributes a unique point λ_N^j to $\partial M_{\tilde{\delta}_N}$ which corresponds in the above coordinates to $a = -\varepsilon$ if the associated index i is odd and to $a = +\varepsilon$ if i is even. The broken strings $\tilde{\delta}_Q S(\lambda_Q^j)$ and $\tilde{\delta}_N S(\lambda_N^j)$ are C^m -close for $jtj > \gamma$, and by Lemma 5.6 for $jtj < \gamma$ they both have a Q -spike and an N -spike with the same first derivatives at the ends. So, using convexity of the space of spikes with fixed ends (Remark 5.5, see also Remark 5.7), we can connect them by a short 1-chain $S^j : [0, 1] \rightarrow \Sigma^{\ell+1}$ with spikes in $[\gamma, \gamma]$ (which we regard as Q -spikes.)

We define $\delta_Q S : M_{\delta_Q} \rightarrow \Sigma^{\ell+1}$ to be $\tilde{\delta}_Q S$ together with the 1-chains S^j , and we set $\delta_N S := \tilde{\delta}_N S : M_{\delta_N} = M_{\tilde{\delta}_N} \rightarrow \Sigma^{\ell+1}$. Recall that the 1-dimensional submanifold $M_{\tilde{\delta}_Q} \cap \Delta_2 \cap (0, 1)$ is the union of the transversely cut out preimages of $K \cap Q$ under the evaluation maps $\operatorname{ev}_{2i}^{-1} S : \Delta_2 \cap (0, 1) \rightarrow Q$. Hence the coorientation of $K \cap Q$ and the orientation of $\Delta_2 \cap (0, 1)$ induce an orientation on $M_{\tilde{\delta}_Q}$, and similarly for $M_{\tilde{\delta}_N}$. (The induced orientations depend on orientation conventions which will be fixed in the proof of Proposition 5.8 below.) We parametrize each connected component of $M_{\tilde{\delta}_Q}$ and $M_{\tilde{\delta}_N}$ by the interval $\Delta_1 = [0, 1]$ proportionally to arclength (with respect to

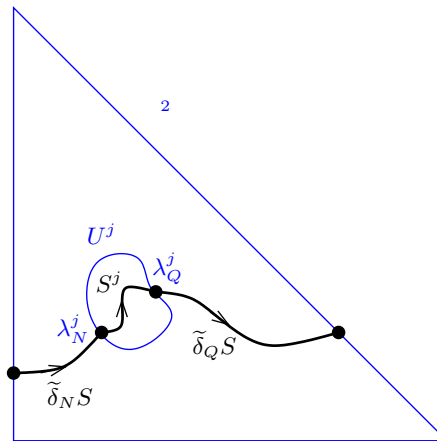


FIGURE 5.4. The definition of $\delta_Q S = \tilde{\delta}_Q S + S^j$ and $\delta_N S = \tilde{\delta}_N S$.

the standard metric on $\Delta_2 \setminus (0, 1)$ and in the direction of the orientation, where for components diffeomorphic to S^1 we choose an arbitrary initial point). So we can view $\delta_Q S : M_{\tilde{\delta}_Q} \rightarrow \Sigma^{\ell+1}$ and $\delta_N S : M_{\tilde{\delta}_N} \rightarrow \Sigma^{\ell+1}$ as generic 1-chains, where we orient the 1-chains S^j such that the points $\tilde{\delta}_Q S(\lambda_Q^j)$ appear with opposite signs in the boundary of S^j and $M_{\tilde{\delta}_Q}$.

PROPOSITION 5.8. — *On generic chains of degree 2, the operations ∂ , δ_Q and δ_N satisfy the relations*

$$\begin{aligned} \partial^2 &= \delta_Q^2 = \delta_N^2 = \delta_Q \delta_N + \delta_N \delta_Q = 0, \\ \partial \delta_Q + \delta_Q \partial + \partial \delta_N + \delta_N \partial &= 0. \end{aligned}$$

In particular, these relations imply

$$(\partial + \delta_Q + \delta_N)^2 = 0.$$

Proof. — Consider a generic 2-chain $S : \Delta_2 \rightarrow \Sigma^\ell$. We continue to use the notation above and denote by $\pi : \Delta_2 \setminus (0, 1) \rightarrow \Delta_2$ the projection. The relation $\partial^2 S = 0$ is clear. Points in $\delta_Q^2 S$ correspond to transverse self-intersections of $\pi(M_{\tilde{\delta}_Q})$, so each point appears twice with opposite signs, hence $\delta_Q^2 S = 0$ and similarly $\delta_N^2 S = 0$. Points in $\delta_Q \delta_N S + \delta_N \delta_Q S$ correspond to transverse intersections of $\pi(M_{\tilde{\delta}_Q})$ and $\pi(M_{\tilde{\delta}_N})$, so again each point appears twice with opposite signs and the expression vanishes. Note that the broken strings corresponding to these points have two preferred spikes inserted at different places, so due to the uniqueness of preferred spikes with given end points the broken strings do not depend on the order in which the spikes are inserted.

In order to achieve $\partial \delta_Q + \delta_Q \partial + \partial \delta_N + \delta_N \partial = 0$, we choose the orientation conventions for $M_{\tilde{\delta}_Q}$ and $M_{\tilde{\delta}_N}$ such that (see Figure 5.4):

- $\partial\tilde{\delta}_Q S + \tilde{\delta}_Q \partial S$ corresponds to the intersection points λ_Q^j of $M_{\tilde{\delta}_Q}$ with the boundaries of the regions U^j , and similarly for $\partial\tilde{\delta}_N S + \tilde{\delta}_N \partial S$;
- the sign of λ_Q^j as a boundary point of $M_{\tilde{\delta}_Q}$ is opposite to the sign of λ_N^j as a boundary point of $M_{\tilde{\delta}_N}$.

Due to the choice of the 1-chains S^j , it follows that $\partial\tilde{\delta}_Q S + \tilde{\delta}_Q \partial S$ is the sum of the points $\delta_N S(\lambda_N^j)$ with suitable signs, and $\partial\tilde{\delta}_N S + \tilde{\delta}_N \partial S$ is the same sum with opposite signs, so the total sum equals zero.

5.4. THE STRING CHAIN COMPLEX. — For $d = 0, 1, 2$ and $\ell > 0$ let $C_d(\Sigma^\ell)$ be the free \mathbb{Z} -module generated by generic d -chains in Σ^ℓ , and set

$$C_d(\Sigma) := \bigoplus_{\ell=0}^1 C_d(\Sigma^\ell), \quad d = 0, 1, 2.$$

The string operations defined in Section 5.3 yield \mathbb{Z} -linear maps

$$\partial : C_d(\Sigma^\ell) \rightarrow C_{d-1}(\Sigma^\ell), \quad \delta_N, \delta_Q : C_d(\Sigma^\ell) \rightarrow C_{d-1}(\Sigma^{\ell+1}).$$

The induced maps $\partial, \delta_Q, \delta_N : C_d(\Sigma) \rightarrow C_{d-1}(\Sigma)$ satisfy the relations in Proposition 5.8, in particular

$$D := \partial + \delta_Q + \delta_N.$$

satisfies $D^2 = 0$. We call $(C(\Sigma), \partial + \delta_Q + \delta_N)$ the *string chain complex* of K , and we define the *degree d string homology* of K as the homology of the resulting complex,

$$H_d^{\text{string}}(K) := H_d(C(\Sigma), \partial + \delta_Q + \delta_N), \quad d = 0, 1, 2.$$

Concatenation of broken strings at the base point x_0 (and the canonical subdivision of $\Delta_1 \times \Delta_1$ into two 2-simplices) yields products

$$C_d(\Sigma^\ell) \times C_{d^0}(\Sigma^{\ell^0}) \rightarrow C_{d+d^0}(\Sigma^{\ell+\ell^0}), \quad d + d^0 \leq 2$$

satisfying the relations

$$(5.4) \quad (a \cdot b) \cdot c = a \cdot (b \cdot c), \quad D(a \cdot b) = Da \cdot b + (-1)^{\deg a} a \cdot Db$$

whenever $\deg a + \deg b + \deg c \leq 2$. In particular, this gives $C_0(\Sigma)$ the structure of a (noncommutative but strictly associative) algebra over \mathbb{Z} and $C_1(\Sigma), C_2(\Sigma)$ the structure of bimodules over this algebra. These structures induce on homology the structure of a \mathbb{Z} -algebra on $H_0^{\text{string}}(K)$, and of bimodules over this algebra on $H_1^{\text{string}}(K)$ and $H_2^{\text{string}}(K)$. By definition, the isomorphism classes of the algebra $H_0^{\text{string}}(K)$ and the modules $H_1^{\text{string}}(K), H_2^{\text{string}}(K)$ are clearly isotopy invariants of the framed oriented knot K .

We can combine these invariants into a single graded algebra as follows. For $d > 2$, we define $C_d(\Sigma^\ell)$ to be the free \mathbb{Z} -module generated by products $S_1 \cdots S_r$ of generic chains S_i of degrees $1 \leq d_i \leq 2$ in Σ^{ℓ_i} such that $d_1 + \cdots + d_r = d$ and $\ell_1 + \cdots + \ell_r = \ell$, modulo the submodule generated by

$$S_1 \cdots S_r - S_1^0 \cdots S_r^0$$

for different decompositions of the same d -chain. Put differently, this submodule is generated by

$$S_1 \quad S_i \quad S_{i+1} \quad S_r \quad S_1 \quad (S_i \quad S_{i+1}) \quad S_r,$$

where S_i and S_{i+1} are generic 1-chains and $(S_i \quad S_{i+1})$ is the associated generic 2-chain. Note that for $d = 2$ this definition of $C_2(\Sigma^\ell)$ agrees with the earlier one. We define $D = \partial + \delta_Q + \delta_N$ on

$$C_d(\Sigma) := \bigoplus_{\ell=0}^1 C_d(\Sigma^\ell), \quad d > 0$$

by the Leibniz rule. This is well-defined in view of the second equation in (5.4) and satisfies $D^2 = 0$. Together with the product this gives $C(\Sigma)$ the structure of a differential graded \mathbb{Z} -algebra. The *total string homology*

$$H^{\text{string}}(K) := H(C(\Sigma), D)$$

inherits the structure of a graded \mathbb{Z} -algebra whose isomorphism class is an invariant of the framed oriented knot K .

REMARK 5.9. — Our definition of string homology of K in degrees > 2 in terms of product chains is motivated by Legendrian contact homology of ΛK when $Q = \mathbb{R}^3$ which is then generated by elements of degrees ≤ 2 . From the point of view of string topology, it would appear more natural to define string homology in arbitrary degrees in terms of higher dimensional generic chains of broken strings in the sense of Definition 5.3. Similarly, for knot contact homology in other ambient manifolds, e.g. for $Q = S^3$, there are higher degree Reeb chords that contribute to the (linearized) contact homology. It would be interesting to see whether such constructions would carry additional information.

5.5. LENGTH FILTRATION. — Up to this point, the constructions have been fairly symmetric in the Q - and N -strings. However, as we will see below, the relation to Legendrian contact homology leads us to assign to Q -strings s_{2i} their geometric length $L(s_{2i})$, and to N -strings length zero. Thus we define the *length* of a broken string $s = (s_1, \dots, s_{2\ell+1})$ by

$$L(s) := \sum_{i=1}^{\ell} L(s_{2i}),$$

where we do not include in the sum those s_{2i} that are Q -spikes in the sense of Definition 5.4. We define the length of a generic i -chain $S : K \rightarrow \Sigma$ by

$$L(S) := \max_{k \in K} L(S(k)).$$

Then the subspaces

$$F^\ell C_i(\Sigma) := \left\{ \sum a_j S_j \in C_i(\Sigma) \mid L(S_j) \leq \ell \text{ whenever } a_j \neq 0 \right\}$$

define a filtration in the sense that $F^k C_i(\Sigma) \subset F^\ell C_i(\Sigma)$ for $k \leq \ell$ and

$$D(F^\ell C_i(\Sigma)) \subset F^\ell C_{i-1}(\Sigma).$$

This *length filtration* will play an important role in the proof of the isomorphism to Legendrian contact homology in Section 7.

REMARK 5.10. — The omission of the length of Q -spikes from the length of a broken string ensures that the operation δ_N , which inserts Q -spikes, does not increase the length. Since Q -spikes do not intersect the knot in their interior, they are not affected by δ_Q and it follows that D preserves the length filtration.

6. THE CHAIN MAP FROM LEGENDRIAN CONTACT HOMOLOGY TO STRING HOMOLOGY

In this section we define a chain map $\Phi: C(R) \rightarrow C(\Sigma)$ from a complex computing Legendrian contact homology to the string chain complex defined in the previous section. The boundary operator on $C(R)$ is defined using moduli spaces of holomorphic disks in $\mathbb{R} \times S^1 \times Q$ with Lagrangian boundary condition $\mathbb{R} \times \Lambda_K$ and the map Φ is defined using moduli spaces of holomorphic disks in T^*Q with Lagrangian boundary condition $Q \sqcup L_K$, where the boundary is allowed to switch back and forth between the two irreducible components of the Lagrangian at corners as in Lagrangian intersection Floer homology. We will describe these spaces and their properties, as well as define the algebra and the chain map. In order not to obscure the main lines of argument, we postpone the technicalities involved in detailed proofs to Sections 8–10.

6.1. HOLOMORPHIC DISKS IN THE SYMPLECTIZATION. — Consider a contact $(2n - 1)$ -manifold (M, λ) with a closed Legendrian $(n - 1)$ -submanifold Λ . For the purposes of this paper we only consider the case that $M = S^1 \times Q$ is the cosphere bundle of $Q = \mathbb{R}^3$ with its standard contact form $\lambda = p dq$ and $\Lambda = \Lambda_K$ is the unit conormal bundle of an oriented framed knot $K \subset Q$, but the construction works more generally for any pair (M, Λ) for which M has no contractible closed Reeb orbits, see Remark 6.4 below.

Denote by R the Reeb vector field of λ . A *Reeb chord* is a solution $a: [0, T] \rightarrow M$ of $\dot{a} = R$ with $a(0), a(T) \in \Lambda$. Reeb chords correspond bijectively to *binormal chords* of K , i.e., geodesic segments meeting K orthogonally at their endpoints. As usual, we assume throughout that Λ is chord generic, i.e., each Reeb chord corresponds to a Morse critical point of the distance function on $K \times K$.

In order to define Maslov indices, one usually chooses for each Reeb chord $a: [0, T] \rightarrow M$ capping paths connecting $a(0)$ and $a(T)$ in Λ to a base point $x_0 \in \Lambda$. Then one can assign to each a completed by the capping paths a *Maslov index* $\mu(a)$, see [5, App. A]. In the case under consideration ($M = S^1 \times \mathbb{R}^3$ and $\Lambda = \Lambda_K$) the Maslov class of Λ equals 0, so the Maslov index does not depend on the choice of capping paths. It is given by $\mu(a) = \text{ind}(a) + 1$, where $\text{ind}(a)$ equals the index of a as a critical point of the distance function on $K \times K$, see [12]. We define the *degree* of a Reeb chord a as

$$|a| := \mu(a) - 1 = \text{ind}(a),$$

and the degree of a word $\mathbf{b} = b_1 b_2 \dots b_m$ of Reeb chords as

$$|\mathbf{b}| := \sum_{j=1}^m |b_j|.$$

Given a and \mathbf{b} , we write $\mathcal{M}^{\text{SY}}(a; \mathbf{b})$ for the moduli space of J -holomorphic disks $u: (D, \partial D) \rightarrow (R \times M, R \times \Lambda)$ with one positive boundary puncture asymptotic to the Reeb chord strip over a at the positive end of the symplectization, and m negative boundary punctures asymptotic to the Reeb chord strips over b_1, \dots, b_m at the negative end of the symplectization. Here J is an R -invariant almost complex structure on $R \times M$ compatible with λ . For generic J , the moduli space $\mathcal{M}^{\text{SY}}(a; \mathbf{b})$ is a manifold of dimension

$$\dim(\mathcal{M}^{\text{SY}}(a; \mathbf{b})) = |a| - |\mathbf{b}| = |a| - \sum_{j=1}^m |b_j|,$$

see Theorem 10.1. In fact, the moduli spaces correspond to the zero set of a Fredholm section of a Banach bundle that can be made transverse by perturbing the almost complex structure, and there exist a system of coherent (or gluing compatible) orientations of the corresponding index bundles over the configuration spaces and this system induces orientations on all the moduli spaces.

By our choice of almost complex structure, R acts on $\mathcal{M}^{\text{SY}}(a; \mathbf{b})$ by translations in the target $R \times M$ and we write $\mathcal{M}^{\text{SY}}(a; \mathbf{b})/R$ for the quotient, which is then an oriented manifold of dimension $|a| - |\mathbf{b}| - 1$.

Finally, we discuss the compactness properties of $\mathcal{M}^{\text{SY}}(a; \mathbf{b})/R$. The moduli space $\mathcal{M}^{\text{SY}}(a; \mathbf{b})/R$ is generally not compact but admits a compactification by multilevel disks, where a multilevel disk is a tree of disks with a top level disk in $\mathcal{M}^{\text{SY}}(a; \mathbf{b}^1)$, $\mathbf{b}^1 = b_1^1 \dots b_{m_1}^1$, second level disks in $\mathcal{M}^{\text{SY}}(b_i^1; \mathbf{b}^{2,i})$ attached at the negative punctures of the top level disk, etc. See Figure 6.9 below. It follows from the dimension formula above that the formal dimension of the total disk that is the union of the levels in a multilevel disk is the sum of dimensions of all its components. Consequently, for generic almost complex structure, if $\dim(\mathcal{M}^{\text{SY}}(a; \mathbf{b})) = 1$ then $\mathcal{M}^{\text{SY}}(a; \mathbf{b})/R$ is a compact 0-dimensional manifold, and if $\dim(\mathcal{M}^{\text{SY}}(a; \mathbf{b})) = 2$ then the boundary of $\mathcal{M}^{\text{SY}}(a; \mathbf{b})/R$ consists of two-level disks where each level is a disk of dimension 1 (and possibly trivial Reeb chord strips).

The simplest version of Legendrian contact homology would be defined by the free \mathbb{Z} -algebra generated by the Reeb chords, with differential counting rigid holomorphic disks. In the following subsection we will define a refined version which also incorporates the boundary information of holomorphic disks.

6.2. LEGENDRIAN CONTACT HOMOLOGY. — In this subsection we define a version of Legendrian contact homology that will be directly related to the string homology of Section 5, see [10] for a similar construction in rational symplectic field theory. The usual definition of Legendrian contact homology is a quotient of our version. We keep the notation from Section 6.1.

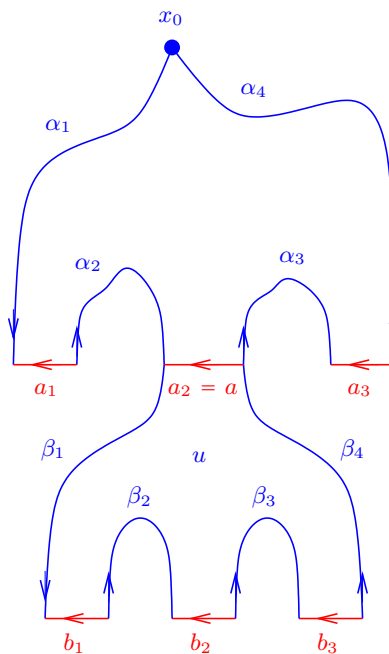


FIGURE 6.1. The definition of $\partial(u)$ and $\partial(u) \cdot \mathbf{a}$.

Fix an integer $m > 3$. For points $x, y \in \Lambda$ we denote by $P_{x,y}\Lambda$ the space of C^m paths $\gamma : [a, b] \rightarrow \Lambda$ with $\gamma(a) = x$ and $\gamma(b) = y$ whose first m derivatives vanish at the endpoints. Here the interval $[a, b]$ is allowed to vary. The condition at the endpoints ensures that concatenation of such paths yields again C^m paths. Fix a base point $x_0 \in \Lambda$ and denote by $\Omega_{x_0}\Lambda = P_{x_0,x_0}\Lambda$ the Moore loop space based at x_0 .

DEFINITION 6.1. — A *Reeb string with ℓ chords* is an expression $\alpha_1 a_1 \alpha_2 a_2 \dots \alpha_\ell a_\ell \alpha_{\ell+1}$, where the $a_i : [0, T_i] \rightarrow M$ are Reeb chords and the α_i are elements in the path spaces

$$\alpha_1 \in P_{x_0 a_1(T_1)}, \quad \alpha_i \in P_{a_{i-1}(0) a_i(T_i)} \text{ for } 2 \leq i \leq \ell, \quad \alpha_{\ell+1} \in P_{a_\ell(0) x_0}.$$

See the top of Figure 6.1. Note that the α_i and the *negatively traversed* Reeb chords a_i fit together to define a loop in M starting and ending at x_0 . Concatenating all the α_i and a_i in a Reeb string with the appropriate capping paths, we can view each α_i as an element in the based loop space $\Omega_{x_0}\Lambda$. However, we will usually not take this point of view.

Boundaries of holomorphic disks in the symplectization give rise to Reeb strings as follows. Consider a holomorphic disk u belonging to a moduli space $\mathcal{M}^{\text{sy}}(\mathbf{a}; \mathbf{b})$ as above, with Reeb chords $a : [0, T] \rightarrow M$ and $b_i : [0, T_i] \rightarrow M$, $i = 1, \dots, \ell$. Its boundary arcs in counterclockwise order and orientation projected to Λ define paths $\beta_1, \dots, \beta_\ell$ in Λ as shown in Figure 6.1, i.e.,

$$\beta_1 \in P_{a(T) b_1(T_1)}, \quad \beta_i \in P_{b_{i-1}(0) b_i(T_i)} \text{ for } 2 \leq i \leq \ell, \quad \beta_{\ell+1} \in P_{b_\ell(0) a(0)}.$$

We denote the alternating word of paths and Reeb chords obtained in this way as the boundary of u by

$$(6.1) \quad \partial(u) := \beta_1 b_1 \beta_2 b_2 \dots \beta_\ell b_\ell \beta_{\ell+1}.$$

Note that the β_i and the negatively traversed Reeb chords b_i fit together to define a path in M from $a(T)$ to $a(0)$. We obtain from $\partial(u)$ a Reeb string if we extend β_1 and $\beta_{\ell+1}$ to the base point x_0 by the capping paths of a .

For $\ell > 0$ we denote by R^ℓ the space of Reeb strings with ℓ chords, equipped with the C^m topology on the path spaces. Note that different collections of Reeb chords correspond to different components. Concatenation at the base point gives

$$R := \coprod_{\ell > 0} R^\ell$$

the structure of an H-space. Note that the sub-H-space $R^0 = \Omega_{x_0} \Lambda$ agrees with the Moore based loop space with its Pontrjagin product. Let

$$C(R) = \bigoplus_{d > 0} C_d(R)$$

be singular chains in R with integer coefficients. It carries two gradings: the degree d as a singular chain, which we will refer to as the *chain degree*, and the degree $\sum_{i=1}^{\ell} j b_{i,j}$ of the Reeb chords, which we will refer to as the *chord degree*. For sign rules we think of the *chain coming first and the Reeb chords last*. The total grading is given by the sum of the two degrees. Recall that it does not depend on the choice of capping paths. Concatenation of Reeb strings at the base point and product of chains gives $C(R)$ the structure of a (noncommutative but strictly associative) graded ring. Note that it contains the subring

$$C(R^0) = C(\Omega_{x_0} \Lambda).$$

Next we define the differential

$$\partial = \partial^{\text{sing}} + \partial^{\text{sy}}: C(R) \rightarrow C(R).$$

Here ∂^{sing} is the singular boundary and ∂^{sy} is defined as follows. Pick a generic compatible cylindrical almost complex structure J on the symplectization $\mathbb{R} \times M$. Consider a punctured J -holomorphic disk $u: D \rightarrow \mathbb{R} \times M$ in $\mathcal{M}^{\text{sy}}(a; \mathbf{b})$. If the Reeb chord $a = a_i$ appears in a Reeb string $\mathbf{a} = \alpha_1 a_1 \dots a_m \alpha_{m+1}$, then we can replace a_i by $\partial(u)$ to obtain a new Reeb string which we denote by

$$\partial(u)_i \mathbf{a} := \alpha_1 a_1 \dots \tilde{\alpha}_i \partial(u) \tilde{\alpha}_{i+1} \dots a_\ell \alpha_{\ell+1}.$$

Here $\partial(u)$ is defined in (6.1) and the paths $\tilde{\alpha}_i, \tilde{\alpha}_{i+1}$ are the concatenations of α_i, α_{i+1} with the paths $\beta_1, \beta_{\ell+1}$ in $\partial(u)$, respectively. See Figure 6.1. For a chain $\mathbf{a} \in C(R)$ of Reeb strings of type $\mathbf{a} = \alpha_1 a_1 \dots a_m \alpha_{m+1}$ we now define

$$\partial^{\text{sy}}(\mathbf{a}) := \sum_{i=1}^{\ell} \sum_{\substack{j a_i j \\ u \in \mathcal{M}^{\text{sy}}(a_i; \mathbf{b})/\mathbb{R}}} \varepsilon(-1)^{d + j a_i j + j a_i - 1} \partial(u)_i \mathbf{a},$$

where d is the chain degree of \mathbf{a} and ε is the sign from the orientation of $\mathcal{M}^{\text{sy}}(a_i; \mathbf{b})/\mathbb{R}$ as a compact oriented 0-manifold (i.e., points with signs). Note that ∂^{sy} preserves the

chain degree and decreases the chord degree by 1, whereas ∂^{sing} preserves the chord degree and decreases the chain degree by 1. In particular, ∂ has degree -1 with respect to the total grading. The main result about the contact homology algebra that we need is summarized in the following theorem.

THEOREM 6.2. — *The differential $\partial : C(R) \rightarrow C(R)$ satisfies $\partial^2 = 0$ and the Legendrian contact homology*

$$H^{\text{contact}}(\Lambda) := \ker \partial / \text{im } \partial$$

is independent of all choices.

Proof. — In the case that we use it, for $M = S^2 \times \mathbb{R}^3$ and $\Lambda = \Lambda_K$, the proof is an easy adaptation of the one in [15, 14] and [7], see also [12].

Consider first the equation for the differential. The equation $\partial^2 = 0$ follows from our description of the boundary of the moduli spaces $\mathcal{M}^{\text{SY}}(a; \mathbf{b})$ of dimension 2 in Section 6.1, which shows that contributions to $(\partial^{\text{SY}})^2$ are in oriented one-to-one correspondence with the boundary of an oriented 1-manifold and hence cancel out. The relations $(\partial^{\text{sing}})^2 = 0$ and $\partial^{\text{sing}}\partial^{\text{SY}} + \partial^{\text{SY}}\partial^{\text{sing}} = 0$ are clear.

To prove the invariance statement we use a bifurcation method similar to [14, §4.3]. Consider a generic 1-parameter family (Λ_s, J_s) , $s \in S = [0, 1]$, of Legendrian submanifolds and almost complex structures. By genericity of the family there is a finite set of points $s_1 < s_2 < \dots < s_m$ such that in $S \setminus \{s_1, \dots, s_m\}$ all Reeb chords of Λ_s are transverse, all Reeb chords have distinct actions, and all holomorphic disks determined by (Λ_s, J_s) have dimension at least 1 (i.e., if we write $\mathcal{M}_s^{\text{SY}}$ for moduli spaces determined by (Λ_s, J_s) then $\dim \mathcal{M}_s^{\text{SY}}(a, \mathbf{b}) > 1$ if the moduli space is nonempty). Furthermore, the points s_j are of three kinds:

- handle slides, where all Reeb chords are nondegenerate but where there is a transversely cut out disk of formal dimension 0 (i.e., there exists one $\mathcal{M}^{\text{SY}}(a; \mathbf{b})$, with $\dim \mathcal{M}^{\text{SY}}(a; \mathbf{b}) = 0$ which contains one \mathbb{R} -family of J_{s_j} -holomorphic disks with boundary on Λ_{s_j} , and this disk is transversely cut out as a solution of the parameterized problem);
- action switches, where two nondegenerate Reeb chords have the same action and their actions interchange;
- birth/death moments where there is one degenerate Reeb chord at which two Reeb chords cancel through a quadratic tangency.

To show invariance we first observe that if $[s^\theta, s^{\theta\theta}] \subset S$ is an interval which does not contain any s_j , then the Reeb chords of Λ_s , $s \in [s^\theta, s^{\theta\theta}]$ form 1-manifolds canonically identified with $[s^\theta, s^{\theta\theta}]$ and the actions of the different Reeb chord manifolds do not cross. Thus for Reeb chords a, b_1, \dots, b_m of Λ_{s^θ} we get corresponding chords on Λ_s for each $s \in [s^\theta, s^{\theta\theta}] = S^\theta$ which we denote by the same symbols, suppressing the s -dependence below. We next define a chain map

$$\Phi: C(R_{s^\theta}) \rightarrow C(R_{s^{\theta\theta}})$$

which counts geometrically induced chains as follows. We introduce the notion of a *disk with lines of Reeb chords*. Such an object has a positive puncture at the Reeb chord a over s^0 and negative punctures at Reeb chords according to \mathbf{b} over s^{00} and is given by a collection of disks u_1, \dots, u_m where the disk u_j is a disk at σ_j , where $s^0 \subset \sigma_1 \subset \sigma_2 \subset \dots \subset \sigma_m \subset s^{00}$ and if $\sigma_j > \sigma_k$ for some k then its positive Reeb chord is connected by a line in a Reeb chord manifold to a Reeb chord at the negative puncture of some u_r for $\sigma_r < \sigma_j$. The collection of such objects naturally forms a moduli space, $\mathcal{M}_{S^0}^{\text{SY}}(a; \mathbf{b})$, where we glue two disks when the length of the line connecting them goes to zero. We define the chain map Φ as

$$\Phi(a) = [\mathcal{M}_{S^0}^{\text{SY}}(a, \mathbf{b})].$$

The chain map equation $\partial_{s^{00}} \Phi = \Phi \partial_{s^0}$ follows immediately once one notices that the codimension one boundary of the moduli space consists of disks over the endpoints with lines of Reeb chords over $[s^0, s^{00}]$ attached. (We point out that this construction is inspired by Morse-Bott arguments, compare [16].)

Consider the filtration in $C(R)$ which associates to a chain of Reeb strings the sum of actions of its Reeb chords. By Stokes' theorem the differential respects the filtration. The pure lines of Reeb chords (without disks) contribute to the map and show that

$$\Phi(a) = a + \Phi_0(a),$$

where the action of $\Phi_0(a)$ is strictly smaller than that of a . It follows that Φ induces an isomorphism on the E_2 -page of the action spectral sequence and hence is a quasi-isomorphism.

In order to show invariance at the bifurcation moments we consider the deformation in a small interval around $[s_j - \varepsilon, s_j + \varepsilon]$. In this case we can construct a Lagrangian cobordism L in the symplectization $\mathbb{R} \times M$ interpolating between the cylinders on $\Lambda_{s_j - \varepsilon}$ and $\Lambda_{s_j + \varepsilon}$, see [10, Lem. A.2]. If a is a Reeb chord of $\Lambda_{s_j + \varepsilon}$ and \mathbf{b} is a word of Reeb chords of $\Lambda_{s_j - \varepsilon}$ then let $\mathcal{M}^{\text{SY}, L}(a; \mathbf{b})$ denote the moduli space of holomorphic disks defined as $\mathcal{M}^{\text{SY}}(a, \mathbf{b})$, see Section 6.1, but with boundary condition given by L instead of $\mathbb{R} \times \Lambda$. (Note that since L is not \mathbb{R} -invariant, \mathbb{R} does generally not act on $\mathcal{M}^{\text{SY}, L}(a; \mathbf{b})$.) We define a chain map

$$\Phi: C(R_+) \rightarrow C(R_-)$$

between the algebras at the positive and the negative ends as follows: Φ is the identity map on chains, and on Reeb chords a of $\Lambda_{s_j + \varepsilon}$ Φ is given by

$$\Phi(a) = \sum_{\mathbf{b}} [\mathcal{M}^{\text{SY}, L}(a; \mathbf{b})],$$

where \mathbf{b} runs over all words of Reeb chords of $\Lambda_{s_j + \varepsilon}$ and $[\mathcal{M}^{\text{SY}, L}(a; \mathbf{b})]$ denotes the chain of Reeb strings carried by the moduli space. SFT compactness and gluing as in [10] shows that the chain map equation $\partial_{s_j + \varepsilon} \Phi = \Phi \partial_{s_j - \varepsilon}$ holds. It remains to show that Φ is a quasi-isomorphism.

Consider first the case that s_j is a handle slide. Taking ε sufficiently small we find that for each Reeb chord a on $\Lambda_{s_j - \varepsilon}$ there is a unique holomorphic strip connecting it to the corresponding Reeb chord a on $\Lambda_{s_j + \varepsilon}$. (These strips converge to trivial strips as $\varepsilon \rightarrow 0$.) It follows that for each generator c (chord or chain),

$$\Phi(c) = c + \Phi_0(c),$$

where the filtration degree of $\Phi_0(c)$ is strictly smaller than that of c . Thus Φ induces an isomorphism on the E_2 -page of the action filtration spectral sequence and is hence a quasi-isomorphism.

Consider second the case of an action switch. In this case we find exactly as in the handle slide case that

$$\Phi(c) = c + \Phi_0(c)$$

for each c . The only difference is that now one action window contains two generators. Since the two Reeb chords have the same action but lie at a positive distance apart, it follows by monotonicity and Stokes' theorem that the chain map induces an isomorphism also in this action window. We find as above that Φ is a quasi-isomorphism.

Finally consider the case that s_j is a birth moment where two new Reeb chords a and b are born (the death case is analogous). For $\varepsilon > 0$ sufficiently small we have

$$\partial^{\text{SY}} a = b + \partial_0^{\text{SY}}(a),$$

where the action of $\partial_0^{\text{SY}}(a)$ is strictly smaller than the action of b , see [13, Lem. 2.14]. As above we find that for any Reeb chord c of $\Lambda_{s_j + \varepsilon}$ we have

$$\Phi(c) = c + \Phi_0(c).$$

If we filter by small action windows that contain one Reeb chord each, except for one that contains both a and b (note that the action of a approaches the action of b as $\varepsilon \rightarrow 0$) we find again that Φ gives an isomorphism on the E_2 -page and hence is an isomorphism. We conclude that we can subdivide the interval S into pieces with endpoints with quasi-isomorphic algebras. The theorem follows.

According to Theorem 6.2, $(C(R), \partial)$ is a (noncommutative but strictly associative) differential graded (dg) ring containing the dg subring

$$(C(R^0), \partial) = (C(\Omega_{x_0} \Lambda), \partial^{\text{sing}}).$$

Thus $(C(R), \partial)$ is a $(C(R^0), \partial)$ -NC-algebra in the sense of the following definition.

DEFINITION 6.3. — Let (R, ∂) be a dg ring. An (R, ∂) -NC-algebra is a dg ring (S, ∂_S) together with a dg ring homomorphism $(R, \partial) \rightarrow (S, \partial_S)$.

It follows that the Legendrian contact homology $H^{\text{contact}}(\Lambda)$ is an NC-algebra over the graded ring

$$H(\Omega_{x_0} \Lambda, \partial^{\text{sing}}) = Z\pi_1(\Lambda) = Z[\lambda^{-1}, \mu^{-1}].$$

Here we have used that in our situation $\Lambda = T^2$ is a $K(\pi, 1)$, so all the homology of its based loop space is concentrated in degree zero and agrees with the group ring of its fundamental group $\pi_1(\Lambda) = Z^2$.

Relation to standard Legendrian contact homology. — Recall that $C(R)$ is a double complex with bidegree (chain degree, chord degree), horizontal differential ∂^{sing} , and vertical differential ∂^{sy} . As observed above, the first page of the spectral sequence corresponding to the chord degree is concentrated in the 0-th column and given by

$$(A := H_0(R, \partial^{\text{sing}}, \partial^{\text{sy}}).$$

Generators of A are words $\alpha_1 a_1 \alpha_2 a_2 \dots \alpha_\ell a_\ell \alpha_{\ell+1}$ consisting of Reeb chords a_i and *homotopy classes* of paths α_i satisfying the same boundary conditions as before. Note that A is an NC-algebra over the subring $A^0 = H_0(R^0) = \mathbb{Z}\pi_1(\Lambda)$ (on which ∂ vanishes), and $A^k = H_0(R^k)$ is the k -fold tensor product of the bimodule A^1 over the ring A^0 .

We denote by

$$\overline{A} := A / I$$

the quotient of A by the ideal I generated by the commutators $[a, \beta]$ of Reeb chords a and $\beta \geq \pi_1(\Lambda)$. Since $\partial(I) \subset I$, the differential descends to a differential $\overline{\partial}^{\text{sy}} : \overline{A} \rightarrow \overline{A}$ whose homology

$$\overline{H}^{\text{contact}}(\Lambda) := \ker \overline{\partial}^{\text{sy}} / \text{im } \overline{\partial}^{\text{sy}}$$

is the usual Legendrian contact homology as defined in [13].

Length filtration. — The complex $(C(R), \partial)$ is filtered by the *length*

$$L(\alpha_1 a_1 \alpha_2 a_2 \dots \alpha_\ell a_\ell \alpha_{\ell+1}) := \sum_{i=1}^{\ell} L(a_i),$$

where $L(a) = \int_a \lambda$ denotes the action of a Reeb chord a , which agrees with its period and also with the length of the corresponding binormal cord. The length is preserved by the singular boundary operator ∂^{sing} and strictly decreases under ∂^{sy} .

REMARK 6.4. — The construction of Legendrian contact homology in this subsection works for any pair (M, Λ) such that M has no contractible closed Reeb orbits. Examples include cosphere bundles S^*Q of n -manifolds Q with a metric of nonpositive curvature that are convex at infinity, with $\Lambda = \Lambda_K$ the unit conormal bundle of a closed connected submanifold $K \subset Q$. However, if Λ is not a $K(\pi, 1)$, then the coefficient ring $H(\Omega_{x_0}\Lambda, \partial^{\text{sing}})$ will not be equal to the group ring of its fundamental group but contain homology in higher degrees.

6.3. SWITCHING BOUNDARY CONDITIONS, WINDING NUMBERS, AND LENGTH. — We continue to consider $Q = \mathbb{R}^3$ equipped with the flat metric and an oriented framed knot $K \subset Q$. In addition, we assume from now on that K is *real analytic*; this can always be achieved by a small perturbation of K not changing its knot type. We equip T^*Q with an almost complex structure J which agrees with an \mathbb{R} -invariant almost complex structure on the symplectization of S^*Q outside a finite radius disk sub-bundle of T^*Q and with the standard almost complex structure J_{st} on T^*Q inside the disk sub-bundle of half that radius. An explicit formula for such J is given in Section 8.2. We point

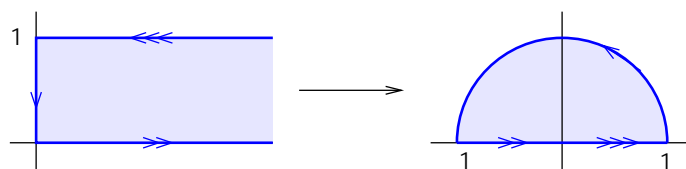


FIGURE 6.2. The biholomorphism χ .

out that the canonical isomorphism $(T \setminus Q, J_{\text{st}}) = (\mathbb{C}^3, i)$ identifies the fibre with \mathbb{R}^3 and the zero section with $i\mathbb{R}^3$. Recall that $L = Q \sqcup L_K$.

Let D be the closed unit disk with a boundary puncture at $1 \in \partial D$ and let $u: (D, \partial D) \rightarrow (T \setminus Q, L)$ be a holomorphic disk with one positive puncture and switching boundary conditions. This means that the map u is asymptotic to a Reeb chord at infinity at the positive puncture 1 and that it is smooth outside an additional finite number of boundary punctures where the boundary switches, i.e., jumps from one irreducible component of L to another (which may be the same one). At these additional boundary punctures, the holomorphic disk is asymptotic to some point in the clean intersection $K \cap L$, i.e., it looks like a corner of a disk in Lagrangian intersection Floer homology.

The real analyticity of χ allows us to get explicit local forms for holomorphic disks near corners. We show in Lemma 8.6 that there are holomorphic coordinates

$$\mathbb{R} \times (0, 0) \xrightarrow{U} \mathbb{C} \times \mathbb{C}^2,$$

in which K corresponds to $\mathbb{R} \times (0, 0)$, the 0 -section Q corresponds to $\mathbb{R} \times \mathbb{R}^2$, and the conormal L_K to $\mathbb{R} \times i\mathbb{R}^2$.

Consider now a neighborhood of a switching point of a holomorphic disk u on the boundary of D , where we use z in a half-disk D_ε^+ around 0 in the upper half-plane as a local coordinate around the switching point in the source. According to Section 4.1, u admits a Taylor expansion around 0 , with $u = (u_1, u_2) \in \mathbb{C} \times \mathbb{C}^2$:

$$(6.2) \quad u_1(z) = \sum_{k \in 2\mathbb{N}} b_k z^k, \quad u_2(z) = \sum_{k \in \frac{1}{2}\mathbb{N}} c_k z^k.$$

Here compared to Section 4.1 we have divided the indices by 2 , so the b_k and c_k correspond to the a_{2k} in Section 4.1. The coefficients b_j are real constants, reflecting smoothness of the tangent component of u . The c_k satisfy one of the conditions in Remark 4.1, i.e., they are either all real or all purely imaginary vectors in \mathbb{C}^2 , and the indices are either all integers or all half-integers.

Equivalently (and more adapted to the analytical study in Sections 8 – 10) one can use z in a neighborhood of infinity in the strip $\mathbb{R} \times [0, 1]$ as a local coordinate in the source. Composing the Taylor expansions (6.2) with the biholomorphism

$$(6.3) \quad \chi: \mathbb{R}_{>0} \times [0, 1] \xrightarrow{\cong} D^+, \quad z \mapsto \exp(-\pi z)$$

(see Figure 6.2) one gets instead the Fourier expansions

$$(6.4) \quad u_1(z) = \sum_{k \in \mathbb{N}} (-1)^k b_k e^{-k\pi z}, \quad u_2(z) = \sum_{k \in \frac{1}{2}\mathbb{N}} (-1)^k c_k e^{-k\pi z}.$$

Recall from Section 4.2 that the *local winding number* at the switch is the positive half-integer or integer which is the index of the first non-vanishing Fourier coefficients in the expansion of u_2 in (6.4). The sum of the local winding numbers at all switching boundary punctures is the *total winding number* of the disk. Since the number of switches from L_K to Q equals that from Q to L_K , the total winding number is an integer.

The following technical result, which is a special case of [5, Th. 1.2], will play a crucial role in the sequel.

THEOREM 6.5 ([5]). — *For a cord-generic real analytic knot $K \subset \mathbb{R}^3$ the total winding number, and in particular the number of switches, of any holomorphic disk $u: (D, \partial D) \rightarrow (T^*Q, L)$ with one positive puncture is uniformly bounded by a constant κ .*

REMARK 6.6. — The necessary energy bound appearing in the corresponding statement in [5] is automatic here, since in our present situation the energy is given by the action of the Reeb chord at the positive puncture, which only varies in a finite set.

In view of this result, when we discuss compactness we need only consider sequences of holomorphic disks with a *fixed* finite number of switches, each of fixed winding number. As we prove in Section 8, each moduli space of such holomorphic disks is for generic data a manifold that admits a natural compactification as a manifold with boundary with corners. We will specifically need such moduli spaces of dimension 0, 1, or 2 and we give brief descriptions in these cases.

Let a be a Reeb chord of Λ_K . Let q_1, \dots, q_m be punctures in ∂D and let $\mathbf{n} = (n_1, \dots, n_m)$ be a vector of local winding numbers, so $n_j \in \{\frac{1}{2}, 1, \frac{3}{2}, 2, \dots\}$ is the local winding number at q_j . We write $\mathcal{M}(a; \mathbf{n})$ for the moduli space of holomorphic disks with positive puncture at the Reeb chord a and switching punctures at q_1, \dots, q_m with winding numbers according to \mathbf{n} . Define the nonnegative integer

$$j\mathbf{n}j := \sum_{j=1}^m 2(n_j - \frac{1}{2}) > 0.$$

THEOREM 6.7. — *For generic almost complex structure J , the moduli space $\mathcal{M}(a; \mathbf{n})$ is a manifold of dimension*

$$\dim \mathcal{M}(a; \mathbf{n}) = ja - j\mathbf{n}j.$$

Furthermore, the choice of a spin structure on L_K together with the spin structure on \mathbb{R}^3 induces a natural orientation on $\mathcal{M}(a; \mathbf{n})$.

Proof. — This is a consequence of [5, Th. A.1] and Lemma 9.5 below.

Note that, due to Theorem 6.5, any moduli space $\mathcal{M}(a; \mathbf{n})$ is empty if \mathbf{n} has more than κ components, i.e., there are more than κ switches.

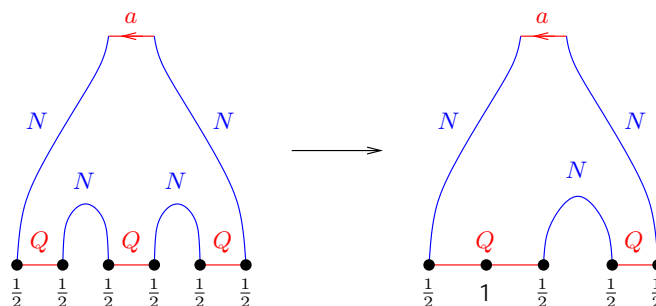


FIGURE 6.3. Type (Lag) boundary where an N -string disappears.

6.4. MODULI SPACES OF DIMENSION ZERO AND ONE. — For moduli spaces of dimension ≤ 1 with positive puncture at a Reeb chord of degree ≤ 1 , we have the following. Theorem 6.7 implies that if $|j_a| = 0$ then $M(a; \mathbf{n})$ is empty if $|j_{\mathbf{n}}| > 0$ and is otherwise a compact oriented 0-manifold. Likewise, if $|j_a| = 1$ then $M(a; \mathbf{n})$ is empty if $|j_{\mathbf{n}}| > 1$ and is an oriented 0-manifold if $|j_{\mathbf{n}}| = 1$. Note that $|j_{\mathbf{n}}| = 1$ implies that there is exactly one switch with winding number 1 and that the winding numbers at all other switches equal $\frac{1}{2}$. Finally, if all entries in \mathbf{n} equal $\frac{1}{2}$ then $\dim(M(a; \mathbf{n})) = 1$.

It follows by Theorem 10.3 that the 1-dimensional moduli spaces of disks with switching boundary condition admit natural compactifications to 1-manifolds with boundary. The next result describes the disk configurations corresponding to the boundary of these compact intervals.

PROPOSITION 6.8. — *If a is a Reeb chord of degree $|j_a| = 1$ and if all entries of \mathbf{n} equal $\frac{1}{2}$, then the oriented boundary of $M(a; \mathbf{n})$ consists of the following:*

(Lag) *Moduli spaces $M(a; \mathbf{n}^0)$, where \mathbf{n}^0 is obtained from \mathbf{n} by removing two consecutive $\frac{1}{2}$ -entries and inserting in their place a 1.*

(sy) *Products of moduli spaces*

$$M^{\text{sy}}(a; \mathbf{b})/\mathbb{R} \quad \coprod_{b_j \geq 2b} M(b_j; \mathbf{n}_j),$$

where \mathbf{n} equals the concatenation of the \mathbf{n}_j .

Proof. — This is a consequence of Theorem 10.3. To motivate the result, note that the first type of boundary corresponds to two switches colliding, see Figures 6.3 and 6.4. The second type corresponds to a splitting into a two level curve with one R-invariant level (of dimension 1) in the symplectization and one rigid curve (of dimension 0) in T^*Q , see Figure 6.5. By transversality, compactness, and the dimension formula this accounts for all the possible boundary phenomena, and by a gluing argument we find that any such configuration corresponds to a unique boundary point.

We conclude this subsection by giving an alternate interpretation of the first boundary phenomenon in Proposition 6.8. Let $\mathcal{M}(a; \mathbf{n})$ denote the moduli space corresponding to $M(a; \mathbf{n})$, but with one extra marked point on the boundary of the disk.

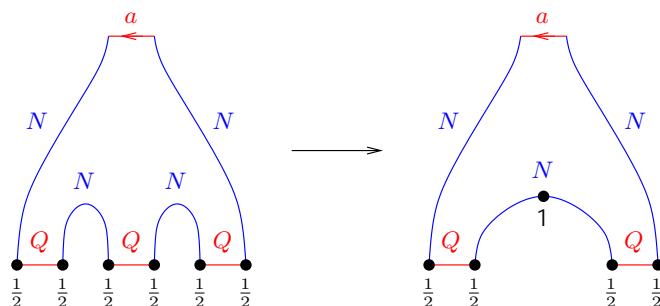


FIGURE 6.4. Type (Lag) boundary where a Q -string disappears.

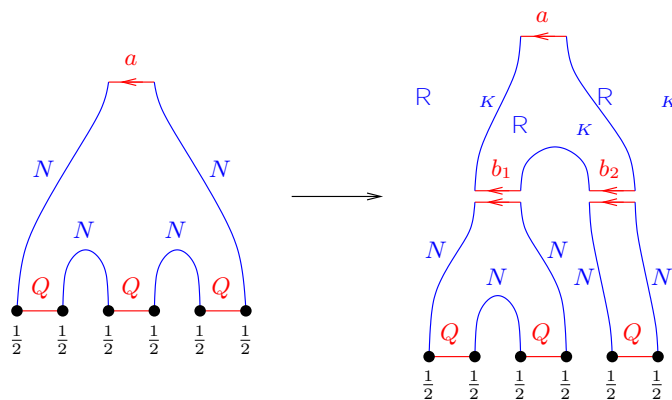


FIGURE 6.5. Type (sy) boundary.

Then $M(a; \mathbf{n})$ fibers over $M(a; \mathbf{n})$ with fiber $\partial D = \{1, q_1, \dots, q_m\}$ and there is an evaluation map $ev: M(a; \mathbf{n}) \rightarrow L$. It follows from Theorem 10.8 that for $\sum j a_j = 1$ and $\sum j n_j = 0$ (and generic data), $ev^{-1}(K)$ is a transversely cut out oriented 0-manifold that projects injectively into $M(a; \mathbf{n})$. We denote its image by

$$\delta M(a; \mathbf{n}).$$

As the notation suggests, this space will be the natural domain for the string operations $\delta = \delta_Q + \delta_N$.

PROPOSITION 6.9. — *If a is a Reeb chord of degree $\sum j a_j = 1$ and if all entries \mathbf{n} equal $\frac{1}{2}$, then there is a natural orientation preserving identification between $\delta M(a; \mathbf{n})$ and $M(a; \mathbf{n}^{00})$, where \mathbf{n}^{00} is obtained from \mathbf{n} by inserting in \mathbf{n} a new entry equal to 1 at the position given by the marked point.*

Proof. — This is a consequence of Theorem 10.8. Here is the idea. Consider local coordinates around the marked point in the source and around K in the target. Then the Taylor expansions (6.2) with $c_{\frac{1}{2}} = 0$ and $c_1 \neq 0$ give the map in $\delta M(a; \mathbf{n})$ with the marked point corresponding to 0. The corresponding Fourier expansions (6.4) present

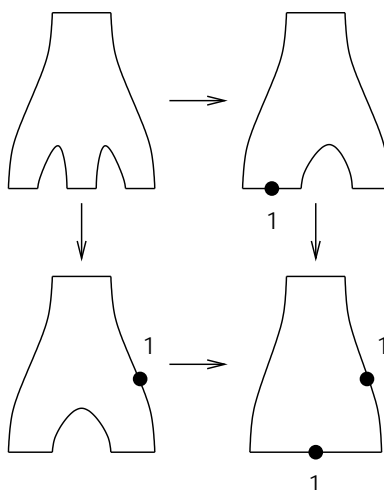


FIGURE 6.6. Type $(\text{Lag}/\text{Lag})^1$ corner.

the map as an element in $M(a; \mathbf{n}^0)$, where the marked point is replaced by a puncture. Conversely, translating the Fourier picture to the Taylor picture proves the other inclusion and hence equality holds. See Section 9.6 for a discussion of orientations of the moduli spaces involved.

6.5. MODULI SPACES OF DIMENSION TWO. — For moduli spaces $M(a; \mathbf{n})$ with positive puncture at a Reeb chord a of degree $jaj = 2$, Theorem 6.7 implies the following:

- If $j\mathbf{n}j > 2$ then $M(a; \mathbf{n}) = ?$.
- If $j\mathbf{n}j = 2$ then $M(a; \mathbf{n})$ is a compact 0-dimensional manifold. This can happen in two ways: either exactly one entry in \mathbf{n} equals $\frac{3}{2}$, or exactly two entries equal 1 and all others equal $\frac{1}{2}$.
- If $j\mathbf{n}j = 1$ then $M(a; \mathbf{n})$ is an oriented 1-manifold, exactly one entry in \mathbf{n} equals 1 and all others equal $\frac{1}{2}$.
- If $j\mathbf{n}j = 0$ then $M(a; \mathbf{n})$ is an oriented 2-manifold and all entries in \mathbf{n} equal $\frac{1}{2}$.

It follows by Theorem 10.6 that the 2-dimensional moduli spaces of disks with switching boundary condition admit natural compactifications to 2-manifolds with boundary and corners. The next result describes the disk configurations corresponding to the boundary and corner points of these compact surfaces, see Figures 6.6, 6.7, 6.8 and 6.9.

PROPOSITION 6.10. — *If a is a Reeb chord of degree $jaj = 2$ and if all entries of \mathbf{n} equal $\frac{1}{2}$, then the 1-dimensional boundary segments in the boundary of $M(a; \mathbf{n})$ consist of the following configurations:*

(Lag) Moduli spaces $M(a; \mathbf{n}^0)$, where \mathbf{n}^0 is obtained from \mathbf{n} by removing two consecutive $\frac{1}{2}$ -entries and inserting in their place a 1.

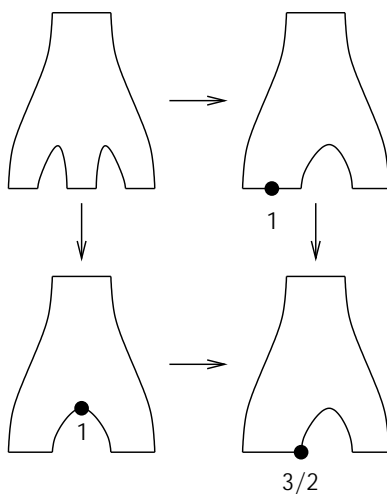


FIGURE 6.7. Type $(\text{Lag}/\text{Lag})^2$ corner.

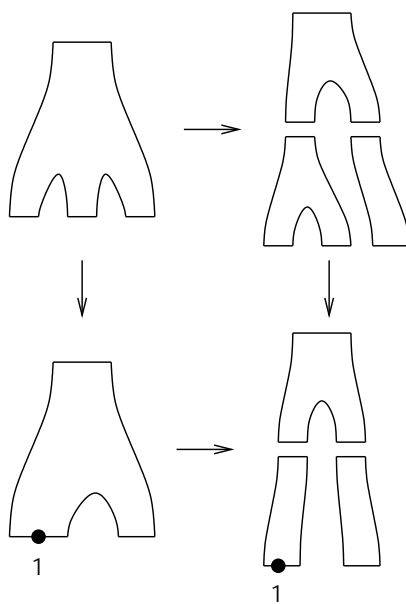


FIGURE 6.8. Type (sy/Lag) corner.

(sy) Products of moduli spaces

$$M^{\text{sy}}(\mathbf{a}; \mathbf{b})/\mathbb{R} \quad \prod_{b_j \geq 2} M(b_j; \mathbf{n}_j),$$

where \mathbf{n} equals the concatenation of the \mathbf{n}_j .

The corner points in the boundary consists of the following configurations:

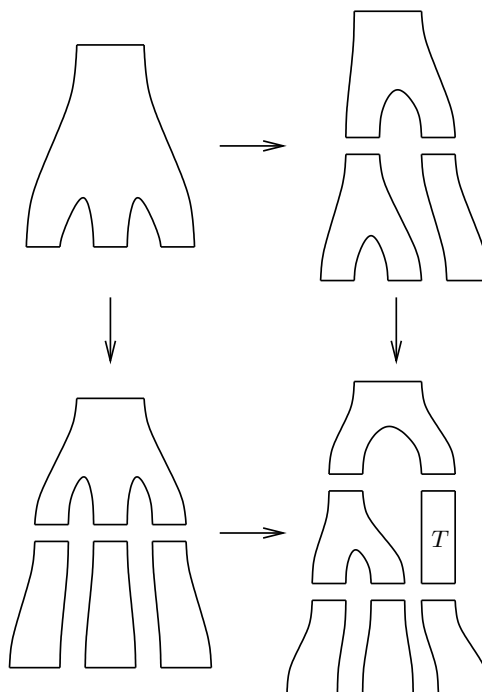


FIGURE 6.9. Type (sy/sy) corner.

(Lag/Lag)¹ Moduli spaces $M(a; \mathbf{n}^0)$, where \mathbf{n}^0 is obtained from \mathbf{n} by removing two pairs of consecutive $\frac{1}{2}$ -entries and inserting 1's in their places.

(Lag/Lag)² Moduli spaces $M(a; \mathbf{n}^{00})$, where \mathbf{n}^{00} is obtained from \mathbf{n} by removing three consecutive $\frac{1}{2}$ -entries and inserting a $\frac{3}{2}$ in their place.

(sy/Lag) Products of moduli spaces

$$M^{\text{sy}}(a; \mathbf{b})/\mathbb{R} \quad \prod_{b_j \geq 2\mathbf{b}} M(b_j; \mathbf{n}_j),$$

where the concatenation of the \mathbf{n}_j gives \mathbf{n} with one consecutive pair of $\frac{1}{2}$ -entries removed and a 1 inserted in their place.

(sy/sy) Products of moduli spaces

$$M^{\text{sy}}(a; \mathbf{b})/\mathbb{R} \quad \prod_{b_j \geq 2\mathbf{b}} \left(M^{\text{sy}}(b_j; \mathbf{c}_j)/\mathbb{R} \quad \prod_{c_{jk} \geq 2\mathbf{c}_j} M(c_{jk}; \mathbf{n}_{jk}) \right),$$

where \mathbf{n} equals the concatenation of the \mathbf{n}_{jk} , and all but one of the $M^{\text{sy}}(b_j; \mathbf{c}_j)$ are trivial strips over the Reeb chords b_j .

Proof. — This is a consequence of Theorem 10.6. The descriptions of the boundary segments are analogous to the boundary phenomena of Proposition 6.8. At a type (Lag/Lag)¹ corner we have two pairs of switches colliding. Local coordinates in the moduli space around this configuration can be taken as the lengths of the corresponding short boundary segments, which is a product of two half-open intervals. At a type

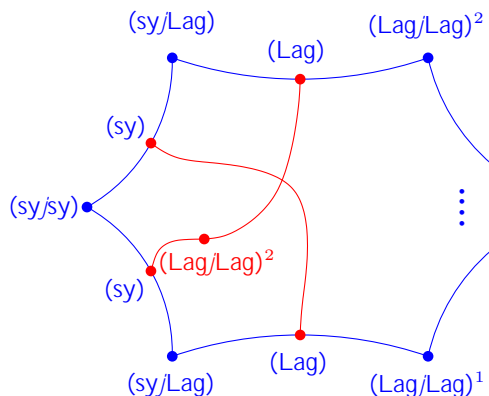


FIGURE 6.10. The 2-dimensional moduli space $M(a; \mathbf{n})$ and the immersed curve $\pi(\delta M(a; \mathbf{n}))$.

$(\text{Lag}/\text{Lag})^2$ corner there are likewise two short boundary segments that give local coordinates on the moduli space, see Figure 6.7. At a type (sy/Lag) corner the two parameters are the length of the short boundary segment and the gluing parameter for the two-level curve. Finally, at a type (sy/sy) corner the two parameters are the two gluing parameter for the three-level curve.

We next give alternate interpretations of the boundary phenomena in Proposition 6.10. Recall the notation $M(a; \mathbf{n})$ for the moduli space corresponding to $M(a; \mathbf{n})$ in which the disks have an additional free marked point on the boundary. It comes with an evaluation map $\text{ev}: M(a; \mathbf{n}) \rightarrow L$ and a projection $\pi: M(a; \mathbf{n}) \rightarrow M(a; \mathbf{n})$ forgetting the marked point, and we denote $\delta M(a; \mathbf{n}) = \text{ev}^{-1}(K)$.

PROPOSITION 6.11. — *If a is a Reeb chord of degree $|aj| = 2$ and if all entries \mathbf{n} equal $\frac{1}{2}$, then there is a natural orientation preserving identification between $\delta M(a; \mathbf{n})$ and $M(a; \mathbf{n}^0)$, where \mathbf{n}^0 is obtained from \mathbf{n} by inserting in \mathbf{n} a new entry equal to 1 at the position given by the marked point.*

The moduli space $\delta M(a; \mathbf{n}) \subset M(a; \mathbf{n})$ is an embedded curve with boundary. Its boundary consists of transverse intersections with the boundary of $M(a; \mathbf{n})$, corresponding to degenerations of type (sy/Lag) and $(\text{Lag}/\text{Lag})^1$ involving the marked point, and to points in the interior of $M(a; \mathbf{n})$, corresponding to degenerations of type $(\text{Lag}/\text{Lag})^2$ involving the marked point.

The projection $\pi(\delta M(a; \mathbf{n})) \subset M(a; \mathbf{n})$ is an immersed curve with boundary and transverse self-intersections. Its boundary consists of transverse intersections with the boundary of $M(a; \mathbf{n})$. See Figure 6.10.

Proof. — This is a consequence of Theorem 10.9. Here is a sketch. The proof of the first statement is analogous to that of Proposition 6.9, looking at Taylor and

Fourier expansions. That $\delta\mathcal{M}(a; \mathbf{n})$ is an embedded curve with boundary follows from transversality of the evaluation map $\text{ev}: \mathcal{M}(a; \mathbf{n}) \rightarrow L$ to the knot K , which holds for generic almost complex structure. More refined transversality arguments show that the projection $\pi(\delta\mathcal{M}(a; \mathbf{n}))$ is an immersed curve with transverse self-intersections corresponding to holomorphic disks that meet the knot twice at non-corner points on their boundary.

For the other statements, note that each stratum of $\delta\mathcal{M}(a; \mathbf{n})$ corresponds to a moduli space $\mathcal{M}(a; \mathbf{n}^\theta)$, where \mathbf{n}^θ is obtained from \mathbf{n} by inserting an entry 1 corresponding to the marked point \bullet . It follows from Proposition 6.10 that boundary points of $\delta\mathcal{M}(a; \mathbf{n})$ correspond to degenerations of types (sy/Lag) , $(\text{Lag}/\text{Lag})^1$ and $(\text{Lag}/\text{Lag})^2$ involving the point \bullet . The first two correspond to transverse intersections of $\pi(\delta\mathcal{M}(a; \mathbf{n}))$ with boundary strata of $\mathcal{M}(a; \mathbf{n})$ of types (sy) and (Lag) , respectively. A dimension argument shows that degenerations of type $(\text{Lag}/\text{Lag})^2$ involving the point \bullet cannot meet the boundary of $\mathcal{M}(a; \mathbf{n})$, so they correspond to boundary points of $\delta\mathcal{M}(a; \mathbf{n})$ in the interior of $\mathcal{M}(a; \mathbf{n})$. They appear in pairs corresponding to holomorphic disks in which the marked point \bullet has approached a corner from the left or right to form a new corner of weight $3/2$. In $\delta\mathcal{M}(a; \mathbf{n})$ the two configurations on a pair are distinct (formally, they are distinguished by the position of the marked point \bullet on the 3-punctured constant disk attached at the weight $3/2$ corner), so they give actual boundary points. In the projection $\pi(\delta\mathcal{M}(a; \mathbf{n}))$ the two configurations become equal and thus give an interior point, hence $\pi(\delta\mathcal{M}(a; \mathbf{n}))$ has no boundary points in the interior of $\mathcal{M}(a; \mathbf{n})$.

See Section 9.6 for a discussion of orientations of the moduli spaces involved in these arguments.

6.6. THE CHAIN MAP. — We can summarize the description of the moduli spaces of punctured holomorphic disks with switching boundary conditions in the preceding subsections as follows. For all Reeb chords a and all integers $\ell > 0$ the compactified moduli spaces

$$\overline{\mathcal{M}}_\ell(a) := \overline{\mathcal{M}}(a; \underbrace{\frac{1}{2}, \dots, \frac{1}{2}}_{2\ell})$$

are compact oriented manifolds with boundary and corners of dimension $|ja|$ whose codimension 1 boundaries satisfy the relations

$$(6.5) \quad \partial \overline{\mathcal{M}}_\ell(a) = \overline{\mathcal{M}}_\ell(\partial a) \sqcup \delta \overline{\mathcal{M}}_{\ell-1}(a),$$

where $\partial a = \partial^{\text{sy}}a$ and $\delta \overline{\mathcal{M}}_{\ell-1}(a)$ is the closure in $\overline{\mathcal{M}}_{\ell-1}(a)$ of the set

$$\delta \mathcal{M}_{\ell-1}(a) := \delta \mathcal{M}(a; \underbrace{\frac{1}{2}, \dots, \frac{1}{2}}_{2\ell-2})$$

introduced in Proposition 6.11. Again we refer to Section 9.6 for a description of the orientations involved.

PROPOSITION 6.12. — *There exist smooth triangulations of the spaces $\overline{M}_\ell(a)$ and generic chains of broken strings*

$$\Phi_\ell(a) : \overline{M}_\ell(a) \rightarrow \Sigma^\ell$$

(understood as singular chains by summing up their restrictions to the simplices of the triangulations) satisfying the relations

$$(6.6) \quad \partial\Phi_\ell(a) = \Phi_\ell(\partial a) - (\delta_Q + \delta_N)\Phi_{\ell-1}(a).$$

Proof. — The idea of the proof is very simple: After connecting the end points of a to the base point x_0 by capping paths, a suitable parametrization (explained below) of the boundary of a holomorphic disk $u \in M_\ell(a)$ determines a broken string $\partial(u) \in \Sigma^\ell$. Thus we get maps

$$\tilde{\Phi}_\ell(a) : M_\ell(a) \rightarrow \Sigma^\ell, \quad u \mapsto \partial(u)$$

and the relations (6.6) should follow from (6.5). However, the map $\tilde{\Phi}_\ell(a)$ in general does not extend to the compactification \overline{M}_ℓ as a map to Σ^ℓ because on the boundary some Q - or N -string can disappear in the limit. We will remedy this by suitably modifying the maps $\tilde{\Phi}_\ell(a)$ near the boundaries (inserting spikes).

Before doing this, let us discuss parametrizations of the broken string $\partial(u)$ for $u \in M_\ell(a)$. Near a switch we can pick holomorphic coordinates on the domain (with values in the upper half-disk) and the target (provided by Lemma 8.6) in which the normal projection of u consists of two holomorphic functions near a corner as in Section 4. The discussion in that section shows that in these coordinates $\partial(u)$ satisfies the matching conditions on the m -jets required in the definition of a broken string. We take near each corner a parametrization of $\partial(u)$ induced by such holomorphic coordinates and extend them arbitrarily away from the corners to make $\partial(u)$ a broken string in the sense of Definition 5.1. Note that the space of such parametrizations is contractible.

Now we proceed by induction over $|a| = 0, 1, 2$.

Case $|a| = 0$. — In this case $\overline{M}_\ell(a)$ consists of finitely many oriented points and we set $\Phi_\ell(a)(u) := \partial(u)$ (picking a parametrization of the boundary as above).

Case $|a| = 1$. — We proceed by induction on $\ell = 0, 1, \dots$. For $\ell = 0$, on the boundary $\partial\overline{M}_0(a) = \overline{M}_0(\partial a)$ we are already given the map $\Phi_0(\partial a)$. We extend it to a map $\Phi_0(a) : \overline{M}_0(a) \rightarrow \Sigma^0$ by sending u to $\partial(u)$ with parametrizations matching the given ones on $\partial\overline{M}_0(a)$, so that $\partial\Phi_0(a) = \Phi_0(\partial a)$ holds.

Now suppose that we have already defined $\Phi_0(a), \dots, \Phi_{\ell-1}(a)$ such that the relations (6.6) hold up to $\ell - 1$. According to (6.5), the boundary $\partial\overline{M}_\ell(a)$ is identified with the union of domains of the maps on the right hand side of (6.6). On the other hand, on the interior $M_\ell(a)$ we are given the map $\tilde{\Phi}_\ell(a)$ described above. Furthermore, by Proposition 6.8 and Remark 8.13, elements u close to the boundary points in $\delta\overline{M}_{\ell-1}(a) \subset \partial\overline{M}_\ell(a)$ have spikes (shrinking as u tends to the boundary) roughly in the same direction as those on the boundary. So near $\partial\overline{M}_\ell(a)$ we can interpolate between the map on the boundary given by the right hand side of (6.6) and the

map $\tilde{\Phi}_\ell(a)$ on the interior to obtain a map $\Phi_\ell(a) : \overline{M}_\ell(a) \rightarrow \Sigma^\ell$ satisfying (6.6). Since the modification of $\tilde{\Phi}_\ell(a)$ can be done away from the finite set $\delta\overline{M}_\ell(a) \cap M_\ell(a)$, $\Phi_\ell(a)$ is a generic 1-chain of broken strings. This concludes the inductive step. Since we are dealing with 1-chains, a smooth triangulation just amounts to a parametrization of the components of $\overline{M}_\ell(a)$ by intervals whose boundary points avoid the set $\delta\overline{M}_\ell(a)$.

Case $|a| = 2$. — We proceed again by induction on $\ell = 0, 1, \dots$. For $\ell = 0$, we again define $\Phi_0(a) : \overline{M}_0(a) \rightarrow \Sigma^0$ by sending u to $\partial(u)$, with parametrizations matching the given ones on $\partial\overline{M}_0(a)$, so that $\partial\Phi_0(a) = \Phi_0(\partial a)$ holds.

Now suppose that we have already defined $\Phi_0(a), \dots, \Phi_{\ell-1}(a)$ and triangulations of their domains such that they are generic 2-chains of broken strings and the relations (6.6) hold up to $\ell - 1$. As in the case of 1-chains, the boundary $\partial\overline{M}_\ell(a)$ is identified via (6.5) with the union of domains of the right hand side of (6.6), so as before we define $\Phi_\ell(a)$ on that boundary via the maps $\Phi_\ell(\partial a)$ resp. $\delta\Phi_{\ell-1}(a)$. By induction hypothesis, these maps coincide at corner points. Note that the map $\delta\Phi_{\ell-1}(a)$ inserts spikes at the intersection points with the knot. According to Proposition 6.10 and Remark 8.13, elements u close to the codimension one boundary strata $\delta\overline{M}_{\ell-1}(a)$ have spikes roughly in the same direction as those on the boundary (shrinking in size as u tends to the boundary). Elements close to a corner point where two boundary strata of $\delta\overline{M}_{\ell-1}(a)$ meet have two spikes roughly in the same directions as those on the nearby boundary strata (which both shrink as u tends to the corner point), see Remark 8.13. So we can interpolate between the given map on the boundary $\partial\overline{M}_\ell(a)$ and the map $\tilde{\Phi}_\ell(a)$ on the interior $M_\ell(a)$ to obtain a map $\Phi_\ell(a) : \overline{M}_\ell(a) \rightarrow \Sigma^\ell$ satisfying (6.6).

Recall that $\delta\overline{M}_\ell(a)$ is an immersed 1-dimensional submanifold with finitely many transverse self-intersections in the interior, and which meets the boundary transversely away from the corners. The modification of $\tilde{\Phi}_\ell(a)$ can be done away from the finite set of self-intersections of $\delta\overline{M}_\ell(a)$ in the interior. Moreover, the modification of $\tilde{\Phi}_\ell(a)$ near the boundary only involves inserting spikes at switching points of broken strings, which can be performed away from the finitely many interior intersection points of the broken strings with the knot and thus does not affect $\delta\overline{M}_\ell(a)$.

We pick a smooth triangulation of $\overline{M}_\ell(a)$ transverse to $\delta\overline{M}_\ell(a)$ (i.e., transverse to its 1-dimensional strata as well as its self-intersection points) and inducing the given triangulation on the boundary. By the discussion in the preceding paragraph, $\Phi_\ell(a)$ (interpreted as the sum over its restriction to simplices) is a generic 2-chain of broken strings. This concludes the inductive step and thus the proof of Proposition 6.12.

Given a Reeb chord a , we define

$$\Phi(a) := \sum_{\ell=0}^{\kappa} \Phi_\ell(a) \in C(\Sigma) = \bigoplus_{\ell=0}^{\infty} C(\Sigma^\ell).$$

Here κ is the constant from the Finiteness Theorem 6.5. The relation (6.6) for the chains $\Phi_\ell(a)$ translates into

$$(6.7) \quad \partial\Phi(a) = \Phi(\partial^{\text{sy}} a) - \delta\Phi(a), \quad \delta = \delta_Q + \delta_N.$$

Given a d -simplex of Reeb strings $\mathbf{a} = \alpha_1 a_1 \dots a_m \alpha_{m+1} : \Delta \rightarrow \mathbb{R}^m$ we define

$$\Phi(\mathbf{a}) := \alpha_1 \Phi(a_1) \dots \alpha_m \Phi(a_m) \alpha_{m+1} \in C(\Sigma).$$

Here the boundary arcs are concatenated in the obvious way to obtain broken strings. For singular simplices Δ_i appearing as domains in $\Phi(a_i)$, the corresponding term in $\Phi(\mathbf{a})$ has by our orientation convention the domain

$$\Delta \quad \Delta_1 \quad \dots \quad \Delta_m$$

in this order of factors.

THEOREM 6.13. — *The map Φ is a chain map from $(C(\mathbb{R}), \partial)$ to $(C(\Sigma), \partial + \delta_Q + \delta_N)$.*

Proof. — Using (6.7) we compute for $\mathbf{a} \in C_d(\mathbb{R})$ as above, with $d = d + j\alpha_1 + \dots + j\alpha_{i-1}$:

$$\begin{aligned} \partial\Phi(\mathbf{a}) &= \Phi(\partial^{\text{sing}}\mathbf{a}) + \sum_{i=1}^m (-1)^{\alpha_1 + \dots + \alpha_{i-1}} \alpha_1 \Phi(a_1) \alpha_2 \dots \partial\Phi(a_i) \dots \alpha_m \Phi(a_m) \alpha_{m+1} \\ &= \Phi(\partial^{\text{sing}}\mathbf{a}) + \sum_{i=1}^m (-1)^{\alpha_1 + \dots + \alpha_{i-1}} \alpha_1 \Phi(a_1) \alpha_2 \dots \left(\Phi(\partial^{\text{sy}}a_i) + \delta\Phi(a_i) \right) \dots \alpha_{m+1} \\ &= \Phi(\partial^{\text{sing}}\mathbf{a}) + \Phi(\partial^{\text{sy}}\mathbf{a}) + \delta\Phi(\mathbf{a}). \end{aligned}$$

Since $\partial = \partial^{\text{sing}} + \partial^{\text{sy}}$, this proves the theorem.

Compatibility with length filtrations. — Holomorphic disks with switching boundary conditions have a length decreasing property that leads to the chain map Φ respecting the length (or action) filtration, which is central for our isomorphism proof. Let $u \in M(a; \mathbf{n})$ be a holomorphic disk with k boundary segments that map to Q . Let $\sigma_1, \dots, \sigma_k$ be the corresponding curves in Q and let $L(\sigma_i)$ denote the length of σ_i . Recall that the Reeb chord a is the lift of a binormal chord on the link K and that the action $\int_a p dq$ of a equals the length of the underlying chord in Q , which we write as $L(a)$. In Section 8.2 we utilize the positivity of a scaled version of the contact form on holomorphic disks to show the following result (Proposition 8.9).

PROPOSITION 6.14. — *If $u \in M(a; \mathbf{n})$ is as above then*

$$\sum_{i=1}^k L(\sigma_i) \leq L(a),$$

with equality if and only if u is a trivial half strip over a binormal chord.

Recall that both chain complexes $(C(\mathbb{R}), \partial)$ and $(C(\Sigma), \partial + \delta_Q + \delta_N)$ carry length filtrations that were defined in Sections 6.2 and 5.5, respectively. Recall also that the length filtration on $C(\Sigma)$ does not count the lengths of Q -spikes. Hence the insertion of Q -spikes in the definition of the chain map Φ does not increase length and Proposition 6.14 implies

COROLLARY 6.15. — *The chain map Φ in Theorem 6.13 respects the length filtrations, i.e., it does not increase length.*

7. PROOF OF THE ISOMORPHISM IN DEGREE ZERO

In the previous section we have constructed a chain map $\Phi : (C(R), \partial) \rightarrow (C(\Sigma), D)$, where $D = \partial + \delta_Q + \delta_N$. In this section we finish the proof of Theorem 1.2 by showing that the induced map $\Phi : H_0(R, \partial) \rightarrow H_0(\Sigma, D)$ in degree zero is an isomorphism. Whereas the results in the previous section hold for any 3-manifold Q with a metric of nonpositive curvature which is convex at infinity, in this section we need to restrict to the case $Q = \mathbb{R}^3$ with its Euclidean metric. This restriction will allow us to obtain crucial control over the straightening procedure for Q -strings described in Proposition 7.6 (see the comment in Remark 7.9 below).

As a first step, we will slightly extend the definition of broken strings to include piecewise linear Q -strings. A relatively simple approximation result will show that the inclusion of broken strings with piecewise linear Q -strings into all broken strings induces an isomorphism on string homology in degree 0.

The central piece of the argument will then consist of deforming the complex of broken strings with piecewise linear Q -strings into the subcomplex of those with *linear* Q -strings.

It is important that both of these reduction steps can be done *preserving the length filtration on Q -strings*. The final step of the argument then consists of comparing the contact homology $H_0(R, \partial)$ with the homology of the chain complex of broken strings with linear Q -strings. At this stage, we will use the length filtrations to reduce to the comparison of homology in degrees 0 and 1 in small length windows containing at most one critical value.

7.1. APPROXIMATION BY PIECEWISE LINEAR Q -STRINGS. — In the following we enlarge the space of broken C^m -strings Σ , keeping the same notation, to allow for Q -strings to be *piecewise C^m* . Here a Q -string $s_{2i} : [a_{2i-1}, a_{2i}] \rightarrow Q$ is called *piecewise C^m* if there exists a subdivision $a_{2i-1} = b_0 < b_1 < \dots < b_r = a_{2i}$ such that the restriction of s_{2i} to each subinterval $[b_{j-1}, b_j]$ is C^m . For a generic d -chain $S : \Delta_d \rightarrow \Sigma^\ell$ ($d = 0, 1, 2$) we require that the number of subdivision points on each Q -string is constant over the simplex Δ_d . The subdivision points can vary smoothly over Σ_d but have to remain distinct. If for some subdivision point b_j the two C^m -strings meeting at b_j fit together in a C^m -fashion for all $\lambda \in \Delta_d$, then we identify S with the generic d -chain obtained by removing the subdivision point b_j .

We allow Q -strings in a generic d -chain S to meet the knot K at a subdivision point b_j , provided at such a point the derivatives from both sides satisfy the genericity conditions in Definition 5.3. If this occurs for some parameter value $\lambda \in \Delta_2$ in a generic 2-chain, then we require in addition that the corresponding Q -string meets K at the subdivision point $b_j(\lambda)$ for all λ in the component of λ in the domain M_{δ_Q} of $\delta_Q S$ defined in Section 5.3. These conditions ensure that the operator $D = \partial + \delta_Q + \delta_N$ extends to generic chains of piecewise C^m strings satisfying the relations in Proposition 5.8.

The subspace $\Sigma_{\text{pl}} \subset \Sigma$ of broken strings whose Q -strings are *piecewise linear* give rise to an inclusion of a D -subcomplex

$$(7.1) \quad C(\Sigma_{\text{pl}}) \xhookrightarrow{i_{\text{pl}}} C(\Sigma).$$

For this to hold, we choose the Q -spikes inserted under the map δ_N to be degenerate 3-gons, i.e., short segments orthogonal to the knot traversed back and forth. Then $C(\Sigma_{\text{pl}})$ becomes a D -subcomplex.

We will also consider the subspace $\Sigma_{\text{lin}} \subset \Sigma_{\text{pl}}$ of broken closed strings whose Q -strings are (essentially) *linear*: any two points $x_1, x_2 \in K$ determine a unique line segment $[x_1, x_2]$ in \mathbb{R}^3 connecting them. For technical reasons, special care has to be taken when such a linear Q -string becomes very short. Indeed, near the diagonal $\Delta \subset K \times K$ we deform the segments to piecewise linear strings with one corner in such a way that at each point of the diagonal, instead of a segment of length zero we have a degenerate 3-gon as above, i.e., a short spike in direction of the curvature of the knot (which we assume vanishes nowhere). Now $\Sigma_{\text{lin}} \subset \Sigma_{\text{pl}}$ consists of all broken closed strings whose Q -strings are constant speed parametrizations of such (possibly deformed) segments. In this way,

$$(7.2) \quad C(\Sigma_{\text{lin}}) \xhookrightarrow{i_{\text{lin}}} C(\Sigma_{\text{pl}})$$

will be an inclusion of a D -subcomplex.

Recall from Section 5.5 that these complexes are filtered by the length $L(\beta)$, i.e., the maximum of the total length of Q -strings over all parameter values of the chain, where in the length we do not count Q -spikes. With these notations, we have the following approximation result.

PROPOSITION 7.1. — *There exist maps*

$$F_0 : C_0(\Sigma) \rightarrow C_0(\Sigma_{\text{pl}}), \quad F_1 : C_1(\Sigma) \rightarrow C_1(\Sigma_{\text{pl}})$$

and

$$H_0 : C_0(\Sigma) \rightarrow C_1(\Sigma), \quad H_1 : C_1(\Sigma) \rightarrow C_2(\Sigma)$$

satisfying with the map i_{pl} from (7.1):

- (i) $F_0 i_{\text{pl}} = \mathbb{1}$ and $DH_0 = i_{\text{pl}} F_0 - \mathbb{1}$;
- (ii) $F_1 i_{\text{pl}} = \mathbb{1}$ and $H_0 D + DH_1 = i_{\text{pl}} F_1 - \mathbb{1}$;
- (iii) F_0, H_0, F_1 and H_1 are (not necessarily strictly) length-decreasing.

Proof. — We first define F_0 and H_0 . Given $\beta \in C_0(\Sigma)$, we pick finitely many subdivision points p_i on the Q -strings in β (which include all end points) and define $H_0 \beta$ to be the straight line homotopy from β to the broken string $F_0 \beta$ whose Q -strings are the piecewise linear strings connecting the p_i . We choose the subdivision so fine that the Q -strings in $H_0 \beta$ remain transverse to K at the end points and do not meet K in the interior. The N -strings are just slightly rotated near the end points to match the

new Q -strings, without creating intersections with K . Then $H_0\beta$ is a generic 1-chain in Σ satisfying

$$\partial H_0\beta = F_0\beta - \beta, \quad \delta_Q H_0\beta = \delta_N H_0\beta = 0.$$

If β is already piecewise linear we include the corner points in the subdivision to ensure $F_0\beta = \beta$, so that condition (i) holds.

To define F_1 and H_1 , consider a generic 1-simplex $\beta : [0, 1] \rightarrow \Sigma$. We pick finitely many smooth paths of subdivision points $p_i(\lambda)$ on the Q -strings in $\beta(\lambda)$ (which include all end points) and define $H_1\beta$ to be the straight line homotopy from β to the 1-simplex $F_1\beta$ whose Q -strings are the piecewise linear strings connecting the $p_i(\lambda)$. Here we choose the $p_i(\lambda)$ to agree with the ones in the definition of H_0 at $\lambda = 0, 1$ as well as at the finitely many values λ_j where some Q -string intersects the knot in its interior (so at such λ_j the intersection point with K is included among the $p_i(\lambda_j)$). Note that for this we may first have to add new subdivision points on the Q -strings on $\beta(\lambda)$ for $\lambda = 0, 1, \lambda_j$, which is allowed due to the identification above. Moreover, we choose the subdivision so fine that the Q -strings in $H_1\beta$ remain transverse to K at the end points and meet K in the interior exactly at the values λ_j above. The N -strings are just slightly rotated near the end points to match the new Q -strings, without creating new intersections with K besides the ones already present in β that are continued along the homotopy. Then $H_1\beta$ is a generic 2-chain in Σ satisfying

$$(\partial H_1 + H_0\partial)\beta = F_1\beta - \beta, \quad (\delta_Q H_1 + H_0\delta_Q)\beta = (\delta_N H_1 + H_0\delta_N)\beta = 0.$$

If β is already piecewise linear we include the corner points in the subdivision to ensure $F_1\beta = \beta$, so that condition (ii) holds.

7.2. PROPERTIES OF TRIANGLES FOR GENERIC KNOTS. — In our arguments, we will assume that the knot K is generic. In particular, we will use that it has the properties listed in the following lemma.

LEMMA 7.2. — *A generic knot $K \subset \mathbb{R}^3$ has the following properties:*

- (i) *There exists an $S \geq N$ such that each plane intersects K at most S times.*
- (ii) *The set $T \subset K$ of points whose tangent lines meet the knot again is finite (and each such tangent line meets the knot in exactly one other point).*

Proof. — We prove part (i). For a generic knot K parametrized by

$$\gamma : S^1 \rightarrow \mathbb{R}^3,$$

the first four derivatives $(\dot{\gamma}, \gamma^{(2)}, \gamma^{(3)}, \gamma^{(4)})$ span \mathbb{R}^3 at each $t \in S^1$. (For this, use the jet transversality theorem [25, Chap. 3] to make the corresponding map $S^1 \rightarrow (\mathbb{R}^3)^4$ transverse to the codimension two subset consisting of quadruples of vectors that lie in a plane.) It follows that there exists an $\varepsilon > 0$ such that γ meets each plane at most four times on a time interval of length ε . (Otherwise, taking a limit of quintuples of times mapped into the same plane whose mutual distances shrink to zero, we would find in the limit an order four tangency of γ to a plane, which we have excluded.) Hence γ can meet each plane at most $4L/\varepsilon$ times.

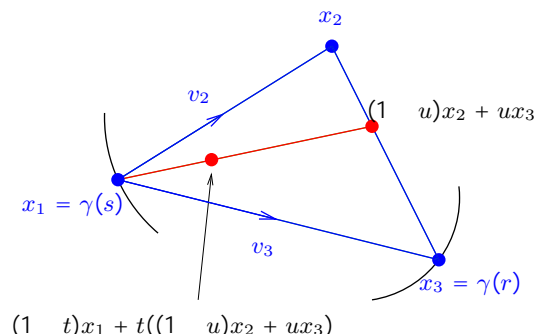


FIGURE 7.1. Parametrization of a triangle.

The proof of part (ii) is contained in the proof of Lemma 7.10(b) below. It relies on choosing K such that its curvature vanishes nowhere.

Now we consider the space of triangles in \mathbb{R}^3 with pairwise distinct corners x_1, x_2, x_3 such that x_1 and x_3 lie on the knot K . Using an arclength parametrization $\gamma : S^1 \rightarrow \mathbb{R}^3 / \mathbb{Z} \cong K$ we identify this space with the open subset

$$T = \{(s, x_2, r) \in S^1 \times \mathbb{R}^3 \mid x_1 = \gamma(s), x_2, x_3 = \gamma(r) \text{ are distinct}\}.$$

We parametrize each triangle $[x_1, x_2, x_3]$ by the map (see Figure 7.1)

$$[0, 1]^2 \rightarrow \mathbb{R}^3, \quad (u, t) \mapsto (1-t)x_1 + t((1-u)x_2 + ux_3).$$

LEMMA 7.3. — *For a generic 1-parameter family of triangles $\beta : [0, 1]^2 \rightarrow T$, $\lambda \nabla (s^\lambda, x_2^\lambda, r^\lambda)$ the following holds.*

(a) *The evaluation map*

$$\text{ev}_\beta : [0, 1]^3 \rightarrow \mathbb{R}^3, \quad (\lambda, u, t) \mapsto (1-t)x_1^\lambda + t((1-u)x_2^\lambda + ux_3^\lambda)$$

is transverse to K on its interior, where we have set $x_1^\lambda = \gamma(s^\lambda)$ and $x_3^\lambda = \gamma(r^\lambda)$.

(b) *The map $(\lambda, u) \nabla \frac{\partial \text{ev}_\beta}{\partial t}(\lambda, u, 0)$ meets the tangent bundle to K transversely in finitely many points. At these points the triangle is tangent to the knot at x_1^λ but not contained in its osculating plane.*

(c) *The points in (b) compactify the set $\text{ev}_\beta^{-1}(K) \setminus [0, 1]^2 \rightarrow (0, 1]$ to an embedded curve in $[0, 1]^3$ transverse to the boundary. Its image in $[0, 1]^2$ under the projection $(\lambda, u, t) \nabla (\lambda, u)$ is an immersed curve with transverse self-intersections.*

Proof. — Part (a) follows from standard transversality arguments. For part (b) we introduce

$$v_2 := x_2 - x_1, \quad v_3 := x_3 - x_1, \quad v := \frac{v_2 \times v_3}{|v_2 \times v_3|}.$$

Thus v_2, v_3 are tangent to the sides of the triangle at x_1 and ν is a unit normal vector to the triangle. So the space of triangles that are tangent to the knot at x_1 is the zero set of the map

$$F : T \rightarrow \mathbb{R}, \quad (s, x_2, r) \mapsto \langle h\dot{\gamma}(s), \nu \rangle = \frac{hv_2, v_3 - \dot{\gamma}(s)j}{jv_2 - v_3j}.$$

The last expression shows that along the zero set the variation of F in direction x_2 (or equivalently v_2) is nonzero provided that $v_3 - \dot{\gamma}(s) \notin 0$. So $F^{-1}(0)$ is a transversely cut out hypersurface in T outside the set T_0 where $v_3 = \gamma(r) - \gamma(s)$ and $\dot{\gamma}(s)$ are collinear. By Lemma 7.2(ii) the set T_0 has codimension 2. Hence a generic curve $\beta : [0, 1] \rightarrow T$ avoids the set T_0 and intersects $F^{-1}(0)$ transversely, which implies the first statement in (b). The second statement in (b) follows similarly from the fact that the set of triangles contained in the osculating plane at x_1 has codimension 2 in T .

For part (c), consider a point (λ_0, u_0) as in (b). To simplify notation, let us shift the parameter interval such that $\lambda_0 = 0$ is an interior point. Then with the obvious notation ν^λ etc the following conditions hold at $\lambda = 0$:

$$a := \langle h\dot{\gamma}(s^0), \nu^0 \rangle = 0, \quad b := \langle h\ddot{\gamma}(s^0), \nu^0 \rangle \neq 0, \quad c := \left. \frac{d}{d\lambda} \right|_{\lambda=0} \langle h\dot{\gamma}(s^\lambda), \nu^\lambda \rangle \neq 0.$$

Here the first condition expresses the fact that the triangle is tangent to the knot at x_1^0 , the second on that the triangle is not contained in the osculating plane, and the third one the transversality of the map in (b) to the tangent bundle of K . Intersections of K with triangles $\beta(\lambda)$ for λ close to zero can be written in the form $\gamma(s^\lambda + s)$ with $s = O(\lambda)$ and must satisfy the equation

$$0 = \langle \gamma(s^\lambda + s) - \gamma(s^\lambda), \nu^\lambda \rangle = \langle s\dot{\gamma}(s^\lambda) + \frac{1}{2}s^2\ddot{\gamma}(s^\lambda) + O(s^3), \nu^\lambda \rangle.$$

Ignoring the trivial solution $s = 0$, we divide by s and obtain using $s = O(\lambda)$:

$$\begin{aligned} 0 &= \left\langle \dot{\gamma}(s^\lambda) + \frac{1}{2}s\ddot{\gamma}(s^\lambda) + O(s^2), \nu^\lambda \right\rangle \\ &= \left\langle \dot{\gamma}(s^0) + \lambda\ddot{\gamma}(s^0) + \frac{1}{2}s\ddot{\gamma}(s^0) + O(\lambda^2), \nu^0 + \lambda\nu^0 + O(\lambda^2) \right\rangle \\ &= \langle h\dot{\gamma}(s^0), \nu^0 \rangle + \lambda \left[\langle h\ddot{\gamma}(s^0), \nu^0 \rangle + \langle h\dot{\gamma}(s^0), \nu^0 \rangle + O(\lambda) \right] + s \left[\frac{1}{2} \langle h\ddot{\gamma}(s^0), \nu^0 \rangle + O(\lambda) \right] \\ &= a + \lambda \left[b + O(\lambda) \right] + s \left[\frac{1}{2}c + O(\lambda) \right]. \end{aligned}$$

Since $a = 0$ and b, c are nonzero, this equation has for each λ a unique solution s of the form

$$s = \frac{2b}{c} \lambda + O(\lambda^2).$$

Now recall that by hypothesis $\dot{\gamma}(s^0)$ is a multiple of $(1 - u^0)v_2^0 + u^0v_3^0$. If it is a positive (resp. negative) multiple, then only solutions with $s > 0$ (resp. $s < 0$) will lie in the triangle. So in either case the solutions describe a curve with boundary and part (c) follows.

REMARK 7.4. — Lemma 7.3 shows that, given a generic 1-parameter family of triangles $\beta : [0, 1] \rightarrow \mathcal{T}$, the associated 2-parameter family $(\lambda, u) \mapsto \text{ev}_\beta(\lambda, u, \cdot)$ can be reparametrized in t to look like the Q -strings in a generic 2-chain of broken strings. To see the last condition (2e) in Definition 5.3, consider a parameter value (λ, u) as in Lemma 7.3(b). Since the triangle is not contained in the osculating plane at x_1^λ , the linear string $t \mapsto \text{ev}_\beta(\lambda, u, t)$ deviates quadratically from the knot, so its projection normal to the knot has nonvanishing second derivative at $t = 0$. Hence we can reparametrize it to make its second derivative vanish and its third derivative nonzero as required in condition (2e). We will ignore these reparametrizations in the following.

REMARK 7.5. — Lemma 7.3 remains true (with a simpler proof) if in the definition of the space of triangles \mathcal{T} we allow x_3 to move freely in \mathbb{R}^3 rather than only on the knot; this situation will also occur in the shortening process in the next subsection.

Let us emphasize that in the space \mathcal{T} we require the points x_1, x_2, x_3 to be distinct. Now in a generic 1-parameter family of triples (x_1, x_2, x_3) with $x_1, x_3 \in K$ the points x_1, x_3 may meet for some parameter values, so this situation is not covered by Lemma 7.3. See Remark 7.7 below on how to deal with this situation.

7.3. REDUCING PIECEWISE LINEAR Q -STRINGS TO LINEAR ONES. — In this subsection we deform chains in Σ_{pl} to chains in Σ_{lin} , not increasing the length of Q -strings in the process. The main result of this subsection is

PROPOSITION 7.6. — *For a generic knot K there exist maps*

$$F_0 : C_0(\Sigma_{\text{pl}}) \rightarrow C_0(\Sigma_{\text{lin}}), \quad F_1 : C_1(\Sigma_{\text{pl}}) \rightarrow C_1(\Sigma_{\text{lin}})$$

and

$$H_0 : C_0(\Sigma_{\text{pl}}) \rightarrow C_1(\Sigma_{\text{pl}}), \quad H_1 : C_1(\Sigma_{\text{pl}}) \rightarrow C_2(\Sigma_{\text{pl}})$$

satisfying with the map i_{lin} from (7.2):

- (i) $F_0 i_{\text{lin}} = \mathbb{1}$ and $DH_0 = i_{\text{lin}} F_0 - \mathbb{1}$;
- (ii) $F_1 i_{\text{lin}} = \mathbb{1}$ and $H_0 D + DH_1 = i_{\text{lin}} F_1 - \mathbb{1}$;
- (iii) F_0, H_0, F_1 and H_1 are (not necessarily strictly) length-decreasing.

Proof. — We assume that K satisfies the genericity properties in Section 7.2. We first construct the maps H_0 and F_0 .

For each simplex $\beta \in C_0^{\text{pl}}(\Sigma)$ we denote by $M(\beta)$ the total number of corners in the Q -strings of β , not counting the corners in Q -spikes (which are by definition 3-gons). Connecting each corner to the starting point of its Q -string, we obtain $M(\beta)$ triangles connecting the various Q -strings to the segments between their end points. We define the *complexity* of $\beta \in C_0^{\text{pl}}(\Sigma)$ to be the pair of nonnegative integers

$$c(\beta) := (M(\beta), I(\beta)),$$

where $I(\beta)$ is the number of interior intersection points of the first triangle with K (we set $I(\beta) = 0$ in the case $M(\beta) = 0$, i.e., if there are no triangles). Note that by part (i) of Lemma 7.2 we know that I is bounded a priori by a fixed constant

$S = S(K)$. We define the maps H_0 and F_0 by induction on the lexicographical order on complexities $c(\beta)$. For $c(\beta) = (0, 0)$ we set $F_0\beta = \beta$ and $H_0\beta = 0$.

For the induction step, let $\beta \geq C_0^{pl}(\Sigma)$ be a 0-simplex and assume that F_0 and H_0 satisfying (i) and (iii) have been defined for all simplices of complexities $c < c(\beta)$. Let the first triangle of β have vertices x_1, x_2, x_3 , where x_1 is the starting point of the first Q -string which is not a segment, and x_2 and x_3 are the next two corners on that Q -string (x_3 might also be the end point). Since there are only finitely many intersections of the knot K with the interior of the triangle (and none with its sides), we can find a segment connecting x_2 to a point x_3^0 on the segment x_1x_3 which is so close to x_3 that the triangle $x_2x_3^0x_3$ does not contain any intersection points with the knot. Let $h\beta \geq C_1^{pl}(\Sigma)$ be the 1-simplex obtained by sweeping the first triangle by the family of segments from x_1 to a varying point $(1 - u)x_2 + ux_3^0$ on the segment $[x_2, x_3^0]$, followed by the segment from that point to x_3 and the remaining segments to x_4 etc. See Figure 7.2 (the point y and the shaded region play no role here and are included for later use). The N -string ending at x_1 (and if there is one, also the

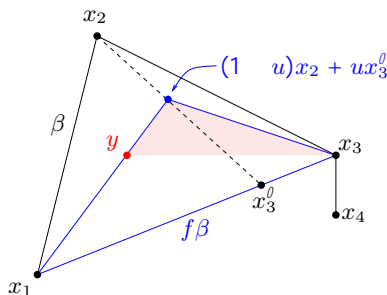


FIGURE 7.2. Reducing the number of corner points.

N -string starting at x_3) is “dragged along” without creating intersections with K , and all remaining N - and Q -strings remain unchanged in the process.

The 1-simplex $h\beta$ has boundary $\partial(h\beta) = \beta^0 - \beta$, where β^0 is the 0-simplex at the end of the sweep with first segment $[x_1, x_3]$. We define

$$f\beta := Dh\beta + \beta = \beta^0 + \delta_Q h\beta + \delta_N h\beta.$$

By construction we have $\delta_N h\beta = 0$ and $M(\beta^0) < M(\beta)$, hence $c(\beta^0) < c(\beta)$. The domain of $\delta_Q h\beta$ consists of those finitely many points where the triangle intersects K in its interior, so that $\delta_Q h\beta$ consists of broken strings with one more Q -string (which is linear) and with the same total number of corners as β . But since the new first triangle is contained in the original first triangle for β , and one of the intersection points is now the starting point of the new Q -string, we have $I(\delta_Q h\beta) < I(\beta)$. Altogether we see that $c(f\beta) < c(\beta)$, so by induction hypothesis F_0 and H_0 are already defined on $f\beta$. We set

$$F_0\beta := F_0f\beta \quad \text{and} \quad H_0\beta := H_0f\beta + h\beta$$

and verify that indeed (using condition (i) on $f\beta$)

$$DH_0\beta = DH_0f\beta + Dh\beta = F_0f\beta \quad f\beta + f\beta \quad \beta = F_0\beta \quad \beta,$$

so condition (i) continues to hold. Condition (iii) holds by induction hypothesis in view of $L(f\beta) \subset L(\beta)$ and $L(h\beta) \subset L(\beta)$. Since every $\beta \in C_0^{\text{pl}}(\Sigma)$ has finite complexity, this finishes the definition of F_0 and H_0 .

We next construct the maps H_1 and F_1 , following the same strategy. For this, we first extend the notion of complexity $c = (M, I)$ to 1-chains with piecewise linear Q -strings. For a 1-simplex $\beta : [0, 1] \rightarrow \Sigma^{\text{pl}}$, we set

$$M(\beta) := \max_{\lambda \in [0,1]} M(\beta(\lambda)), \quad I(\beta) := \max_{\lambda \in [0,1]} I(\beta(\lambda)).$$

Note that $I(\beta)$ is still bounded by the constant $S = S(K)$ in Lemma 7.2. Note also that, according to our definition of chains of piecewise linear strings, the number $M(\beta(\lambda))$ of corner points of Q -strings in $\beta(\lambda)$ is actually constant equal to the maximal number $M(\beta)$. Observe that with this definition of complexity for 1-chains, the maps $h_0 := h$ and H_0 used in the argument for 0-chains do not increase complexity.

Again our definition of F_1 and H_1 proceeds by induction on the lexicographic order on complexity. For simplices $\beta \in C_1^{\text{pl}}(\Sigma)$ with $M = 0$ we set $F_1\beta = \beta + H_0D\beta$ and $H_1\beta = 0$. Then (ii) holds by construction, and (iii) holds since H_0 and D are length-decreasing.

For the induction step, let $\beta \in C_1^{\text{pl}}(\Sigma)$ be a 1-simplex, and assume that F_1 and H_1 satisfying (ii) and (iii) have been defined for all 1-simplices of complexity $c < c(\beta)$. Using a parametrized version of sweeping the first triangle, we obtain a 2-chain $h_1\beta \in C_2^{\text{pl}}(\Sigma)$. By construction its boundary satisfies $\partial h_1\beta + h_0\partial\beta = \beta^0 - \beta$, where β^0 is the 1-simplex at the end of the sweep with first segment $[x_1, x_3]$, see Figure 7.3. We now

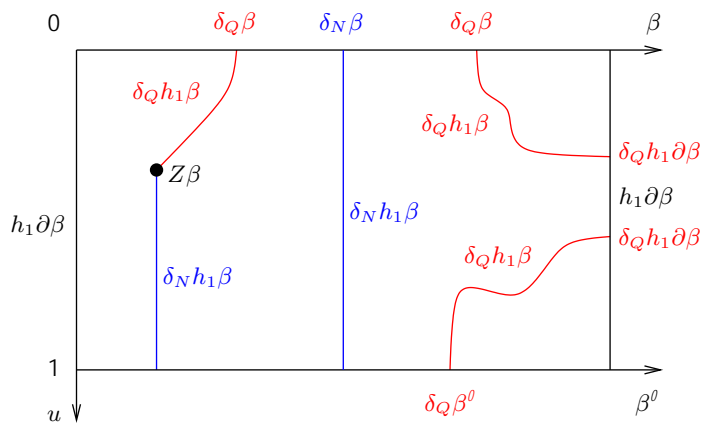


FIGURE 7.3. The domain of $h_1\beta$.

define

$$\begin{aligned} f_1\beta &:= Dh_1\beta + h_0D\beta + \beta \\ &= \beta^0 + (\delta_Q h_1 + h_0\delta_Q)\beta + (\delta_N h_1 + h_0\delta_N)\beta. \end{aligned}$$

We claim that $c(f_1\beta) < c(\beta)$. To see this, we need to show that the three terms on the right hand side of the last displayed equation have complexity lower than $c(\beta)$. For β^0 this holds because its Q -strings have one fewer corner, i.e., $M(\beta^0) < M(\beta)$. The domain of $(\delta_Q h_1 + h_0\delta_Q)\beta$ consists of the finitely many curves in which the first triangle intersects K at an interior point y , so that $(\delta_Q h_1 + h_0\delta_Q)\beta$ consists of broken strings with one more Q -string (which is linear) and with the same total number of corners as β . But since the new first triangle (the shaded region in Figure 7.2) is contained in the original first triangle for each parameter value in β , and one of the intersection points is now the starting point of the new Q -string, we have $I((\delta_Q h_1 + h_0\delta_Q)\beta) < I(\beta)$. The domain of $(\delta_N h_1 + h_0\delta_N)\beta$ consists of the finitely many straight line segments $[u, 1] \times \mathcal{L}g$ emanating from the parameter values (u, λ) corresponding to the tangencies of the triangle $[x_1, x_2, x_3]$ to the knot at x_1 , see Figure 7.3 where one such point of tangency is shown as $Z\beta$. So $(\delta_N h_1 + h_0\delta_N)\beta$ consists of broken strings with one more Q -spike and with the same total number of corners as β . But since the new triangle with corners $x_1, (1-u)x_2 + ux_3, x_3$ is contained in the original first triangle at parameter value λ , and one of the intersection points with the knot is the corner point x_1 of the new triangle (which does not count towards I), we have $I((\delta_N h_1 + h_0\delta_N)\beta) < I(\beta)$ and the claim is proved.

According to the claim, F_1 and H_1 are defined on $f_1\beta$ and we set

$$F_1\beta := F_1f_1\beta \quad \text{and} \quad H_1\beta := H_1f_1\beta + h_1\beta.$$

To distinguish the proposed extensions from the maps given by induction hypothesis, we temporarily call the extended versions H_1 and F_1 , so we can write

$$F_1 := F_1f_1 \quad \text{and} \quad H_1 := H_1f_1 + h_1$$

without ambiguity. Recall also that in this notation $H_0 = H_0f_0 + h_0$. Now using $f_1 = h_0D + Dh_1 + \mathbb{1}$ we compute

$$\begin{aligned} DH_1 + H_0D &= DH_1f_1 + Dh_1 + H_0f_0D + h_0D \\ &= (F_1f_1 - f_1 - H_0Df_1) + (f_1 - \mathbb{1} - h_0D) + H_0f_0D + h_0D \\ &= F_1 - \mathbb{1} + H_0(f_0D - Df_1). \end{aligned}$$

Using $f_1 = h_0D + Dh_1 + \mathbb{1}$ again and $f_0 = Dh_0 + \mathbb{1}$, we find $Df_1 = Dh_0D + D = (Dh_0 + \mathbb{1})D = f_0D$, so that the last term in the displayed equation vanishes and the extensions H_1, F_1 have the required properties. This completes the induction step and hence the proof of Proposition 7.6.

REMARK 7.7. — If in a 1-simplex β as in the preceding proof the third point x_3 of the first triangle is the end point of the corresponding Q -string and thus constrained to lie on the knot, then the points x_1 and x_3 can cross each other for some parameter values λ in the chain. The homotopy $h_1\beta$ then shrinks the corresponding degenerate

triangle at parameter λ to a constant Q -string, which according to our convention from Section 7.1 we interpret as a linear Q -spike in the direction of the degenerate triangle. Incidentally, the segment $[x_2, x_3]$ is always short throughout the shortening process, so if x_1 and x_3 agree then the triangle is already a linear Q -spike without further shrinking.

REMARK 7.8. — Definition 5.4 implies that if a Q -string in β in the preceding proof is a (piecewise linear) Q -spike, then it never intersects the knot in its interior and remains a Q -spike throughout the shortening process (which ends with a degenerate triangle as in Remark 7.7). This property ensures that H_0 and H_1 indeed do not increase length, which does not count Q -spikes.

REMARK 7.9. — The proof relies crucially on the (trivial) fact that the new triangle $[y, (1 - u)x_2 + ux_3, x_3]$ (the shaded region in Figure 7.2) obtained by splitting the Q -string at an intersection point y with K is contained in the old triangle $[x_1, x_2, x_3]$. This is the only place where we use that the metric is Euclidean; the rest of the proof works equally well for any metric of nonpositive curvature.

7.4. PROPERTIES OF LINEAR Q -STRINGS FOR GENERIC KNOTS. — Now we consider the space of 2-gons, i.e., straight line segments starting and ending on the knot. This space is canonically identified with $K \times K$ by associating to each 2-gon its endpoints on K . We consider the squared distance function

$$E : K \times K \rightarrow \mathbb{R}, \quad E(x, y) = \frac{1}{2} |x - y|^2.$$

LEMMA 7.10. — For a generic knot $K \subset \mathbb{R}^3$ the following holds for the space $K \times K$ of 2-gons (see Figure 7.4).

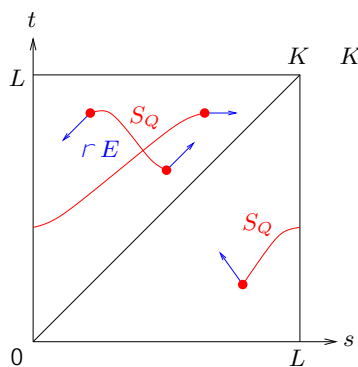


FIGURE 7.4. The space of 2-gons.

(a) E attains its minimum 0 along the diagonal, which is a Bott nondegenerate critical manifold; the other critical points are nondegenerate binormal chords of index 0, 1, 2.

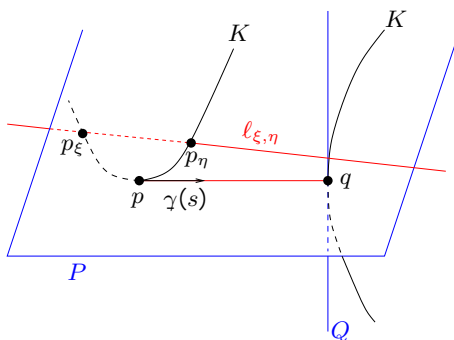


FIGURE 7.5. A 2-gon becoming tangent to K at an endpoint.

(b) The subset $S_Q \subset K \times K$ of 2-gons meeting K in their interior is a 1-dimensional submanifold with boundary consisting of finitely many 2-gons tangent to K at one endpoint, and with finitely many transverse self-intersections consisting of finitely many 2-gons meeting K twice in their interior.

(c) The negative gradient $-\nabla E$ is not pointing into S_Q at the boundary points.

Proof

(a) In terms of an arclength parametrization γ of K we write the energy as a function $E(s, t) = \frac{1}{2}j\gamma(s) - \gamma(t)j^2$. We compute its partial derivatives

$$(7.3) \quad \begin{aligned} \frac{\partial E}{\partial s} &= h\gamma(s) - \gamma(t), \dot{\gamma}(s)i, & \frac{\partial E}{\partial t} &= h\gamma(t) - \gamma(s), \dot{\gamma}(t)i, \\ \frac{\partial^2 E}{\partial s^2} &= j\dot{\gamma}(s)j^2 + h\gamma(s) - \gamma(t), \ddot{\gamma}(s)i, & \frac{\partial^2 E}{\partial s \partial t} &= h\dot{\gamma}(s), \dot{\gamma}(t)i, \\ & & \frac{\partial^2 E}{\partial t^2} &= j\dot{\gamma}(t)j^2 + h\gamma(t) - \gamma(s), \ddot{\gamma}(t)i. \end{aligned}$$

We see that critical points of E are points on the diagonal $s = t$ and binormal chords (where $s \neq t$), and the Hessian of E at $s = t$ equals $\begin{pmatrix} 1 & \\ & 1 \end{pmatrix}$. Its kernel is the tangent space to the diagonal and it is positive definite in the transverse direction. This proves Bott nondegeneracy of the diagonal. Nondegeneracy of the binormal chords is achieved by a generic perturbation of K .

(b) We choose K so that its curvature is nowhere 0 (which holds generically). Then there exists $\delta > 0$ such that no 2-gon of positive length $< \delta$ intersects the knot in an interior point. Consider the tangential variety τ_K of K (where $\gamma: [0, L] \rightarrow \mathbb{R}^3$ is a parametrization of K)

$$\tau_K := \{r\gamma(s) + r\dot{\gamma}(s) \mid s \in [0, L], r \in \mathbb{R}\} \subset \mathbb{R}^3.$$

Since the curvature of K is nowhere zero, there exists $\delta > 0$ such that for each s the line segment $r\gamma(s) + r\dot{\gamma}(s) \mid r \in (-\delta, \delta)$ intersects K only at $r = 0$. Let $N(\delta)$ denote the union of these line segments. After small perturbation, the surface $\tau_K \cap N(\delta)$

intersects K transversely. This shows that there are finitely many 2-gons that are tangent to K at one endpoint and that this is a transversely cut out 0-manifold. Moreover, transversality implies that for each 2-gon that is tangent to K at one endpoint p , the tangent line Q to K at the other endpoint q does not lie in the osculating plane P (the plane spanned by the first two derivatives of γ) at p ; see Figure 7.5.

We claim that the 2-gon $[p, q]$ is the boundary point of a unique local embedded curve of 2-gons intersecting K in their interior. To see this, we choose a set of coordinates (x, y, z) on \mathbb{R}^3 in which $p = (0, 0, 0)$, $q = (1, 0, 0)$, P is the (x, y) -plane, and Q is parallel to the z -axis. Then K can be written near p as a graph over the x -axis in the form

$$y = \kappa x^2 + O(x^3), \quad z = O(x^3),$$

and near q as a graph over the z -axis in the form

$$x = 1 + O(z^2), \quad y = O(z^2).$$

Here $2\kappa \neq 0$ is the curvature of K at p , and after a further reflection we may assume that $\kappa > 0$. We fix a small $\varepsilon > 0$ (to be chosen later) and consider points ξ, η on the x -axis with $-\varepsilon < \xi < \eta < 2\varepsilon$. Let p_ξ, p_η be the points of K near p with x -coordinates ξ, η and let $\ell_{\xi, \eta}$ be the line through p_ξ and p_η . Let $\pi(x, y, z) = (x, z)$ be the projection onto the (x, z) -plane. Since the line $\ell_{\xi, \eta}$ is close to the x -axis and K is tangent to the z -axis at q , the projected curves $\pi(\ell_{\xi, \eta})$ and $\pi(K)$ intersect in a unique point $r_{\xi, \eta}$ in the (x, z) -plane near $\pi(q) = (1, 0)$. Let $f_\xi(\eta)$ denote the difference in the y -values between the points of K and $\ell_{\xi, \eta}$ lying over $r_{\xi, \eta}$. Thus $f_\xi(\eta)$ is the “distance in the y -direction” between $\ell_{\xi, \eta}$ and K near q . To compute the function $f_\xi(\eta)$, note that the slope of the line through the points $(\xi, \kappa\xi^2)$ and $(\eta, \kappa\eta^2)$ on the parabola $y = \kappa x^2$ equals $\kappa(\xi + \eta)$, so the y -value of this line at $x = 1$ is of the form $\kappa(\xi + \eta) + O(\xi^2 + \eta^2)$. The linear term persists for the function $f_\xi(\eta)$, hence

$$f_\xi(\eta) = \kappa(\xi + \eta) + O(\xi^2 + \eta^2).$$

For ε sufficiently small, we see that if $\xi > 0$, then $f_\xi(\eta) > 0$ for all $\eta \geq (\xi, 2\varepsilon)$. Suppose therefore that $\xi < 0$. Then for ε sufficiently small we have $f_\xi(0) = \kappa\xi + O(\xi^2) < 0$, $f_\xi(-2\xi) = -\kappa\xi + O(\xi^2) > 0$, and $f_\xi^0(\eta) = \kappa + O(|\xi| + |\eta|) > 0$. Thus for every $\xi \geq (-\varepsilon, 0)$ there exists a unique $\eta(\xi) \geq (\xi, 2\varepsilon)$ such that $f_\xi(\eta(\xi)) = 0$, i.e., the line $\ell_{\xi, \eta(\xi)}$ intersects K near q . Moreover, the point $\eta(\xi)$ depends smoothly on ξ and satisfies $0 < \eta(\xi) < 2\xi$. This shows that the 2-gons with endpoints near p, q intersecting K in their interior form a smooth curve parametrized by $\xi \geq (-\varepsilon, 0)$, consisting of the corresponding segments of the lines $\ell_{\xi, \eta(\xi)}$. As this curve extends smoothly to $\xi = 0$ by the 2-gon $[p, q]$, the claim is proved.

So we have shown that the subset $S_Q \subset K \times K$ avoids a neighborhood of the diagonal and is a 1-manifold with boundary near the finitely many 2-gons that are tangent to K at an endpoint. Away from these sets, a generic perturbation of K makes the evaluation map at the interior of the 2-gons transverse to K . Since the condition that a chord meets K in the interior is codimension one, and the condition that the

tangent line at the intersection is parallel to the chord is of codimension three and can thus be avoided for generic K , we conclude that (b) holds.

(c) Consider a boundary point of S_Q , i.e., a 2-gon $[p, q]$ tangent to K at one endpoint, say at p . Let $p = \gamma(s)$ and $q = \gamma(t)$ for an arclength parametrization of K such that $\dot{\gamma}(s)$ is a positive multiple of $q - p$; see Figure 7.5. By equation (7.3) we have $\partial E/\partial s = \langle \dot{\gamma}(s), \gamma(s) - p \rangle < 0$, so the parameter s strictly increases in the direction of ∇E . On the other hand, the description in (b) shows that s strictly decreases as we move into S_Q . Hence ∇E is not pointing into S_Q at $[p, q]$.

More generally, for an integer $\ell > 1$ we consider the space $(K \times K)^\ell$ of ℓ -tuples of 2-gons with the energy and length functions $E^\ell, L^\ell : (K \times K)^\ell \rightarrow \mathbb{R}$,

$$E^\ell(x_1, y_1, \dots, x_\ell, y_\ell) := \frac{1}{2} \sum_{i=1}^{\ell} |x_i - y_i|^2,$$

$$L^\ell(x_1, y_1, \dots, x_\ell, y_\ell) := \sum_{i=1}^{\ell} |x_i - y_i|.$$

As a consequence of Lemma 7.10, E^ℓ is a Morse-Bott function whose critical manifolds are products $C_1 \times \dots \times C_\ell$ of critical manifolds of E , so each C_i is either a binormal chord or the corresponding diagonal. Note that the symmetric group S_ℓ acts on $(K \times K)^\ell$ preserving E^ℓ as well as the product metric.

For $a > 0$ we denote by $M^a \subset (K \times K)^\ell$ the collection of tuples $c = (c_1, \dots, c_\ell)$ of binormal chords of total length $L(c) = a$, and by W^a the disjoint union of the unstable manifolds of points in M^a under the flow of $-\nabla E^\ell$ (here M^a and thus W^a may be empty). Let $\phi^T : (K \times K)^\ell \rightarrow (K \times K)^\ell$ be the time- T map of the flow of $-\nabla E^\ell$.

LEMMA 7.11. — For a generic knot $K \subset \mathbb{R}^3$ and each $a > 0$ there exist $\varepsilon_a > 0$ and $T_a > 0$ with the following property. For each $\varepsilon < \varepsilon_a$, $T > T_a$ and $\ell \geq 2 \in \mathbb{N}$ we have

$$\phi^T(fL^\ell \subset a + \varepsilon g) \subset fL^\ell \subset a - \varepsilon g \cup V^a,$$

where V^a is a tubular neighborhood of $W^a \setminus fL^\ell \subset a - \varepsilon g$ in $f(a - \varepsilon) \subset L^\ell \subset a + \varepsilon g$. Moreover, tuples of Q -strings in V^a do not intersect the knot K in their interior.

Proof. — Note that on $K \times K$ the length and energy are related by $L = \sqrt{2E}$, so they have the same critical points and L is strictly decreasing under the flow of $-\nabla E$ outside the critical points. Since the flow of $-\nabla E^\ell$ is the product of the flows of $-\nabla E$ in each factor, the same relation holds for any $\ell \geq 2 \in \mathbb{N}$: L^ℓ and E^ℓ have the same critical points and L^ℓ is strictly decreasing under the flow of $-\nabla E^\ell$ outside the critical points.

Next recall from above that E^ℓ is a Morse-Bott function. In particular, the set of critical values of E^ℓ , and thus also of L^ℓ , is discrete. Given $a \geq 2R$, we pick $\varepsilon_a > 0$ such that a is the only critical value of L^ℓ in the interval $[a - \varepsilon_a, a + \varepsilon_a]$. (Since only finitely many binormal chords can appear in tuples of critical points of total length a , the constant ε_a can be chosen independently of ℓ .) For $\varepsilon < \varepsilon_a$, the familiar argument from Morse theory shows that $\phi^T(fL^\ell \subset a + \varepsilon g) \subset fL^\ell \subset a - \varepsilon g \cup V_{\varepsilon, T}^a$, where $V_{\varepsilon, T}^a$

for large T are tubular neighborhoods of $W^a \setminus fL^\ell > a - \varepsilon g$ in $f_a - \varepsilon \subset L^\ell \subset a + \varepsilon g$ that shrink to $W^a \setminus fL^\ell > a - \varepsilon g$ as $T \rightarrow 1$.

For the last statement recall that, for a generic knot K , binormal chords do not meet K in their interior. So for each $\ell \geq \mathbb{N}$ there exists a neighborhood U^a of M^a in $(K \times K)^\ell$ such that tuples of Q -strings in U^a do not intersect K in their interior. We pick T_a large enough and ε_a small enough so that V_{ε_a, T_a}^a is contained in U^a . By the argument as in the previous paragraph, the constants ε_a and T_a can be chosen independently of ℓ and the lemma is proved.

7.5. SHORTENING LINEAR Q -STRINGS. — We will need some homological algebra. Suppose we have the following algebraic situation:

- a chain complex $(C, D = \partial + \delta)$ satisfying the relations

$$\partial^2 = \delta^2 = \partial\delta + \delta\partial = 0, \quad \text{and}$$

- a chain map $f : (C, \partial) \rightarrow (C, \partial)$ and a chain homotopy $H : (C, \partial) \rightarrow (C, \partial)$ satisfying

$$(7.4) \quad \partial H + H\partial = f - \mathbb{1},$$

such that for every $c \in C$ there exists a positive integer $S(c)$ with

$$(7.5) \quad (\delta H)^{S(c)}(c) = 0.$$

In our applications below, we will have $\delta = \delta_Q + \delta_N$, and the equation $\delta^2 = 0$ will follow from

$$\delta_Q^2 = \delta_N^2 = [\delta_Q, \delta_N] = 0,$$

which is part of the statement that $D^2 = 0$ in our chain complex. Here, as usual, we denote the graded commutator of two maps A, B by

$$[A, B] := AB - (-1)^{j_A j_B} BA.$$

Set $H_0 := H$ and $f_0 := f$, and more generally for $d > 1$ define the maps

$$(7.6) \quad H_d := H(\delta H)^d, \quad f_d := \sum_{i=0}^d (H\delta)^i f (\delta H)^{d-i}.$$

It is also convenient to set $H_{-1} = 0$. Note that the maps f_d satisfy the recursion relation $f_{d+1} = f_d \delta H + H_d \delta f$.

LEMMA 7.12. — For each $d > 1$ we have

$$(7.7) \quad [\partial, H_d] + [\delta, H_{d-1}] = f_d.$$

Proof. — We prove this by induction on d . The case $d = 1$ is an immediate consequence of (7.4) and $[\delta, \partial] = 0$. For the induction step we observe that

$$\begin{aligned} [\partial, H_{d+1}] &= \partial H_d \delta H + H_d \delta H \partial \\ &= [\partial, H_d] \delta H - H_d \partial \delta H - H_d \delta \partial H + H_d \delta f - H_d \delta \\ &= f_d \delta H - [\delta, H_{d-1}] \delta H + H_d \delta f - H_d \delta \\ &= f_{d+1} - \delta H_d - H_{d-1} \delta^2 H - H_d \delta \\ &= f_{d+1} - [\delta, H_d]. \end{aligned}$$

Here in the second equality we have used (7.4), in the third equality the induction hypothesis and $[\delta, \partial] = 0$, in the fourth equality the recursion relation above, and in the fifth equality we have used $\delta^2 = 0$.

In view of equation (7.5), for each $c \in C$ we have $H_d c = 0$ and $f_d c = 0$ for

$$d > S(c) + \max\{S(fc), S(f\delta Hc), \dots, S(f(\delta H)^{S(c)-1}c)\}.$$

So the sums

$$(7.8) \quad H := \sum_{d=0}^1 H_d, \quad F := \sum_{d=0}^1 f_d$$

are finite on every $c \in C$. Summing up equation (7.7) for $d = 1, \dots, e$ and using equation (7.4), we obtain

$$[\partial, H_e] + [D, H_0 + \dots + H_{e-1}] = f_0 + \dots + f_e \mathbb{1}$$

for all e , and hence

$$[D, H] = F \mathbb{1}.$$

This concludes the homological algebra discussion.

We now apply this construction to the space Σ_{lin} of broken strings with linear Q -strings as follows. We fix a large time $T > 0$ and consider a generic i -chain β in Σ_{lin} , for $i = 0, 1$. Moving the Q -strings in β by the flow of τ_E for times $t \in [0, T]$ we obtain an $(i + 1)$ -chain in $(K - K)^\ell$. We make this an $(i + 1)$ -chain $H^T \beta$ in Σ_{lin} by dragging along the N -strings without creating new intersections with the knot. In the case $i = 1$, we moreover grow new N -spikes starting from the finitely many points $Z\beta$ where some Q -string becomes tangent to the knot at one end point, as shown in Figure 7.3. We define $f^T \beta$ as the boundary component of $H^T \beta$ at time T .

REMARK 7.13. — Technically, we should be careful to arrange that H maps generic chains to generic chains. This is easy for 0-chains, but some care should be taken for 1-chains, especially near the points $Z\beta$ where some Q -string becomes tangent to K at one of its end points.

PROPOSITION 7.14. — *For a generic knot K , the operations defined above yield for $i = 0, 1$ maps*

$$f^T : C_i(\Sigma_{\text{lin}}) \rightarrow C_i(\Sigma_{\text{lin}}), \quad H^T : C_i(\Sigma_{\text{lin}}) \rightarrow C_{i+1}(\Sigma_{\text{lin}})$$

satisfying conditions (7.4) and (7.5).

Proof. — Standard transversality arguments show that f^T and H^T map generic chains to generic chains, provided that we impose suitable genericity conditions on generic chains with respect to linear strings. Now condition (7.4) is clear by construction.

For condition (7.5), we use Lemma 7.10(c). It implies that there exists a neighborhood $U \cap K \cap K$ of the finitely many 2-gons ∂S_Q that are tangent to K at one endpoint and an $\varepsilon > 0$ with the following property: Each 2-gon in $U \setminus S_Q$ decreases in length by at least ε under the flow of τE before it meets S_Q again, and the same holds for the longer 2-gon resulting from splitting it at its intersection with the knot. On the other hand, if a 2-gon in $S_Q \cap U$ is split at its intersection with the knot, then both pieces are shorter by at least some fixed amount $\delta > 0$. Hence each application of $H^T \delta_Q$ decreases the total length of Q -strings by at least $\min(\varepsilon, \delta)$, and since $L(\beta)$ is finite this can happen only finitely many times.

Applying definition (7.8) to the maps f^T and H^T , we obtain for $i = 0, 1$ length decreasing maps

$$F^T : C_i(\Sigma_{\text{lin}}) \rightarrow C_i(\Sigma_{\text{lin}}), \quad H^T : C_i(\Sigma_{\text{lin}}) \rightarrow C_{i+1}(\Sigma_{\text{lin}})$$

satisfying

$$(7.9) \quad DH_0^T = F_0^T \circ \mathbb{1}, \quad H_0^T D + DH_1^T = F_1^T \circ \mathbb{1}$$

We now use these maps to compute the homology of $(C_i(\Sigma_{\text{lin}}), D)$ in small length intervals. For $a \geq R$ and $i = 0, 1$ we denote by A_i^a the free \mathbb{Z} -module generated by words $\gamma_1 c_1 \dots \gamma_\ell c_\ell \gamma_{\ell+1}$, $\ell > 0$, where c_1, \dots, c_ℓ are binormal chords of total length a and of total index i , and the γ_j are homotopy classes of paths in ∂N connecting the c_j to broken strings and not intersecting K in their interior. We define linear maps

$$\Theta : A_i^a \rightarrow H_i^{[a-\varepsilon, a+\varepsilon]}(\Sigma_{\text{lin}}, D)$$

as follows. For $i = 0$, Θ sends $\gamma_1 c_1 \dots \gamma_\ell c_\ell \gamma_{\ell+1}$ to the homology class of the broken string $\tilde{\gamma}_1 c_1 \dots \tilde{\gamma}_\ell c_\ell \tilde{\gamma}_{\ell+1}$, where $\tilde{\gamma}_j$ are representatives of the classes γ_j . For $i = 1$, consider a word $\gamma_1 c_1 \dots \gamma_\ell c_\ell \gamma_{\ell+1}$ with exactly one binormal chord c_k of index 1 and all others of index 0. Then Θ sends this word to the homology class of the 1-chain $\tilde{\gamma}_1 c_1 \dots \tilde{c}_k \dots \tilde{\gamma}_\ell c_\ell \tilde{\gamma}_{\ell+1}$, where $\tilde{\gamma}_j$ are representatives of the classes γ_j and \tilde{c}_k is the unstable manifold of c_k in $(K \cap K) \setminus \mathcal{L} > a - \varepsilon g$, viewed as a 1-chain by fixing some parametrization.

COROLLARY 7.15. — *For $a \geq R$ let ε_a be the constant from Lemma 7.11. Then for each $\varepsilon < \varepsilon_a$ the map $\Theta : A_i^a \rightarrow H_i^{[a-\varepsilon, a+\varepsilon]}(\Sigma_{\text{lin}}, D)$ is an isomorphism for $i = 0$ and surjective for $i = 1$.*

Proof. — We first consider the case $i = 1$. Fix $\varepsilon < \varepsilon_a$ and $T > T_a$, where ε_a, T_a are the constants from Lemma 7.11. Consider a relative 1-cycle $\beta \in C_1^{[a-\varepsilon, a+\varepsilon]}(\Sigma_{\text{lin}})$. In view of (7.9), β is homologous to $F^T \beta$. Recall from its definition in (7.6) and (7.8) that each tuple of Q -strings appearing in $F^T \beta$ is obtained by flowing some tuple of

Q -strings for time T (and maybe applying $H\delta_Q$ several times to the resulting tuple). Now we distinguish two cases.

Case 1. — a is not the length of a word of binormal chords. Then in Lemma 7.11 the set V^a is empty and it follows that all tuples of Q -strings in $F^T\beta$ have length at most $a - \varepsilon$. This shows that $H_1^{[a - \varepsilon, a + \varepsilon]}(\Sigma_{\text{lin}}) = 0$ and the map Θ is an isomorphism.

Case 2. — a is the length of a word of binormal chords. For simplicity, let us assume that up to permutation there is only one word w of length a (the general case differs just in notation). By Lemma 7.11, $F^T\beta$ is a finite sum $\beta_1^0 + \beta_2^0 + \dots$ of relative 1-cycles β_ℓ^0 in tubular neighborhoods V^a of the unstable manifolds $W^a \setminus fL^\ell > a - \varepsilon g$ of critical ℓ -tuples of length a . Recall that critical ℓ -tuples consist of binormal chords and Q -spikes (corresponding to constant 2-gons). Using the operation δ_N , we can replace Q -spikes by differences of N -strings to obtain a relative 1-cycle β^{00} in V^a homologous to $F^T\beta$ which contains no Q -spikes. So each 1-simplex β_j^{00} in β^{00} is a relative 1-chain whose Q -strings lie in the tubular neighborhood V_j of the unstable manifold of some permutation w_j of w . Then the N -strings in β^{00} do not intersect the knot in their interior, and by Lemma 7.11 neither do the Q -strings. Thus each β_j^{00} is a relative cycle in V_j with respect to the singular boundary ∂ . We distinguish three subcases.

(i) If the total degree of the word w is bigger than 1, then its stable manifold for the flow of rE has codimension bigger than 1. So, after a small perturbation, each β_j^{00} will avoid the stable manifold of w_j and will therefore have length at most $a - \varepsilon$ for sufficiently large T . This shows that, as in Case 1, both groups vanish and Θ is an isomorphism.

(ii) If the degree of the word w is 0, then its unstable manifold is a point and thus each V_j is contractible relative to $fL \subset a - \varepsilon g$. It follows that each relative cycle β_j^{00} is ∂ -exact, and since no δ_Q and δ_N occurs also D -exact. Again we see that both groups vanish and Θ is an isomorphism.

(iii) If the degree of the word w is 1, then each V_j deformation retracts relative to $fL \subset a - \varepsilon g$ onto the 1-dimensional unstable manifold \tilde{w}_j of w_j . It follows that each relative cycle β_j^{00} is ∂ -homologous, and since no δ_Q and δ_N occurs also D -homologous, to a multiple of the 1-chain of Q -strings \tilde{w}_j connected by suitable N -strings. By definition of Θ , this shows that the D -homology class $[\beta^{00}] = [\beta]$ lies in the image of Θ . So Θ is surjective, which concludes the case $i = 1$.

In the case $i = 0$, the proof of surjectivity is analogous but simpler than in the case $i = 1$. For injectivity one considers $F^T\beta$ for a 1-chain β in Σ_{lin} with $D\beta = \alpha$ for a given 0-chain α and argues similarly. Note that this last step does not work to prove injectivity for $i = 1$ because it would require considering $F^T\beta$ for a 2-chain β , which we have not defined (although this should of course be possible).

7.6. PROOF OF THE ISOMORPHISM. — Let $\Phi : (C(\mathcal{R}), \partial) \rightarrow (C(\Sigma), D)$ be the chain map constructed in the previous section. We now use the fact (Corollary 6.15) that the map Φ preserves the length filtrations. Thus for $a < b < c$ we have the commuting

diagram with exact rows of length filtered homology groups

$$\begin{array}{ccccccccc}
 H_1^{[b,c]}(R) & \rightarrow & H_0^{[a,b]}(R) & \rightarrow & H_0^{[a,c]}(R) & \rightarrow & H_0^{[b,c]}(R) & \rightarrow & 0 \\
 \downarrow \Phi & & \downarrow \Phi & & \downarrow \Phi & & \downarrow \Phi & & \downarrow \\
 H_1^{[b,c]}(\Sigma) & \rightarrow & H_0^{[a,b]}(\Sigma) & \rightarrow & H_0^{[a,c]}(\Sigma) & \rightarrow & H_0^{[b,c]}(\Sigma) & \rightarrow & 0.
 \end{array}$$

The main result of this section asserts that Φ is an isomorphism (resp. surjective) for sufficiently small action intervals:

PROPOSITION 7.16. — *For each $a \geq R$ there exists an $\varepsilon_a > 0$ such that for each $\varepsilon < \varepsilon_a$ the map*

$$\Phi : H_0^{[a-\varepsilon, a+\varepsilon]}(R) \rightarrow H_0^{[a-\varepsilon, a+\varepsilon]}(\Sigma)$$

is an isomorphism and the map

$$\Phi : H_1^{[a-\varepsilon, a+\varepsilon]}(R) \rightarrow H_1^{[a-\varepsilon, a+\varepsilon]}(\Sigma)$$

is surjective.

This proposition implies Theorem 1.2 as follows. Since $H_0(R) = \lim_{R \downarrow} H_0^{[0,R]}(R)$ and $H_0(\Sigma) = \lim_{R \downarrow} H_0^{[0,R]}(\Sigma)$, it suffices to show that

$$\Phi : H_0^{[0,R]}(R) \rightarrow H_0^{[0,R]}(\Sigma)$$

is an isomorphism for each $R > 0$. Now the compact interval $[0, R]$ is covered by finitely many of the open intervals $(a - \varepsilon_a, a + \varepsilon_a)$, with $a \geq [0, R]$ and ε_a as in Proposition 7.16. Thus, according to Proposition 7.16, there exists a partition $0 = r_0 < r_1 < \dots < r_N = R$ such that the maps

$$\Phi : H_0^{[r_{i-1}, r_i]}(R) \rightarrow H_0^{[r_{i-1}, r_i]}(\Sigma)$$

are isomorphisms and

$$\Phi : H_1^{[r_{i-1}, r_i]}(R) \rightarrow H_1^{[r_{i-1}, r_i]}(\Sigma)$$

are surjective for all $i = 1, \dots, N$. To prove by induction that

$$\Phi : H_0^{[0, r_i]}(R) \rightarrow H_0^{[0, r_i]}(\Sigma)$$

is an isomorphism for each $i = 1, \dots, N$, consider the commuting diagram above with $a = 0$, $b = r_{i-1}$ and $c = r_i$. By induction hypothesis for $i - 1$ the second, fourth and fifth vertical maps are isomorphisms and the first one is surjective, so by the five lemma the third vertical map is an isomorphism as well. This proves the inductive step and hence Theorem 1.2.

Proof of Proposition 7.16. — Let us denote the maps provided by Proposition 7.1 by $F_i^{\text{pl}}, H_i^{\text{pl}}$ and the maps in Proposition 7.6 by $F_i^{\text{lin}}, H_i^{\text{lin}}, i = 0, 1$. A short computation shows that the maps

$$\begin{aligned}
 F_i &:= F_i^{\text{lin}} - F_i^{\text{pl}} : C_i(\Sigma) \rightarrow C_i(\Sigma_{\text{lin}}), \\
 H_i &:= H_i^{\text{pl}} + i_{\text{pl}} - H_i^{\text{lin}} - F_i^{\text{pl}} : C_i(\Sigma) \rightarrow C_{i+1}(\Sigma)
 \end{aligned}$$

for $i = 0, 1$ satisfy with the map $i := i_{\text{pl}} - i_{\text{lin}} : C(\Sigma_{\text{lin}}) \rightarrow C(\Sigma)$:

- (i) $F_0 i = \mathbb{1}$ and $DH_0 = iF_0 - \mathbb{1}$;
- (ii) $F_1 i = \mathbb{1}$ and $H_0 D + DH_1 = iF_1 - \mathbb{1}$;
- (iii) F_0, H_0, F_1 and H_1 are (not necessarily strictly) length-decreasing.

Conditions (i) and (ii) imply $DF_1 = F_0 D$ and $iF_1 D = D(\mathbb{1} + H_1 D)$, and therefore

$$F_0(\text{im } D) \subset \text{im } D, \quad F_1(\ker D) \subset \ker D, \quad F_1(\text{im } D) \subset i^{-1}(\text{im } D).$$

Hence the F_i define chain maps between the chain complexes (where the left horizontal maps are the obvious inclusions)

$$\begin{array}{ccccc} \text{im } D & \hookrightarrow & C_1(\Sigma) & \xrightarrow{D} & C_0(\Sigma) \\ \downarrow F_1 & & \downarrow F_1 & & \downarrow F_0 \\ i^{-1}(\text{im } D) & \hookrightarrow & C_1(\Sigma^{\text{lin}}) & \xrightarrow{D^{\text{lin}}} & C_0(\Sigma^{\text{lin}}). \end{array}$$

Note that the upper complex computes the homology groups $H_0(\Sigma)$ and $H_1(\Sigma)$, while the lower complex has homology groups $H_0(\Sigma^{\text{lin}})$ and

$$\widehat{H}_1(\Sigma^{\text{lin}}) := \ker D^{\text{lin}} / i^{-1}(\text{im } D).$$

Conditions (i) and (ii) show that F_0, F_1 induce isomorphisms between these homology groups (with inverses i^{-1}), and in view of condition (iii) the same holds for length filtered homology groups. Setting

$$\Psi := F_i \circ \Phi : (C_i(R), \partial) \rightarrow (C_i(\Sigma^{\text{lin}}), D), \quad i = 0, 1,$$

it therefore suffices to prove: *For each $a \in \mathbb{R}$ there exists an $\varepsilon_a > 0$ such that for each $\varepsilon < \varepsilon_a$ the map*

$$\Psi : H_0^{[a-\varepsilon, a+\varepsilon]}(R) \rightarrow H_0^{[a-\varepsilon, a+\varepsilon]}(\Sigma^{\text{lin}})$$

is an isomorphism and the map

$$\Psi : H_1^{[a-\varepsilon, a+\varepsilon]}(R) \rightarrow \widehat{H}_1^{[a-\varepsilon, a+\varepsilon]}(\Sigma^{\text{lin}})$$

is surjective.

We take for ε_a the constant from Lemma 7.11 and consider $\varepsilon < \varepsilon_a$. Then we have canonical isomorphisms

$$\Gamma : H_i^{[a-\varepsilon, a+\varepsilon]}(R) \cong A_i^a, \quad i = 0, 1$$

to the groups A_i^a introduced in the previous subsection. Recall the maps $\Theta : A_i^a \rightarrow H_i^{[a-\varepsilon, a+\varepsilon]}(\Sigma^{\text{lin}}, D)$ from Corollary 7.15 which are an isomorphism for $i = 0$ and surjective for $i = 1$.

We consider first the case $i = 0$. By Proposition 6.14, for a binormal chord c of index 0 and length a the moduli space of holomorphic disks with positive puncture c and switching boundary conditions contains one component corresponding to the half-strip over c , and on all other components the Q -strings in the boundary have total length less than $a - \varepsilon$. This shows that the map $\Psi : H_0^{[a-\varepsilon, a+\varepsilon]}(R) \rightarrow H_0^{[a-\varepsilon, a+\varepsilon]}(\Sigma^{\text{lin}})$ agrees with $\Theta \circ \Gamma$ and is therefore an isomorphism.

For $i = 1$ we have a diagram

$$\begin{array}{ccc}
 H_1^{[a-\varepsilon, a+\varepsilon]}(R) & \xrightarrow{\Psi} & \widehat{H}_1^{[a-\varepsilon, a+\varepsilon]}(\Sigma^{\text{lin}}) \\
 \downarrow \Gamma & & \uparrow \Pi \\
 A_1^a & \xrightarrow{\Theta} & H_1^{[a-\varepsilon, a+\varepsilon]}(\Sigma^{\text{lin}}),
 \end{array}$$

where $\Pi : H_1(\Sigma^{\text{lin}}) = \ker D^{\text{lin}} / \text{im } D^{\text{lin}} \rightarrow \ker D^{\text{lin}} / i^{-1}(\text{im } D) = \widehat{H}_1(\Sigma^{\text{lin}})$ is the canonical projection. Since Π and Θ are surjective, surjectivity of Ψ follows once we show that the diagram commutes.

To see this, consider a word $w = b_1 \dots b_k c$ of binormal chords of indices $j b_j = 0$ and $j c_j = 1$ and total length a . The 1-dimensional moduli space of holomorphic strips with positive puncture asymptotic to c and one boundary component on the zero section contains a unique component M_c passing through the trivial strip over c . By Proposition 6.14, for each other element in M_c the boundary on the zero section has length strictly less than $L(c)$. So, for ε sufficiently small, the moduli space represents a generator of the local first homology at c . Since on all other components of the moduli space the Q -strings in the boundary have total length less than $a - \varepsilon$, the product of M_c with the half-strips over the b_j gives $\Phi(w) \in C_1^{[a-\varepsilon, a+\varepsilon]}(\Sigma)$. Its image $\Psi(w) = \Gamma^{-1} \Phi(w) \in C_1^{[a-\varepsilon, a+\varepsilon]}(\Sigma^{\text{lin}})$ is obtained from $\Phi(w)$ by shortening the Q -strings to linear ones. Since the tuples of Q -strings in $\Phi(w)$ were either C^1 -close to w (depending on ε) or had total length less than $a - \varepsilon$, the same holds for $\Psi(w)$. Hence $\Psi(w)$ is homologous (with respect to ∂ , and therefore with respect to D) in $C_1^{[a-\varepsilon, a+\varepsilon]}(\Sigma^{\text{lin}})$ to the unstable manifold of w in Σ^{lin} , which by definition equals $\Pi \circ \Theta \circ \Gamma(w)$.

In the previous argument we have ignored the N -strings, always connecting the ends of Q -strings to the base point by capping paths. More generally, a generator of $H_1^{[a-\varepsilon, a+\varepsilon]}(R) = A_1^a$ is given by a word $\gamma_1 c_1 \dots \gamma_\ell c_\ell \gamma_{\ell+1}$, where the c_j are binormal chords with one of them of index 1 and all others of index 1, and the γ_j are homotopy classes of N -strings connecting the end points and not intersecting K in the interior. Now we apply the same arguments as above to the Q -strings, dragging along the N -strings, to prove commutativity of the diagram. This concludes the proof of Proposition 7.16, and thus of Theorem 1.2.

8. PROPERTIES OF HOLOMORPHIC DISKS

In this section we begin our analysis of the holomorphic disks involved in the definition of the chain map from Legendrian contact homology to string homology. For the remainder of the paper, we consider the following setup:

- Q is a real analytic Riemannian 3-manifold without closed geodesics and convex at infinity (the main example being $Q = \mathbb{R}^3$ with the flat metric);
- $K \subset Q$ is a real analytic knot with nondegenerate binormal chords;

– $L_K \cap T^*Q$ is the conormal bundle, $Q \subset T^*Q$ is the 0-section, and

$$L = L_K \cup Q$$

is the singular Lagrangian with clean intersection $L_K \setminus Q = K$.

The reader will notice that much of the discussion naturally extends to higher dimensional manifolds Q and submanifolds $K \subset Q$.

8.1. ALMOST COMPLEX STRUCTURES. — Consider the subsets

$$S^*Q = \{(q, p) \mid |p| = 1\} \subset T^*Q \quad D^*Q = \{(q, p) \mid |p| < 1\} \subset T^*Q$$

of the cotangent bundle. The canonical isomorphism

$$R : S^*Q \rightarrow T^*Q \cap Q, \quad (s, (q, p)) \mapsto (q, e^s p)$$

intertwines the R -actions given by translation resp. rescaling. Let $\lambda = p dq$ be the canonical Liouville form on T^*Q with Liouville vector field $p\partial_p$. Its restriction λ_1 to S^*Q is a contact form with contact structure $\xi = \ker \lambda_1$ and Reeb vector field R . We denote the R -invariant extensions of λ_1, ξ, R to $T^*Q \cap Q$ by the same letters. In geodesic normal coordinates q_i and dual coordinates p_i they are given by

$$\lambda_1 = \frac{p dq}{|p|}, \quad R = \sum_i p_i \frac{\partial}{\partial q_i}, \quad \xi_{(q,p)} = \ker \lambda_1 \setminus \ker(p dp) = \text{span} \left\{ R, p \frac{\partial}{\partial p} \right\}^{\perp_{d\lambda_1}}.$$

Around each Reeb chord $c : [0, T] \rightarrow S^*Q$ with end points on $\Lambda_K = L_K \setminus S^*Q$ we pick a neighborhood $U \subset (\varepsilon, T + \varepsilon) \subset S^*Q$, where U is a neighborhood of the origin in \mathbb{C}^2 , with the following properties:

- the Reeb chord c corresponds to $f \circ g : [0, T]$;
- the Reeb vector field R is parallel to ∂_t , where t is the coordinate on $(\varepsilon, T + \varepsilon)$ and the contact planes project isomorphically onto U along R ;
- along $f \circ g : (\varepsilon, T + \varepsilon)$ the contact planes agree with $\mathbb{C}^2 \setminus f \circ g$ and the form $d\lambda_1$ with $\omega_{\text{st}} = dx_1 \wedge dy_1 + dx_2 \wedge dy_2$;
- the Legendrian Λ_K intersects $U \subset (\varepsilon, T + \varepsilon)$ in two linear subspaces contained in $U \setminus f \circ g$ and $U \setminus f T g$, respectively, whose projections to U are transversely intersecting Lagrangian subspaces of $(\mathbb{C}^2, \omega_{\text{st}})$.

DEFINITION 8.1. — An almost complex structure J on T^*Q is called *admissible* if it has the following properties.

(i) J is everywhere compatible with the symplectic form $dp \wedge dq$. Moreover, Q admits an exhaustion $Q_1 \subset Q_2 \subset \dots$ by compact sets with smooth boundary such that the pullbacks $\pi^{-1}(\partial Q_i)$ under the projection $\pi : T^*Q \rightarrow Q$ are J -convex hypersurfaces.

(ii) Outside D^*Q , J agrees with an R -invariant almost complex structure J_1 on the symplectization that takes the Liouville field $p\partial_p$ to the Reeb vector field R , restricts to a complex structure on the contact distribution ξ , and is compatible with the symplectic form $d\lambda_1$ on ξ .

(iii) Outside the zero section, J preserves the subspace $\text{span}\{p\partial_p, Rg\}$ as well as ξ and is compatible with the symplectic form $d\lambda_1$ on ξ . Along the zero section, J agrees with the canonical structure $\partial/\partial p_i \nabla \partial/\partial q_i$.

(iv) J is integrable near K such that Q and K are real analytic.

(v) On each neighborhood $U = (\varepsilon, T + \varepsilon)$ around a Reeb chord as above, the restriction of J_1 to the contact planes is the pullback of the standard complex structure on $U = \mathbb{C}^2$ under the projection.

REMARK 8.2. — Conditions (i) and (ii) are standard conditions for studying holomorphic curves in TQ and its symplectization $\mathbb{R} \times S^1 Q$. Condition (iii) ensures the crucial length estimate for holomorphic curves in the next subsection. Condition (iv) is needed for the Finiteness Theorem 6.5 to hold. Condition (v) is added to facilitate our study of spaces of holomorphic disks and is convenient for fixing gauge when finding smooth structures on moduli spaces; it can probably be removed with a more involved analysis of asymptotics.

REMARK 8.3. — Note that an admissible almost complex structure remains so under arbitrary deformations satisfying (ii) that are supported outside DQ and away from the Reeb chords. This gives us enough freedom to achieve transversality within the class of admissible structures in Section 9.

The Riemannian metric on Q induces a canonical almost complex structure J_{st} on TQ which in geodesic normal coordinates q_i at a point q and dual coordinates p_i is given by

$$J_{\text{st}} \left(\frac{\partial}{\partial q_i} \right) = \frac{\partial}{\partial p_i}, \quad J_{\text{st}} \left(\frac{\partial}{\partial p_i} \right) = \frac{\partial}{\partial q_i}.$$

More generally, for a positive smooth function $\rho: [0, 1) \rightarrow (0, 1)$ we define an almost complex structure J_ρ by

$$J_\rho \left(\frac{\partial}{\partial q_i} \right) = \rho(j\rho) \frac{\partial}{\partial p_i}, \quad J_\rho \left(\frac{\partial}{\partial p_i} \right) = \rho(j\rho)^{-1} \frac{\partial}{\partial q_i}.$$

If $\rho(r) = r$ for large r , then it is easy to check that J_ρ satisfies the first part of condition (i) as well as conditions (ii) and (iii) in Definition 8.1. If the metric is flat (i.e., Q is \mathbb{R}^3 or a quotient of \mathbb{R}^3 by a lattice), then J_{st} is integrable and J_ρ also satisfies the second part of (i) (choosing Q_i to be round balls) and condition (iv). Condition (v) can then be arranged by deforming J_ρ near infinity within the class of almost complex structures satisfying (ii). So we have shown the following.

LEMMA 8.4. — *For $Q = \mathbb{R}^3$ with the Euclidean metric there exist admissible almost complex structures in the sense of Definition 8.1.*

REMARK 8.5. — In fact, the almost complex structure J_{st} induced by the metric is integrable if and only if the metric is flat (this observation is due to M. Grüneberg, unpublished). So the preceding proof of Lemma 8.4 does not carry over to general manifolds Q (although the conclusion should still hold).

The next result provides nice holomorphic coordinates near $K \subset TQ$.

LEMMA 8.6. — *Suppose that J satisfies condition (iv) in Definition 8.1. Then for $\delta > 0$ small enough there exists a holomorphic embedding from $S^1 \times (\delta, \delta) \rightarrow B_\delta^4$, where $B_\delta^4 \subset \mathbb{C}^2$ is the ball of radius δ , with its standard complex structure onto a neighborhood of K in TQ with complex structure J with the following properties:*

- $S^1 \times \{0\} \rightarrow TQ$ maps onto K ;
- $S^1 \times (\mathbb{R}^2 \setminus B_\delta^4)$ maps to Q ;
- $S^1 \times (i\mathbb{R}^2 \setminus B_\delta^4)$ maps to L_K .

Alternatively, we can arrange the last two properties with the roles of Q and L_K interchanged.

Proof. — This is proved in more generality in [5, Rem. 3.2]; for convenience we repeat the proof in the situation at hand. Consider the real analytic embedding $\gamma: S^1 \rightarrow Q$ representing K . Pick a real analytic vector field v on Q which is nowhere tangent to K along K . Let v_1 be the unit vector field along K in the direction of the component of v perpendicular to $\dot{\gamma}$. Then v_1 is a real analytic vector field along K . Let $v_2 = \dot{\gamma} \times v_1$ be the unit vector field along K which is perpendicular to both $\dot{\gamma}$ and v_1 and which is such that $(\dot{\gamma}, v_1, v_2)$ is a positively oriented basis of TQ . Consider $S^1 \times D^2$ with coordinates (s, σ_1, σ_2) , $s \in \mathbb{R}/\mathbb{Z}$, $\sigma_j \in \mathbb{R}$. Since K is an embedding there exists $\rho > 0$ such that

$$(8.1) \quad \phi(s, \sigma_1, \sigma_2) = \gamma(s) + \sigma_1 v_1(s) + \sigma_2 v_2(s)$$

is an embedding for $\sigma_1^2 + \sigma_2^2 < \rho$. Note that the embedding is real analytic. Equip $S^1 \times D^2$ with the flat metric and consider the induced complex structure on $T(S^1 \times D^2)$. The real analyticity of ϕ in (8.1) implies that it extends to holomorphic embedding Φ from a neighborhood of $S^1 \times D^2$ in $T(S^1 \times D^2)$ to a neighborhood of K in TQ (here we use integrability of J near K). In fact, locally Φ is obtained by replacing the real variables (s, σ_1, σ_2) in the power series corresponding in the right hand side of (8.1) by their complexifications $(s + it, \sigma_1 + i\tau_1, \sigma_2 + i\tau_2)$. This proves the first assertion of the lemma. The alternative assertion follows from this one by precomposing Φ with multiplication by i on B_δ^4 .

REMARK 8.7. — The coordinate system gives a framing of K determined by the normal vector field v . By real analytic approximation we can take v to represent any class of framings.

8.2. LENGTH ESTIMATES. — In this subsection we show that the chain map Φ respects the length filtrations. This was shown in [6] for the absolute case, i.e., without the additional boundary condition L_K , and the arguments carry over immediately to the relative case. For completeness, we provide the proof in this subsection and we keep the level of generality of [6], which is slightly more than what we use in this paper.

For preparation, consider a smooth function $\tau: [0, 1) \rightarrow [0, 1)$ with $\tau^0(s) > 0$ everywhere and $\tau(s) = 0$ near $s = 0$. Then

$$\lambda_\tau := \frac{\tau(jpj)p dq}{j pj}$$

defines a smooth 1-form on TQ .

LEMMA 8.8. — *Let J be an admissible almost complex structure on TQ and τ a function as above. Then for all $v \in T_{(q,p)}TQ$ we have*

$$d\lambda_\tau(v, Jv) > 0.$$

At points where $\tau(jpj) > 0$ and $\tau^0(jpj) > 0$ equality holds only for $v = 0$, whereas at points where $\tau(jpj) > 0$ and $\tau^0(jpj) = 0$ equality holds if and only if v is a linear combination of the Liouville field $p\partial_p$ and the Reeb vector field $R = p\partial_q$.

Proof. — By condition (iii) in Definition 8.1, J preserves the splitting

$$T(TQ) = \text{span}\{\bar{p}\partial_p, R\} \oplus \xi$$

and is compatible with $d\lambda_1$ on ξ . Let us denote by $\pi_1: T(TQ) \rightarrow \text{span}\{\bar{p}\partial_p, R\}$ and $\pi_2: T(TQ) \rightarrow \xi$ the projections onto the direct summands. Since $\ker(d\lambda_1) = \text{span}\{\bar{p}\partial_p, R\}$, for $v \in T_{(q,p)}TQ$ we conclude

$$d\lambda_1(v, Jv) = d\lambda_1(\pi_2v, J\pi_2v) > 0,$$

with equality if $v \in \text{span}\{\bar{p}\partial_p, R\}$. Next, we consider

$$d\lambda_\tau = \tau(jpj)d\lambda_1 + \frac{\tau^0(jpj)}{j pj} p dp \wedge \lambda_1.$$

Since the form $p dp \wedge \lambda_1$ vanishes on ξ and is positive on $\text{span}\{\bar{p}\partial_p, R\}$, we conclude

$$d\lambda_\tau(v, Jv) = \tau(jpj)d\lambda_1(\pi_2v, J\pi_2v) + \frac{\tau^0(jpj)}{j pj} p dp \wedge \lambda_1(\pi_1v, J\pi_1v) > 0,$$

with equality if both summands vanish. From this the lemma follows.

Let now J be an admissible almost complex structure on TQ and

$$u: (\Sigma, \partial\Sigma) \rightarrow (TQ, Q \cup L_K)$$

be a J -holomorphic curve with finitely many positive boundary punctures asymptotic to Reeb chords a_1, \dots, a_s and with switching boundary conditions on $Q \cup L_K$. Let $\sigma_1, \dots, \sigma_k$ be the boundary segments on Q . Recall that $L(\sigma_i)$ denotes the Riemannian length of σ_i and $L(a_j) = \int_{a_j} \lambda_1$ denotes the action of the Reeb chord a_j , which agrees with the length of the corresponding binormal chord.

PROPOSITION 8.9. — *With notation as above we have*

$$\sum_{i=1}^k L(\sigma_i) \leq \sum_{j=1}^s L(a_j),$$

and equality holds if and only if u is a branched covering of a half-strip over a binormal chord.

Proof. — The idea of the proof is straightforward: integrate $u \, d\lambda_1$ over Σ and apply Stokes' theorem. However, some care is required to make this rigorous because the 1-form λ_1 is singular along the zero section.

Fix a small $\delta > 0$. For $i = 1, \dots, s$ pick biholomorphic maps $\phi_i : [0, \delta] \rightarrow [0, 1] \setminus \{1\}$ $N_i \subset \Sigma$ onto neighborhoods N_i in Σ of the i^{th} boundary segment mapped to Q , so that $\phi_i(0, t)$ is a parametrization of the i^{th} boundary segment. We choose δ so small that $N_i \cap N_j = \emptyset$ if $i \neq j$ and $u \circ \phi_i(\delta, \cdot)$ does not hit the zero section (the latter is possible because otherwise by unique continuation u would be entirely contained in the zero section, which it is not by assumption). For fixed i we denote the induced parametrization of σ_i by $q(t) := u \circ \phi_i(t) \in Q$, so we can write

$$u \circ \phi_i(s, t) = (q(t) + v(s, t), s\dot{q}(t) + w(s, t))$$

with $v(0, t) = 0 = w(0, t)$, and therefore $\partial v / \partial t(0, t) = 0 = \partial w / \partial t(0, t)$. The hypothesis that J is standard near the zero section (condition (iii) in Definition 8.1) implies that $\partial v / \partial s(0, t) = 0 = \partial w / \partial s(0, t)$. Denoting $v_\delta = v(\delta, \cdot)$ and $w_\delta = w(\delta, \cdot)$ we compute

$$\begin{aligned} (u \circ \phi_i) \lambda_1|_{s=\delta} &= \frac{h\delta\dot{q} + w_\delta, \dot{q} + v_\delta}{j\delta\dot{q} + w_\delta} dt \\ &= \frac{h\dot{q} + w_\delta/\delta, \dot{q} + v_\delta}{j\dot{q} + w_\delta/\delta} dt \\ &= (j\dot{q} + O(\delta)) dt, \end{aligned}$$

where in the last line we have used that $v_\delta = O(\delta)$ and $w_\delta = O(\delta^2)$.

Pick $\varepsilon > 0$ smaller than the minimal norm of the p -components of $u \circ \phi_i(\delta, \cdot)$ for all i . Pick a function $\tau : [0, 1] \setminus \{1\} \rightarrow [0, 1]$ with $\tau^\theta > 0$, $\tau(s) = 0$ near $s = 0$, and $\tau(s) = 1$ for $s > \varepsilon$. By Lemma 8.8, the form $\lambda_\tau = (\tau(jp)/jp) \, pdq$ on $T \setminus Q$ satisfies $u \circ (d\lambda_\tau) > 0$. Note that λ_τ agrees with $\lambda_1 = (p/jp) \, dq$ on the subset $\{jpj > \varepsilon g\} \subset T \setminus Q$, so the preceding computation yields

$$\int_{\widehat{r}_s = \delta g} (u \circ \phi_i) \lambda_\tau = \int_{\widehat{r}_s = \delta g} (j\dot{q} + O(\delta)) dt = L(\sigma_i) + O(\delta)$$

for all i . Next, consider polar coordinates (r, φ) around 0 in the upper half plane H^+ near the j^{th} positive puncture. Then the asymptotic behavior of u near the punctures yields

$$\int_{\widehat{r} = \delta g \setminus H^+} u \lambda_\tau = L(a_j) + O(\delta).$$

Now let $\Sigma_\delta \subset \Sigma$ be the surface obtained by removing the neighborhoods $\widehat{r} \in \delta g \setminus H^+$ around the positive punctures and the neighborhoods N_i of the boundary segments mapped to Q , see Figure 8.1.

The boundary of Σ_δ consists of the arcs $\widehat{r} = \delta g \setminus H^+$ around the positive punctures, the arcs $\phi_i(\widehat{r} = \delta g)$ near the boundary segments mapped to Q (negatively oriented), and the remaining parts of $\partial\Sigma$ mapped to L_K . Since λ_τ vanishes on L_K , the latter boundary parts do not contribute to its integral and Stokes' theorem combined with

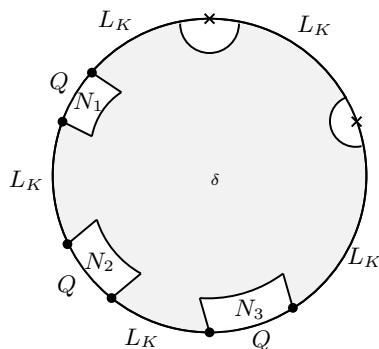


FIGURE 8.1. The domain Σ_δ is obtained from Σ by removing small neighborhoods of the boundary arcs mapping to Q and of the positive punctures. The punctures are denoted by x , and switches are denoted by dots.

the preceding observations yields

$$0 \leq \int_\delta u \, d\lambda_\tau = \int_\delta u \, \lambda_\tau = \sum_{j=1}^s L(a_j) \sum_{i=1}^k L(\sigma_i) + O(\delta).$$

Taking $\delta \neq 0$ this proves the inequality in Proposition 8.9. Equality holds if $u \, d\lambda_\tau$ vanishes identically, which by Lemma 8.8 is the case if u is everywhere tangent to $\text{span}\{\tilde{f}_p, \partial_p\}$, Rg . In view of the asymptotics at the positive punctures, this is the case precisely for a half-strip over a binormal chord.

8.3. HOLOMORPHIC HALF-STRIPS. — We consider the half-strip $\mathbb{R}_+ \times [0, 1]$ with coordinates (s, t) and its standard complex structure. Let J be an admissible almost complex structure on $T \times Q$ and J_1 the associated structure on $\mathbb{R} \times S \times Q$. A *holomorphic half-strip* in $\mathbb{R} \times S \times Q$ is a holomorphic map

$$u: \mathbb{R}_+ \times [0, 1] \rightarrow (\mathbb{R} \times S \times Q, J_1)$$

mapping the boundary segments $\mathbb{R} \times \{0\} \times g$ and $\mathbb{R} \times \{1\} \times g$ to $\mathbb{R} \times \Lambda_K$. Similarly, a holomorphic half-strip in $T \times Q$ is a holomorphic map

$$u: \mathbb{R}_+ \times [0, 1] \rightarrow (T \times Q, J)$$

mapping the boundary to $L = L_K \cup Q$. We write the components of a map u into $\mathbb{R} \times S \times Q$ (or into $T \times Q \cap D \times Q = \mathbb{R}_+ \times S \times Q$) as

$$u = (a, f).$$

Recall from [3] (see also [5]) that to any smooth map u from a surface to $\mathbb{R} \times S \times Q$ or $T \times Q$ we can associate its *Hofer energy* $E(u)$. It is defined as the sum of two terms, the ω -energy and the λ -energy, whose precise definition will not be needed here. The following result follows from [10, Lem. B.1], see also [3, Prop. 6.2], in combination with well-known results in Lagrangian Floer theory, see e.g. [21].

PROPOSITION 8.10. — *For each holomorphic half-strip u in $\mathbb{R} \times S \times Q$ or $T \times Q$ of finite Hofer energy exactly one of the following holds:*

- *There exists a Reeb chord $c: [0, T] \rightarrow S \times Q$ and a constant $a_0 \in \mathbb{R}$ such that*

$$a(s, t) - Ts - a_0 \neq 0, \quad f(s, t) \neq c(Tt)$$

uniformly in t as $s \rightarrow 1^-$. We say that the map has a positive puncture at c .

- *There exists a Reeb chord $c: [0, T] \rightarrow S \times Q$ and a constant $a_0 \in \mathbb{R}$ such that*

$$a(s, t) + Ts - a_0 \neq 0, \quad f(s, t) \neq c(-Tt)$$

uniformly in t as $s \rightarrow 1^-$. We say that the map has a negative puncture at c .

- *There exists a point x_0 on $\mathbb{R} \times \Lambda_K$ (resp. L) such that*

$$u(s, t) \rightarrow x_0$$

uniformly in t as $s \rightarrow 1^-$. In this case $u = \chi^{-1}$, where $\chi: \mathbb{R}_+ \times [0, 1] \rightarrow D^+$ is the map from (6.3), extends to a holomorphic map on the half-disk mapping the boundary to $\mathbb{R} \times \Lambda_K$ (resp. L). If $x_0 \notin K$ then we say that u has a removable puncture at x_0 , and if $x_0 \in K$ then we say that u has a Lagrangian intersection puncture at x_0 . (These are the standard situations in ordinary Lagrangian intersection Floer homology.)

Because of our choice of almost complex structure we can say more about the local forms of the maps as follows.

Consider first a Reeb chord puncture where the map approaches a Reeb chord c . Let $U = (\varepsilon, T + \varepsilon)$ be the neighborhood of c as in Definition 8.1 (v) and note that the holomorphic half-strip is uniquely determined by the local projection to $U \subset \mathbb{C}^2$ where the complex structure is standard. By a complex linear change of coordinates on \mathbb{C}^2 we can arrange that the two branches of the Legendrian Λ_K through the end points of c project to \mathbb{R}^2 and to the subspace spanned by the vectors $(e^{i\theta_1}, 0)$ and $(0, e^{i\theta_2})$, for some angles θ_1, θ_2 . The \mathbb{C}^2 -component v of the map u then has a Fourier expansion

$$(8.2) \quad v(z) = \sum_{n>0} (c_{1;n} e^{(\theta_1+n)z}, c_{2;n} e^{(\theta_2+n)z}),$$

where $c_{j;n}$ are real numbers. We call the smallest n such that $(c_{1;n}, c_{2;n}) \neq 0$ the *order of convergence* to the Reeb chord c .

We have similar expansions near the Lagrangian intersection punctures. Lemma 8.6 gives holomorphic coordinates $(z_0, z_1) = (x_0 + iy_0, x_1 + iy_1)$ in $\mathbb{C} \times \mathbb{C}^2$ around any point $q_0 \in K$ such that the Lagrangian submanifold $Q = T \times Q$ corresponds to $\hat{r}y_0 = y_1 = 0g$, the Lagrangian submanifold L_K corresponds to $\hat{r}y_0 = x_1 = 0g$, and the almost complex structure J corresponds to the standard complex structure i on \mathbb{C}^3 . Consider a holomorphic map $u: [0, 1) \rightarrow [0, 1] \times T \times Q$ such that $u(z) \rightarrow q \in K$ as $z \rightarrow 1^-$ where q lies in a small neighborhood of q_0 in K . We write u in the local coordinates described above as $v = (v_0, v_1)$. Now Remark 4.2 yields the following Fourier expansions for v .

If $v([0, 1) \setminus f_0g) \subset Q$ and $v([0, 1) \setminus f_1g) \subset L_K$ then

$$(8.3) \quad v(z) = \left(\sum_{m>0} c_{0,m} e^{-m\pi z}, \sum_{n+1/2>0} c_{1;n+1/2} e^{-(n+1/2)\pi z} \right),$$

where $c_{0,m} \in \mathbb{R}$ for all $m \in \mathbb{Z}_{>0}$ and where $c_{1;n+1/2} \in \mathbb{R}^2$ for all $n \in \mathbb{Z}_{>0}$, in a neighborhood of 1 . If $v([0, 1) \setminus f_0g) \subset L_K$ and $v([0, 1) \setminus f_1g) \subset Q$ then

$$(8.4) \quad v(z) = \left(\sum_{m>0} c_{0,m} e^{-m\pi z}, i \sum_{n+1/2>0} c_{1;n+1/2} e^{-(n+1/2)\pi z} \right),$$

where notation is as in (8.3). If $v([0, 1) \setminus f_0g) \subset Q$ and $v([0, 1) \setminus f_1g) \subset Q$ then

$$(8.5) \quad v(z) = \left(\sum_{n>0} c_{0,m} e^{-m\pi z}, \sum_{n>0} c_{1;n} e^{-n\pi z} \right),$$

where $c_{0,m}$ is as in (8.3) and $c_{1;n} \in \mathbb{R}^2$ all $n \in \mathbb{Z}_{>0}$. If $v([0, 1) \setminus f_0g) \subset L_K$ and $v([0, 1) \setminus f_1g) \subset L_K$ then

$$(8.6) \quad v(z) = \left(\sum_{n>0} c_{0,m} e^{-m\pi z}, i \sum_{n>0} c_{1;n} e^{-n\pi z} \right),$$

where notation is as in (8.5). We say that the smallest half-integer $n + 1/2$ in (8.3) or (8.4) such that $c_{1;n+1/2} \neq 0$ or the smallest integer n in (8.5) or (8.6) such that $c_{1;n} \neq 0$ is the *asymptotic winding number* of u at its Lagrangian intersection puncture.

8.4. HOLOMORPHIC DISKS. — Consider the closed unit disk $D \subset \mathbb{C}$ with $m+1$ cyclically ordered distinct points z_0, \dots, z_m on ∂D . Set $\dot{D} := D \setminus \{z_0, \dots, z_m\}$. Consider a J -holomorphic map $u: \dot{D} \rightarrow (S^1 \times Q)$ resp. (T^*Q) which maps $\partial D \setminus \{z_0, \dots, z_m\}$ to $\mathbb{R} \times \Lambda_K$ resp. $L = Q \cap L_K$ and which has finite ω -energy and λ -energy. Proposition 8.10 shows that near each puncture z_j the map u either extends continuously, or it is positively or negatively asymptotic to a Reeb chord. We will use the following notation for such disks.

A *symplectization disk (with $m > 0$ negative punctures)* is a J -holomorphic map

$$u: (\dot{D}, \partial\dot{D}) \rightarrow (\mathbb{R} \times S^1 \times Q, \mathbb{R} \times \Lambda_K)$$

with positive puncture at z_0 and negative punctures at z_1, \dots, z_m . A *cobordism disk (with $m > 0$ Lagrangian intersection punctures)* is a J -holomorphic map

$$u: (\dot{D}, \partial\dot{D}) \rightarrow (T^*Q, L)$$

with positive puncture at z_0 and Lagrangian intersection punctures at z_1, \dots, z_m .

Let $\mathbf{b} = b_1 b_2 \dots b_m$ be a word of m Reeb chords. We write

$$\mathcal{M}^{\text{sy}}(a, n_0; b_1, \dots, b_m) = \mathcal{M}^{\text{sy}}(a, n_0; \mathbf{b})$$

for the moduli space of symplectization disks with positive puncture asymptotic to the Reeb chord a where the order of convergence is n_0 and m negative punctures (in counterclockwise order) asymptotic to the Reeb chords b_1, \dots, b_m . Here the points z_0, \dots, z_m on ∂D are allowed to vary and we divide by the action of Möbius transformations on D . Note that \mathbb{R} acts by translation on these moduli spaces.

Similarly, let $\mathbf{n} = (n_1, \dots, n_m)$ be a vector of half-integers or integers. We write

$$M(a, n_0; n_1, \dots, n_m) = M(a, n_0; \mathbf{n})$$

for the moduli space of cobordism disks with positive puncture asymptotic to the Reeb chord a with degree of convergence n_0 and $m > 0$ Lagrangian intersection punctures with asymptotic winding numbers given by the integers or half-integers n_j . Note that the number of half-integers must be even for topological reasons (at each half-integer the boundary of u switches from Q to L_K or vice versa).

In both cases when $n_0 = 0$ we will suppress it from notation and simply write

$$M^{\text{sy}}(a; \mathbf{b}) \text{ and } M(a; \mathbf{n}),$$

respectively.

For a Reeb chord $c: [0, T] \rightarrow S^1 \times Q$ of length T , the map $u_c: \mathbb{R} \rightarrow [0, 1] \times \mathbb{R} \rightarrow S^1 \times Q$ given by $u_c(s + it) = (Ts, c(Tt))$ is a J -holomorphic parametrization of $\mathbb{R} \rightarrow c$ and thus a symplectization disk with positive and negative puncture asymptotic to c . We call it the *Reeb chord strip* over c .

8.5. COMPACTNESS IN $\mathbb{R} \rightarrow S^1 \times Q$ AND $T \times Q$. — In this subsection we review the compactness results proved in [5] that concern compactness of the moduli spaces of holomorphic disks discussed in Section 8.4.

Let us denote by a *source disk* D_m the unit disk with some number $m + 1 > 1$ of punctures z_0, \dots, z_m on its boundary; we call z_0 the positive and z_1, \dots, z_m the negative punctures. A *broken source disk* \dot{D}_m with $r > 1$ levels with $m + 1$ boundary punctures is represented as a finite disjoint union of punctured disks,

$$\dot{D}_m = D^{1,1} \sqcup (D^{2,1} \sqcup \dots \sqcup D^{2,\ell_2}) \sqcup \dots \sqcup (D^{r,1} \sqcup \dots \sqcup D^{r,\ell_r}),$$

where $(D^{j,1} \sqcup \dots \sqcup D^{j,\ell_j})$ are the disks in the j^{th} level and we require the following properties:

- Each negative puncture q of a disk $D^{j,k}$ in the j^{th} level for $j < r$ is formally joined to the positive puncture of a unique disk $D^{j+1,s}$ in the $(j + 1)^{\text{th}}$ level. We say that $D^{j+1,s}$ is attached to $D^{j,k}$ at the negative puncture q .

- The total number of negative punctures on level r is m .

Note that a broken source disk with one level is just a source disk.

We consider first compactness for curves in the symplectization. Let \dot{D}_m be a broken source disk as above. A *broken symplectization disk with r levels* with domain \dot{D}_m is a collection \dot{v} of J -holomorphic maps $v^{j,k}$ defined on $D^{j,k}$ with the following properties:

- For each $1 \leq j \leq r$ and $1 \leq k \leq \ell_j$, $v^{j,k}$ represents an element in

$$M^{\text{sy}}(a^{j,k}; b_1^{j,k}, \dots, b_s^{j,k}).$$

Moreover, for $j > 1$, the Reeb chord $a^{j,k}$ at the positive puncture of $v^{j,k}$ matches the Reeb chord b^{j-1,k^0} at the negative puncture of v^{j-1,k^0} in D^{j-1,k^0} at which $D^{j,k}$ is attached.

- For each level $1 \leq j \leq r$, at least one of the maps $v^{j,k}$ is not a Reeb chord strip.

An *arc* in a source disk is an embedded curve that intersects the boundary only at its end points and away from the punctures. We say that a sequence of symplectization disks

$$f u_j g \in M^{\text{SY}}(a; b_1, \dots, b_m)$$

converges to a broken symplectization disk if there are disjoint arcs $\gamma_1, \dots, \gamma_k$ in the domains of u_j which give the decomposition of the domain into a broken source disk in the limit and such that in the complement of these arcs, the maps u_j converge to the corresponding map of the broken disk uniformly on compact subsets.

THEOREM 8.11. — Any sequence $f u_j g \in M^{\text{SY}}(a, b_1, \dots, b_m)$ of symplectization disks has a subsequence which converges to a broken symplectization disk \dot{v} with $r > 1$ levels.

Proof. — Follows from [3] (see also [5, Th. 1.1]).

In order to describe the compactness result for moduli spaces of holomorphic disks in $T \times Q$ we first introduce a class of constant holomorphic disks and then the notion of convergence to a constant disk. A *constant holomorphic disk* is a source disk \mathbf{D}_m , $m > 3$, a constant map into a point $q \in K$, and the following extra structure: Each boundary component is labeled by L_K or by Q and at each puncture z_j there is an asymptotic winding number $n_j \in \mathbb{Z} \setminus \{\frac{1}{2}, 1, \frac{3}{2}, \dots\}$ such that n_j is a half-integer if the adjacent boundary components of \mathbf{D}_m are labeled by different components of $L = L_K \sqcup Q$ and an integer otherwise, and such that $n_0 = \sum_{j=1}^m n_j$.

A sequence of holomorphic maps $v_j: \mathbf{D}_m \rightarrow T \times Q$ with boundary on L converges to a constant holomorphic disk if it converges uniformly to the constant map on any compact subset and if for all sufficiently large j , v_j takes any boundary component labeled by L_K or Q to L_K or Q , respectively, and if the asymptotic winding numbers at the negative punctures of the maps v_j agree with those of the constant limit map at corresponding punctures.

Let $\dot{\mathbf{D}}_m$ be a broken source disk with r levels and suppose $1 \leq r_0 \leq r$. A *broken cobordism disk with r_0 non-constant levels* and domain $\dot{\mathbf{D}}_m$ is a collection \dot{v} of J -holomorphic maps $v^{j,k}$ defined on $\mathbf{D}^{j,k}$ with the following properties.

- For $j < r_0$ and $1 \leq k \leq \ell_j$, $v^{j,k}$ represents an element in

$$M^{\text{SY}}(a^{j,k}; b_1^{j,k}, \dots, b_s^{j,k}).$$

Moreover, for $j > 1$, the Reeb chord $a^{j,k}$ at the positive puncture of $v^{j,k}$ matches the Reeb chord b^{j-1,k^0} at the negative puncture of v^{j-1,k^0} in \mathbf{D}^{j-1,k^0} at which $\mathbf{D}^{j,k}$ is attached.

- For each level $j < r_0$, at least one of the maps $v^{j,k}$ is not a Reeb chord strip.
- For $j = r_0$ and $1 \leq k \leq \ell_j$, $v^{j,k}$ represents an element in

$$M(a^{j,k}; n_1^{j,k}, \dots, n_s^{j,k})$$

and the Reeb chord at the positive puncture of $v^{j,k}$ matches the Reeb chord at the negative puncture of v^{j-1,k^0} in \mathbf{D}^{j-1,k^0} at which $\mathbf{D}^{j,k}$ is attached.

– For $j > r_0$, $v^{j,k}$ is a constant map to $q \in K$, where $q \in K$ is the image of the negative puncture of v^{j-1,k^0} in D^{j-1,k^0} , at which $D^{j,k}$ is attached. Moreover, $D^{j,k}$ has at least 3 punctures and the winding number and labels at its positive puncture agree with those of the negative puncture where it is attached. (From the point of view of the source disk these constant levels encode degenerations of the conformal structure corresponding to colliding Lagrangian intersection punctures, see Section 10.3 for more details.)

We say that the disks in levels $j < r_0$ are the *symplectization disks*, that the disks in level r_0 are the *cobordism disks*, and that disks in levels $j > r_0$ are the *constant disks* of the broken disk.

We define convergence to a broken cobordism disk completely parallel to the symplectization case.

THEOREM 8.12. — *Let $\{u_j\} \subset M(a; n_1, \dots, n_m)$ be a sequence of cobordism disks. Then $\{u_j\}$ has a subsequence which converges to a broken cobordism disk.*

Proof. — This is a consequence of [5, Th. 1.1]. Note that the levels of constant disks are recovered by the sequence of source disks that converges to a broken source disk.

REMARK 8.13. — We consider the convergence implied by the Compactness Theorem 8.12 in more detail in a special case relevant to the description of our moduli spaces below. Consider a sequence of holomorphic disks u_j as in the theorem that converges to a broken cobordism disk with top level v and such that all disks on lower levels are constant. Let q_ℓ be a negative puncture of the top level v and let D_ℓ be the (possibly broken) constant disk attached with its positive puncture at q_ℓ .

Consider the sequence of domains of u_j as a sequence of strips with slits S_j , see the discussion of standard domains in Section 9.1 and Figure 9.1. It follows from the proof of [5, Th. 1.1] that there is a strip region $[\rho_j, 0] \subset [0, 1] \setminus S_j$, where $\rho_j \rightarrow 1$ as $j \rightarrow \infty$ such that in the limit the negative puncture q_ℓ of v corresponds to $(1, 0) \in [0, 1]$ and the positive puncture of the domain D_ℓ corresponds to $[0, 1) \subset [0, 1]$ attached at this puncture. Assume that q_ℓ maps to $x \in K$ and consider the Fourier expansion of v near q_ℓ in the local coordinates near K perpendicular to the knot:

$$v(s + it) = e^{k_0\pi(s+it)} \sum_{k=0}^{\infty} c_k e^{k\pi(s+it)},$$

where $k_0 > 1/2$ is a half-integer and c_k are vectors in \mathbb{R}^2 or $i\mathbb{R}^2$, $c_0 \neq 0$. We say that the complex line spanned by c_0 is the limiting tangent plane of v at q_ℓ . Writing v using Taylor expansion as a map from the upper half plane with the puncture q_ℓ at the origin and taking the complex line of c_0 as the first coordinate we find that the normal component of v at x is given by

$$v(z) = (z^{k_0}, O(z^{k_0+1})),$$

after suitable rescaling of the first coordinate.

We next restrict to the case relevant to our applications, of a sequence of disks u_j with a constant disk with three or four punctures splitting o . The three punctured disk is simpler, so we consider the case of a disk with four punctures splitting o . In this case, consider a vertical segment $f\rho^0g$ $[0, 1]$ in the stretching strip $[\rho_j, 0]$ $[0, 1]$. It subdivides the domain of u_j in two components D_+ containing the positive puncture and its complement D_- . Consider the Fourier expansion of u_j near this vertical segment. We have

$$u_j(s + it) = \sum_{k > k_0} c_{j;k} e^{k\pi(s+it)},$$

where k are half-integers and $c_{j;k} \in \mathbb{R}^2$ (or $i\mathbb{R}^2$). Since the winding number along the vertical segment is equal to the sum of the winding numbers of the negative punctures in the component of D_- that it bounds, we find that, for j sufficiently large, $c_{j;k} = 0$ for all $k < 3/2$, hence $k_0 > 3/2$. Moreover, $c_{j;k_0}$ converges to a vector in the limiting tangent plane of u_0 at the newborn negative puncture. In the generic case, see Lemma 9.5, this limiting vector is non-zero. We assume for definiteness in what follows that it is equal to $(1, 0)$.

Pick a conformal map taking D to the half disk of radius 1 in the upper half plane, with the vertical segment corresponding to the half circular arc and with the middle boundary puncture mapping to 0. Then as $j \rightarrow \infty$ the locations of the other two punctures both converge to 0 and, for large j , the projection to the first complex coordinate determines the location of the other two punctures. Moreover, the sum of the winding numbers at these three punctures equals $\frac{3}{2}$ (i.e., the winding number along the half circle of radius 1). Consequently, we have, with z a coordinate on the upper half plane, for all j large enough

$$u_j(z) = \sqrt{z(z - \delta_j)(z - \varepsilon_j)} ((1, 0) + v_j + O(z)),$$

where $\delta_j, \varepsilon_j \rightarrow 0$ and $v_j \rightarrow 0 \in \mathbb{R}^2$ as $j \rightarrow \infty$. It follows that disks in a limiting sequence eventually lie close to the model disk (4.2) discussed in Section 4.3.

There is a completely analogous and simpler analysis of the case when two punctures collide which shows that disks in a limiting sequence are close to the model disk (4.1) of Section 4.3 in the same sense.

9. TRANSVERSELY CUT OUT SOLUTIONS AND ORIENTATIONS

In this section we show that the moduli spaces in Section 8 are manifolds for generic almost complex structure J . To accomplish this, we first express each moduli space as the zero locus of a section of a bundle over a Banach manifold and then show, using an argument from [15], that one may make any section transverse to the 0-section by perturbing the almost complex structure. Here cases of disks with unstable domains require extra care: we stabilize their domains using extra marked points on the boundary. We control these marked points using disks with higher order of convergence to Reeb chords.

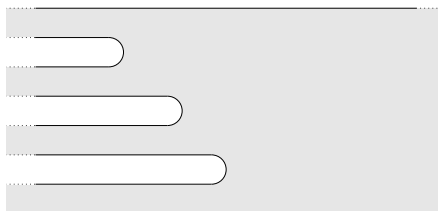


FIGURE 9.1. A standard domain.

9.1. CONFORMAL REPRESENTATIVES AND BANACH MANIFOLDS. — In order to define suitable Banach spaces for our study of holomorphic curves we endow the domains of our holomorphic disks with cylindrical ends. For convenience we choose a particular such model for each conformal structure on the punctured disks. (The precise choice is not important since the space of possible choices of cylindrical ends is contractible.)

- A *standard domain* Δ_0 with one puncture is the unit disk in the complex plane with a puncture at 1 and fixed cylindrical end $[0, 1) \rightarrow [0, 1]$ at this puncture.

- A *standard domain* Δ_1 with two punctures is the strip $\mathbb{R} \rightarrow [0, 1]$.

- A *standard domain* $\Delta_m([a_1, \dots, a_{m-1}])$ with $m + 1 > 2$ boundary punctures is a strip $\mathbb{R} \rightarrow [0, m] \subset \mathbb{C}$ with slits of small fixed width (and fixed shape) around half-infinite lines $(-\infty, a_j] \rightarrow f_j g$, where $0 < j < m$ is an integer, removed. See Figure 9.1. We say that $a_j \in \mathbb{R}$ is the j^{th} boundary maximum of $\Delta_m([a_1, \dots, a_{m-1}])$.

The space of conformal structures \mathcal{C}_m on the $(m + 1)$ -punctured disk is then represented as $\mathbb{R}^m / \mathbb{R}$ where \mathbb{R} acts on vectors of boundary maxima by overall translation, see [9, §2.1.1]. The boundary of the space of conformal structures on an $(m + 1)$ -punctured disk in its compactification $\partial \mathcal{C}_m \rightarrow \overline{\mathcal{C}}_m$ can then be understood as consisting of the several level disks which arise as some differences $|a_j - a_k|$ between boundary maxima approach 1 . We sometimes write Δ_m for a standard domain, suppressing its conformal structure $[a_1, \dots, a_{m-1}]$ from the notation.

The breaking of a standard domain into a standard domain of several levels is compatible with the compactness results Theorems 8.11 and 8.12. In the proof of these results given in [5], after adding a finite number of additional punctures the derivatives of the maps are uniformly bounded and each component in the limit has at least two punctures and can thus be represented as a standard domain. In particular, the domain right before the limit is the standard domain obtained by gluing these in the natural way and the arcs in the definition of convergence can be represented by vertical segments. Here a *vertical segment* in a standard domain $\Delta_m \subset \mathbb{C}$ is a line segment in Δ_m parallel to the imaginary axis which connects two boundary components of Δ_m .

9.2. CONFIGURATION SPACES. — In this section we construct Banach manifolds which are configuration spaces for holomorphic disks. In order to show that all moduli spaces we use are manifolds we need to stabilize disks with one and two punctures by adding punctures in a systematic way. To this end we will use Sobolev spaces with extra

weights. This is the reason for introducing somewhat more complicated spaces below. The constructions in this section parallels corresponding constructions in [15] and [19].

We first define the configuration space for holomorphic disks in $T \times Q$ and then find local coordinates for this space showing that it is a Banach manifold. We then repeat this construction for disks in the symplectization.

Below we are interested in the moduli spaces $\mathcal{M}(a, n_0; \mathbf{n})$ of holomorphic disks for $n_0 = 0$ or $n_0 = 1$ and $\mathbf{n} = (n_1, \dots, n_m)$, which we will describe as subsets of suitable configuration spaces $W = W(a; \delta_0; \mathbf{n})$. Here $\delta_0 > 0$ and n_0 are related as follows: Consider the standard neighborhood $(-\varepsilon, T + \varepsilon) \times U$ (with $U \subset \mathbb{C}^2$) of the Reeb chord $a : [0, T] \rightarrow S \times Q$ which we introduced on page 739. The projections of the contact planes at the two end points of a to \mathbb{C}^2 intersect transversally, and we denote by $0 < \theta^0 \pi < \theta^m \pi < \pi$ the two complex angles between them. Now for $n_0 = 0$ we choose $0 < \delta_0 < \theta^0$ and for $n_0 = 1$ we choose $\theta^m < \delta_0 < 1$.

The space W fibers over the product space

$$B = \mathbb{R}^{m-2} \times \mathbb{R} \times J(K).$$

The first factor \mathbb{R}^{m-2} is the space of conformal structures on the disk with $m+1$ boundary punctures. We represent the disk as a standard domain with the first boundary maximum at 0 and \mathbb{R}^{m-2} as the coordinates of the remaining $m-2$ boundary maxima. The second factor \mathbb{R} corresponds to the shift in parameterization of the asymptotic trivial strips at the positive puncture. The third factor is itself a product with one factor for each negative puncture:

$$J(K) = J^{(r_1)}(K) \times \dots \times J^{(r_m)}(K).$$

Here r_j is the smallest integer $< n_j$ and $J^{(r_j)}(K)$ denotes the r_j^{th} jet-space of K . A point $(q_0, q_1, \dots, q_{r_j}) \in J^{(r_j)}(K)$ corresponds to the first Fourier (Taylor) coefficients of the map at the j^{th} negative puncture. Note that $J(K)$ depends on $\mathbf{n} = (n_1, \dots, n_m)$, but we omit this dependence from the notation.

Fix a parameterization of each Reeb chord strip. If $\gamma \in \mathbb{R}^{m-2}$ then we write $\Delta[\gamma]$ for the standard domain with first boundary maximum at 0 and the following boundary maxima according to the components of γ . If $\mathbf{n} = (n_1, \dots, n_m) \in (\frac{1}{2}\mathbb{Z})^m$ with $\sum_j n_j \in \mathbb{Z}$ then we decorate the boundary components of $\Delta[\gamma]$ according to \mathbf{n} as follows. Start at the positive puncture and follow the boundary of $\Delta[\gamma]$ in the positive direction. Decorate the first boundary component by L_K and then when we pass the j^{th} negative puncture we change Lagrangian (from L_K to Q or vice versa) if n_j is a half integer and do not change if it is an integer.

Fix a smooth family of smooth maps

$$w_\beta : (\Delta[\beta_1], \partial\Delta[\beta_1]) \rightarrow (T \times Q, L), \quad \beta = (\beta_1, \beta_2, \beta_3) \in B,$$

with the following properties:

– w_β respects the boundary decoration, i.e., it takes boundary components decorated by L_K resp. Q to the corresponding Lagrangian submanifold.

– w_β agrees with the Reeb chord strip of a shifted by β_2 in a neighborhood of the positive puncture.

– Consider standard coordinates $\mathbb{C} \subset \mathbb{C}^2$ near the first component of $\beta_3^j \in J^{(r_j)}(K)$. Then in a strip neighborhood of the j^{th} negative puncture, the \mathbb{C}^2 -component of w_β vanishes and the \mathbb{C} -component is given by

$$w_\beta(z) = \sum_{l=0}^{r_j} q_l e^{l\pi z},$$

where the j^{th} component β_3^j of β_3 is

$$\beta_3^j = (q_0, q_1, \dots, q_{r_j}) \in J^{(r_j)}(K).$$

Let $0 < \delta < 1/2$ and as before let either $0 < \delta_0 < \theta^0$ or $\theta^{00} < \delta_0 < 1$, where θ^0 describes the smallest non-zero complex angle at the Reeb chord a and θ^{00} the largest. Let $H_{\delta_0, \delta}(\beta_1)$ denote the Sobolev space of maps

$$w: \Delta[\beta_1] \rightarrow T\mathbb{R}^3 = \mathbb{R}^6$$

with two derivatives in L^2 and finite weighted 2-norm with respect to the weight function η_δ with the following properties.

- $\eta_{\delta_0, \delta}$ equals 1 outside a neighborhood of the punctures.
- $\eta_{\delta_0, \delta}(s + it) = e^{\delta_0 \pi i s j}$ near the positive puncture.
- $\eta_{\delta_0, \delta}(s + it) = e^{(n_j - \delta) \pi i s j}$ near the j^{th} negative puncture.

Consider the bundle $E \rightarrow B$ with fiber over $\beta \in B$ given by $H_{\delta_0, \delta}(\beta_1)$. Define the configuration space $W = W(a; \delta_0; \mathbf{n}) \subset E$ of (β, w) such that $u = w_\beta + w$ satisfies the following

- u takes the boundary of $\Delta[\beta_1]$ to L respecting the boundary decoration.
- u is holomorphic on the boundary, i.e., the restriction (trace) of $\bar{\partial}_J u$ to $\partial\Delta[\beta_1]$ vanishes.

It is not hard to see that W is a closed subspace of E . In fact it is a Banach submanifold of the Banach manifold E . We will next explain how to find local coordinates on W . Let $(\beta, w) \in W$, and assume that $u = w_\beta + w$ is a map in W .

In order to find local coordinates around u we first consider the finite dimensional directions. Pick diffeomorphisms of the source $\Delta[\beta_1]$,

$$(9.1) \quad \phi_\gamma, \gamma \in \mathbb{R}; \quad \psi_{\eta_1}, \eta_1 \in \mathbb{R}^{m-1},$$

corresponding to the second and first finite dimensional factors. Here ϕ_γ equals the identity outside a neighborhood of the positive end where it equals translation by γ , and $\psi_{\eta_1}: \Delta[\beta_1] \rightarrow \Delta[\beta_1 + \eta_1]$ moves the boundary maxima according to η_1 , see [9, §6.2.3].

We next turn to the translations along the knot and the infinite dimensional component of the space. Using the coordinate map of Lemma 8.6 we import the flat metric on $T(S^1 \times D^2)$ to TQ , we extend this metric to a metric h^1 on all of TQ so that L_K is totally geodesic and flat near Reeb chord endpoints, see 8.1 (v), and such that $h^1 = ds^2 + g$ on $TQ \cap DQ = \mathbb{R}_+ \times SQ$, where g is a metric on SQ .

Consider the standard almost complex structure in a neighborhood of the zero section of $Q = \mathbb{R}^3$ in TQ . Note that this almost complex structure agrees with the standard almost complex structure in the holomorphic neighborhood of K . Using the construction in [13, Prop. 5.3], we extend it to an almost complex structure J over all of TQ with the following additional property near L_K . If V is a vector field along a geodesic in the metric h^1 in L_K then V satisfies the Jacobi equation if and only if the vector field JV does. To achieve this we might have to alter h^1 slightly near but not on L_K , see [13, Eq. (5.7)] for the precise form of h^1 (corresponding to \hat{g} in that equation). Note that this construction gives the standard almost complex structure near the knot. Let h^0 denote the standard flat metric on TQ and note that it has the Jacobi field property discussed above along Q . Let

$$(9.2) \quad h^\sigma, \quad 0 \leq \sigma \leq 1$$

be the linear interpolation between the metrics h^0 and h^1 .

Consider the pullback bundle $u^*T(TQ)$. Note that the Riemannian metrics h^t on TQ induce connections on this bundle which we denote by r^t .

Let $\dot{H}_\delta(u)$ denote the linear space of sections v of $u^*T(TQ)$ with the following properties.

- The partial derivatives of v up to second order lie in $L^2_{\text{loc}}(\Delta[\beta], u^*T(TQ))$.
- The restriction of $r^\sigma v + J r^\sigma v$ to the boundary (sometimes called the trace of $r^\sigma v + J r^\sigma v$) vanishes, where $\sigma = 1$ for a boundary component mapping to L_K and $\sigma = 0$ for a component mapping to Q .
- With $k_{k,\delta,\mathbf{n}}$ denoting the Sobolev 2-norm weighted by $\eta_{\delta,\delta_0,\mathbf{n}}$, $k v k_{k,\delta,\mathbf{n}} < 1$.

Then $\dot{H}_{2,\delta,\delta_0,\mathbf{n}}(w)$ equipped with the norm $k_{k,\delta,\delta_0,\mathbf{n}}$ is a Banach space.

Also fix $m + \sum_{j=1}^m r_j$ smooth vector fields s_k^j , $1 \leq j \leq m$ and $0 \leq k \leq r_j$ along u with properties as above and with the following additional properties.

- The vector field s_k^j is supported only near the j^{th} negative puncture in a half strip neighborhood which maps into the analytic neighborhood of the knot.
- In standard coordinates along the knot $\mathbb{C} \rightarrow \mathbb{C}^2$, the \mathbb{C}^2 -component of s_k^j equals 0 and the \mathbb{C} -component is $s_k^j = e^{k\pi z}$.

We are now ready to define the local coordinate system. Write \exp^σ for the exponential map in the Riemannian metric h^σ , $0 \leq \sigma \leq 1$, from (9.2). The local coordinate system around u has the form

$$\Phi_u : U_1 \times U_2 \times U_3 \times U \rightarrow W,$$

where $U_1 \subset \mathbb{R}^{m-2}$, $U_2 \subset \mathbb{R}$, $U_3 \subset \prod_{j=1}^m \mathbb{R}^{r_j+1}$, and $U \subset \dot{H}_{2,\delta,\delta_0,\mathbf{n}}(w)$ are small neighborhoods of the origin with coordinates $\gamma_j \in U_j$. Let $\sigma : \Delta[\beta_1] \rightarrow [0, 1]$ be a smooth function that equals 0 resp. 1 in a neighborhood of any boundary component that maps to Q resp. L_K and that equals 0 on $u^{-1}(DQ)$. For u as above we then consider

$$\Psi_u(\gamma_1, \gamma_2, \gamma_3, v)(z) = \exp_{u(z^\theta)}^{\sigma(z^\theta)} \left(v(z^\theta) + \sum_{j=1}^m \sum_{k=0}^{r_j} \gamma_{3k}^j s_k^j(z^\theta) \right), \quad z^\theta = \phi_{\gamma_1}(\psi_{\gamma_2}(z)),$$

see (9.1) for the diffeomorphisms ϕ_{γ_1} and ψ_{γ_2} . Here γ_1 corresponds to shifts near the positive puncture, γ_2 corresponds to variations of the conformal structure, γ_3 is related to variations of the map near Lagrangian intersection punctures, and v is a vector field along the curve. We use the exponential map to go from linearized variations to actual maps.

LEMMA 9.1. — *The space W is a Banach manifold with local coordinates around u given by Ψ_u .*

Proof. — This is straightforward, see [15, Lem. 3.2] for an analogous result.

Consider the bundle E over the configuration space W with fiber over u the complex anti-linear maps

$$T\Delta[\beta_1] \rightarrow T(T \times Q).$$

The $\bar{\partial}_J$ -operator gives a section of this bundle $u \mapsto (du + J \circ du \circ i)$ and the moduli space $\mathcal{M}(a; n_0, \mathbf{n})$ is the zero locus of this section, where $n_0 = 0$ if $0 < \delta_0 < \theta^0$ and $n_0 = 1$ if $\theta^{00} < \delta_0 < 1$. The section is Fredholm and the formal dimension of the solution spaces is given by its index. We have the following dimension formula.

LEMMA 9.2. — *The formal dimension of $\mathcal{M}(a, n_0; \mathbf{n})$ is given by*

$$\dim(\mathcal{M}(a, n_0; \mathbf{n})) = |a| - 2n_0 - \sum \mathbf{n}_j.$$

Proof. — The case $n_0 = 0$ follows from [5, Th. A.1 & Rem. A.2]. The fact that the index jumps when the exponential weight crosses the eigenvalues of the asymptotic operator is well known and immediately gives the other case, see e.g. [13, Prop. 6.5].

We next consider a completely analogous construction of a configuration space for holomorphic disks in $M^{SY}(a, n_0; \mathbf{b})$. We discuss mainly the points where this construction differs from that above. Consider first the finite dimensional base. Here the situation is simpler and we take instead

$$B = \mathbb{R}^{m-2} \times \mathbb{R}^{m+1},$$

where the first factor corresponds to conformal structures on the domain exactly as before and where the second factor corresponds to re-parameterizations of the trivial Reeb chord strips exactly as for the positive puncture before. We fix a smooth family of maps $w_\beta: \Delta[\beta_1] \rightarrow \mathbb{R} \times S \times Q$ which agrees with the prescribed Reeb chord strips near the punctures. We next fix an isometric embedding of $S \times Q$ into \mathbb{R}^N and consider the bundle of weighted Sobolev spaces with fiber over $\beta \in B$ the Sobolev space $H_{n_0, \delta}$ of functions with two derivatives in L^2 with respect to the norm weighed by a function which equals $e^{\delta|s|}$ in the negative ends and $e^{(\delta+n_0)|s|}$ in the positive end.

In analogy with the above we then fix (commuting) re-parameterization diffeomorphisms ψ_{β_1} corresponding to changes of the conformal structure and ϕ_{β_2} corresponding to translation in the half strip neighborhoods. Again this then leads to a Fredholm section and its index gives the formal dimension of the moduli space.

LEMMA 9.3. — *The formal dimension of $\mathcal{M}^{\text{sy}}(a, n_0; \mathbf{b})$ is given by*

$$\dim \mathcal{M}^{\text{sy}}(a, n_0; \mathbf{b}) = |a| - 2n_0 - |\mathbf{b}|.$$

Proof. — See [5, Th. A.1 & Rem. A.2] and use the relation between weights and index, see e.g. [13, Prop. 6.5].

REMARK 9.4. — We consider for future reference the conformal variations of the domain with more details. In the local coordinates around a map $w: \Delta_{m+1} \rightarrow T \times Q$ or $w: \Delta_{m+1} \rightarrow \mathbb{R} \times S \times Q$ defined above, the conformal variations correspond to a diffeomorphism that moves the boundary maxima of the domain. We take such a diffeomorphism to be a shift along a constant (and hence holomorphic) vector field τ in the real direction around the boundary maximum and then cut it off in nearby strip regions. Hence the corresponding linearized variation $L\bar{\partial}_J(\gamma)$ at w , where γ is the first order variation of the complex structure corresponding in the domain is

$$L\bar{\partial}_J(\gamma) = \partial_J w \cdot \bar{\partial}\tau.$$

We will sometimes use other ways of expressing conformal variations, where the variations are supported near a specific negative puncture rather than near a specific boundary maximum. To this end we first note that we may shift the conformal variation by any element $L\bar{\partial}_J(v)$ where v is a vector field along w in the Sobolev space H_δ . In particular we can shift γ by $\bar{\partial}\sigma$ where σ is a vector field along Δ_{m+1} that is constant near the punctures. In this way we get equivalent conformal variations γ_q of the form

$$L\bar{\partial}_J(\gamma_q) = \partial_J w \cdot \bar{\partial}\tau_q.$$

where τ_q is a vector field of the form

$$\tau_q(z) = \beta(s + it)e^{\pi(s+it)},$$

where $s + it$ is a standard coordinate in the strip neighborhood of the negative puncture a and β is a cut-off function equal to 1 near the puncture and 0 outside a strip neighborhood of the puncture. We refer to [9, §2.1.1] for details.

9.3. TRANSVERSALITY. — We next use the special form of our almost complex structure near Reeb chords in combination with an argument from [15, Lem. 4.5] to show that we can achieve transversality for $\bar{\partial}_J$ -section of E over W by perturbing the almost complex structure. In other words we need to show that the linearization $L\bar{\partial}_J$ of the section $\bar{\partial}_J$ is surjective.

LEMMA 9.5. — *For generic J any solutions in $\mathcal{M}(a, n_0; \mathbf{n})$ and $\mathcal{M}^{\text{sy}}(a, n_0; \mathbf{b})$ are transversely cut out.*

Proof. — To see this we perturb the almost complex structure near the positive puncture. Consider the local projection to \mathbb{C}^2 near the Reeb chord. Here the Lagrangians correspond to two Lagrangian planes. Furthermore the holomorphic disks admit local Taylor expansions near the points that map to their intersection. The lemma now follows from the proof of [15, Lem. 4.5]. We sketch the argument.

Let U denote a neighborhood of the Reeb chord strip C_a of a for M^{sy} or of the Reeb chord strip in $T^*Q \rightarrow D^*Q$ for M . If u is a holomorphic disk then $u^{-1}(U \setminus C_a)$ is the pre-image under u composed with the projection to \mathbb{C}^2 of the intersection point of the two Lagrangian planes. It follows by monotonicity that the preimage is a finite collection of points $f_{q_0, q_1, \dots, q_r}g$, where q_0 is the positive puncture. If q_i is an interior point, let E_i denote a small disk around q_i , if q_j is a boundary point let E_j denote a half-disk neighborhood of q_j . If the map u has an injective point near the double point then a standard argument perturbing the almost complex structure there establishes the necessary transversality. We therefore assume that this is not the case. Consider the image of a small half disk E_0 near the positive puncture q_0 , and note that the boundary arcs end at the positive punctures. Since the map is not injective there are neighborhoods (after renumbering) E_1, \dots, E_m where u agrees with the image γ under u of one of the boundary arcs of E_0 . By analytic continuation, the images of these neighborhoods then intersect the Lagrangian sheet of the boundary arc γ^∂ that contains γ . Consequently, the map has multiplicity $m + 1$ along γ and multiplicity m along $\gamma^\partial \setminus \gamma$. Consider a vector field in the cokernel of the linearized operator $L\bar{D}$. Perturbing the almost complex structure near $\gamma^\partial \setminus \gamma$ we see that the contributions from the anti-holomorphic cokernel vector field on E_1, \dots, E_m must vanish. By unique continuation, the contributions from E_1, \dots, E_m must then also cancel along γ and it follows that there is nothing that cancels the perturbation in E_0 (just as if the map was injective in E_0). The desired transversality follows.

9.4. STABILIZATION OF DOMAINS. — For disks with more than three punctures the transversality results in Section 9.3 directly give the solution spaces the structure of C^1 -smooth manifolds. For the case of unstable domains this is not as direct since the solutions admit re-parameterizations that do not act with any uniformity on the associated configuration spaces. This is a well-known phenomenon and we resolve the problem by a gauge fixing procedure, adding marked points near the positive puncture. This construction was studied in detail in [18, App. A.2] and in [19, §5.2 and 6] and we will refer to these articles for details.

As we shall see below we need only consider moduli spaces of dimension ≤ 2 . Recall the neighborhood $a \in U, U \rightarrow \mathbb{C}^2$ of the Reeb chord a , see the discussion before Definition 8.1 on page 739, and the corresponding Fourier expansion of the \mathbb{C}^2 -component of any holomorphic disk near a , see (8.2) on page 745.

Consider a space $M(a, \mathbf{n})$ of formal dimension ≤ 1 . Then by Lemmas 9.2 and 9.5 the corresponding space $M(a, 1; \mathbf{n})$ is empty. Consequently, for any solution $u \in M(a, \mathbf{n})$, the first Fourier coefficient of the \mathbb{C}^2 -component of the map near a is non-vanishing. Let $S_{0;\varepsilon}$ and $S_{1;\varepsilon}$ be spheres in Λ_K of radii $\varepsilon > 0$ around the Reeb chord endpoints of a . Non-vanishing of the first Fourier coefficient in combination with compactness then implies that for each solution u there are two unique points in the boundary of the domain closest to the positive puncture that map to $S_{j;\varepsilon}$, $j = 0, 1$, see Figure 9.2. We add punctures at these points. More precisely, we consider standard domains with two more punctures and require that the maps are

asymptotic to points in $S_{j;\varepsilon}$ at the extra punctures. In the above notation these would be “Lagrangian intersection punctures” in $S_{j;\varepsilon}$ of local winding number 1 in the direction normal to $S_{j;\varepsilon}$. The transversality result 9.5 holds as before also for the solution spaces with extra punctures, so that they are C^1 -manifolds. The asymptotic properties above then imply that the solutions with extra punctures capture all holomorphic disks.

Consider next a space $M^{\text{sy}}(a; \mathbf{b})$ of formal dimension ≤ 2 . Since any holomorphic curve in the symplectization can be translated we find that the corresponding space $M^{\text{sy}}(a, 1; \mathbf{b})$ is again empty and we get a manifold structure by adding two marked points near the Reeb chord endpoints exactly as above.

It remains then to consider the case of spaces $M(a; \mathbf{n})$ of formal dimension 2. Here the corresponding space $M(a, 1; \mathbf{n})$ has dimension 0. There are then a finite number of solutions with this decay condition. Considering the Fourier expansion we can fix unique marked points for all solutions in a neighborhood V (in the configuration space) of these isolated solutions as above. For solutions outside V the Fourier coefficients do not vanish and we can fix marked points as above. Note however, that these will generally not be the same marked points. This way we however get two types of manifold charts: one for solutions inside V and one for solutions in a neighborhood of any map u^θ with nonvanishing first Fourier coefficient which lies outside a smaller neighborhood V^θ of u . To get a manifold structure for the moduli space we then need to study the transition maps, and to that end we use four marked points, see Figure 9.2 and [19, §5.2] for details.

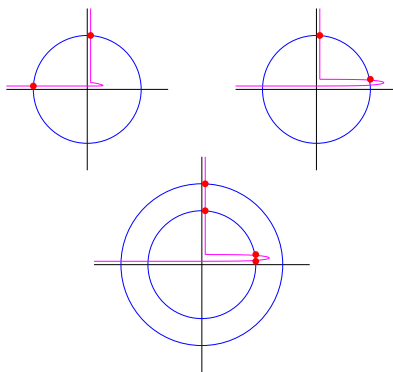


FIGURE 9.2. Top left: marked points for a disk near the disk with degenerate asymptotics. Top right: marked points for a disk outside a neighborhood of the disk with degenerate asymptotics. Lower: the four marked points in the intermediate region used to define the coordinate change. The (black) lines represent the projections of the branches of $L_K \setminus Q$ to \mathbb{C}^2 , and the (blue) circles represent 3-spheres S_ε^3 , which cut these local branches along the circles $S_{j;\varepsilon}$ appearing in the text.

A priori, the smooth structures on the moduli spaces above depend on the choice of gauge condition. However, using the fact that the C^0 -norm of a holomorphic map controls all other norms, it is not hard to see that different gauge conditions lead to the same smooth structure.

We also need to show that the compactness result where sequences of curves converge to several level curves are compatible with additional marked points. This is similar to the above. The compactness result we already have implies uniform convergence on compact sets and in particular it is possible to add marked points on the curves near the limit that correspond to the extra marked points on the unstable curves in the limit. As before we show that these extra marked points do not affect the moduli spaces. See [19, §A.3] for details. In conclusion, by adding marked points also on curves near broken limits we obtain versions of the compactness results Theorems 8.11 and 8.12 where all domains involved are stable with marked points compatible with the several level breaking.

9.5. INDEX BUNDLES AND ORIENTATIONS. — Viewing the $\bar{\partial}_J$ -operator as a Fredholm section of a Banach bundle, its linearization defines an index bundle over the configuration space and an orientation of this index bundle gives an orientation of transverse solution spaces. Following Fukaya, Oh, Ohta, Ono [22, §8.1] one defines a coherent system of such orientations as follows. Fix spin structures of the two Lagrangians L_K and Q , which we here can think of as trivializations of the respective tangent bundles. Consider a closed disk with boundary in one of the two Lagrangians and the linearized $\bar{\partial}_J$ -operator acting on vector fields along this disk that are tangent to the Lagrangian along the boundary. Using the trivialization of the boundary condition, such an operator can be deformed to an operator on the disk with values in \mathbb{C}^3 and constant \mathbb{R}^3 boundary condition, with a copy of CP^1 attached at the center with a complex linear operator. The first operator has trivial cokernel and a kernel that consists only of constant vector fields, and the orientation of \mathbb{C} induces an orientation on the determinant bundle of the operator over CP^1 . This gives a canonical orientation over closed disks with trivialized boundary condition (that depends only on the homotopy class of the trivialization).

Here we need to orient moduli spaces of disks with punctures. This was done in the setting of Legendrian contact homology in [14]; we will give a sketch and refer to that reference for details. We reduce to the case of closed disks by picking so-called capping operators at all Reeb chords and along the Lagrangian intersection K with an orientation of the corresponding determinant bundles. Here it is important that the capping operators are chosen in a consistent way. At Reeb chords there is a positive and a negative capping operator and we require that they glue to the standard orientation on the closed disk. We also pick positive and negative capping operators at the Lagrangian intersection punctures satisfying the same conditions. Now, given a holomorphic disk in the symplectization or in T^*Q we glue the capping operators to it and produce a closed disk. The standard orientation of the closed disk and the chosen orientations on the capping operators then give an orientation of the determinant line

of the linearized operator over the disk, which, together with an orientation of the finite dimensional space of conformal structures on the punctured disk, in turn gives an orientation of the moduli space if it is transversely cut out. The gluing condition for the capping operators ensures that the resulting orientations of the moduli spaces are compatible with splittings into multi-level curves.

In what follows we assume that spin structures on the Lagrangians and capping operators have been fixed and thus all our moduli spaces are oriented manifolds.

9.6. SIGNS AND THE CHAIN MAP EQUATION. — Recall the chain map

$$\Phi: (C(R), \partial) \rightarrow (C(\Sigma), \partial + \delta_Q + \delta_N)$$

from Theorem 6.13. Here we consider the signs of the operations δ_Q and δ_N in this formula. These operations are defined on chains of broken strings by taking the oriented preimage of K under the evaluation map. In the map Φ , the oriented chain is given by a moduli space of holomorphic disks. In order to deal with the evaluation maps on such spaces we present them as bundles over Q as follows. Consider first the operation δ_Q . Fix a point $q \in Q$ and an additional puncture on the boundary that we require maps to q . Concretely, we work on strips with slits and add a small positive exponential weight at the puncture mapping to q . Then we consider the bundle of such maps over Q when we let q vary in Q . The orientation of this space is induced from capping operators as described above. When we consider the corresponding boundary condition on the closed disk we find a vanishing condition for linearized variations at the marked point corresponding to the positive exponential weight. Thus if σ denotes the orientation of the index bundle induced as above, then the orientation on the bundle with marked point mapping to q is given by the orientation of the formal difference $\sigma \otimes TQ$. (The formal difference should be interpreted as in K -theory: the difference $\xi \otimes \eta$ of two bundles ξ and η is represented by a bundle ζ such that the direct sum $\zeta \oplus \eta$ is equivalent to ξ .)

We point out that here and throughout this section orientations depend on ordering conventions, whether the point condition goes before or after the index bundle, etc. In calculations below we put point conditions after the index bundle, and put the fiber of bundles over Q before the base.

The orientation of the bundle corresponding to a point constraint q varying over Q is then given by $\sigma \otimes TQ \otimes TQ$. Finally, the orientation of the chain given by the preimage of K under the evaluation map is then

$$(9.3) \quad \sigma \otimes TQ \otimes TQ \otimes TK \otimes TQ = \sigma \otimes TK \otimes TQ.$$

In order to show that the chain map equation holds we must then show that there are choices of capping operators and orientations on Q and N so that this orientation agrees with the boundary orientation of the disk viewed as the boundary in the moduli space of disks with two colliding Lagrangian punctures.

Consider the capping operators c_{QN} and c_{NQ} for such a puncture going from Q to N and vice versa. These capping operators are standard $\bar{\partial}$ -operators on a once

punctured disk D_1 acting on C^3 -valued functions in a weighted Sobolev space that satisfy a Lagrangian boundary condition.

We first describe the boundary conditions. For c_{QN} the Lagrangian boundary condition $\lambda: \partial D_1 \rightarrow \text{Lag}_3$, where Lag_3 denotes the Lagrangian Grassmannian of Lagrangian subspaces of C^3 , starts at the tangent space of Q and ends at the tangent space of N . For c_{NQ} the boundary condition instead starts at the tangent space of N and ends at the of Q . More specifically, the tangent spaces of Q and N intersect along TK and are perpendicular in the normal directions of K . We think of the normal directions to K as C^2 and the tangent spaces of Q and N as iR^2 and R^2 , respectively. We take both capping operators c_{QN} and c_{NQ} to fix TK , to be a rotation by $\pi/2$ in one of the complex lines normal to the knot, and a rotation by $3\pi/2$ in the other.

We next describe the weights at the puncture in D_1 . We use a half strip neighborhood of the puncture and a Sobolev space with small positive exponential weight δ , $0 < \delta < \pi/2$, in this strip neighborhood.

The index of the $\bar{\partial}$ -operator with this boundary condition and weight equals 3, see e.g. [13, Prop. 6.5].

Recall from Section 9.5 that an orientation of the moduli space is induced from the capping operators together with an orientation on the space of conformal structures on the punctured disk. Here we think of variations of the conformal structure as vector fields moving the punctures along the boundary of the disk. We have one such vector field for each puncture which give an additional one dimensional oriented vector space associated to each puncture, see [14, §3.4.1] for details. For simplicity we write simply c_{QN} and c_{NQ} for the sum of the index bundles of the capping operator described above and one dimensional conformal variations associated to the respective punctures. Thus, in the calculations below c_{QN} and c_{NQ} have index $3 + 1 = 4$.

We choose the orientations on Q and N so that the linear transformations between tangent spaces TQ and TN induced by the Lagrangian boundary conditions of c_{QN} and c_{NQ} take the orientation on Q to that on N and vice versa.

The boundary orientation of the two-level disk (second level constant) is the fiber product over K of the orientations of its levels. We view the top level disk as having a small positive exponential weight at the puncture mapping to K and a cut-off local solution in the direction of K . In analogy with the above, its orientation is thus given by $\sigma \otimes TQ \otimes TK$. The orientation of the constant disk (which has small negative weights at its positive puncture) is then $\sigma^\theta \otimes c_{QN} \otimes c_{NQ}$, where σ^θ is the standard orientation on the closed up boundary condition of the constant three punctured disk. The boundary orientation is thus

$$(9.4) \quad (\sigma \otimes TQ \otimes TK) \otimes (\sigma^\theta \otimes c_{QN} \otimes c_{NQ}) \otimes TK.$$

Now choose the orientation on c_{QN} and c_{NQ} so that the orientation of the index one problem on the constant disk with kernel in direction of the knot induced by $\sigma^\theta \otimes c_{QN} \otimes c_{NQ}$ is opposite to the orientation of TK . Then the orientation in (9.4) is

$\sigma \text{ TK } TQ$ (there is an orientation change when one permutes the odd-dimensional summands TK and TQ), in agreement with (9.3).

For the sign of the operation δ_N we argue exactly as above replacing Q with N and we must compare the orientations $\sigma \text{ TK } TN$ and

$$(\sigma \text{ TN } TK) (\sigma^{\theta} c_{NQ} c_{QN}) TK.$$

Compared to the above the main difference is that the summands c_{NQ} and c_{QN} have been permuted. However, as explained above, the index of each of these operators is 4, so the orientation remains as before and the positive sign for δ_N is correct for the chain map.

10. COMPACTIFICATION OF MODULI SPACES AND GLUING

In this section we show that the moduli spaces $\mathcal{M}(a; \mathbf{n})$ and $\mathcal{M}^{\text{sy}}(a; \mathbf{b})/\mathbb{R}$ admit compactifications as manifolds with boundary with corners. Furthermore, we describe the boundary explicitly in terms of broken holomorphic disks. The smoothness of individual strata of the compactified moduli spaces are governed by the Transversality Lemma 9.5. The Compactness Theorems 8.12 and 8.11 describe disk configurations in the boundary of the compactification. The main purpose of this section is thus to show how to glue these configurations on the boundary to curves in the smooth part of the moduli space and thereby obtain boundary charts in the sense of manifolds with boundary with corners. Such gluing theorems were proved before in closely related situations and we will discuss details only when they differ from the standard cases.

We first state the structural theorems in Section 10.1 and then turn to the gluing results and their proofs in the following subsections.

We work throughout this section with an almost complex structure J so that Lemma 9.5 holds. Furthermore we assume that the domains of all holomorphic disks are stable, which can be achieved by adding marked points as explained in Section 9.4.

10.1. STRUCTURE OF THE MODULI SPACES. — In this subsection we state the results on moduli spaces of holomorphic disks. As before there are two cases to consider, disks in the symplectization and disks in the cotangent bundle. The structural results all have the same flavor. Basically we show that a specified moduli space is a manifold with boundary with corners of dimension ≤ 2 , and we describe the boundary strata as well as certain submanifolds important for our study. The proofs of the results are the main goal for the rest of the section.

Recall from Sections 9.2 and 9.3 (with $n_0 = 0$) that for generic J the moduli spaces $\mathcal{M}(a; \mathbf{n})$ and $\mathcal{M}^{\text{sy}}(a; \mathbf{b})$ are manifolds of dimensions

$$\dim \mathcal{M}(a; \mathbf{n}) = |aj| - |\mathbf{n}j|, \quad \dim \mathcal{M}^{\text{sy}}(a; \mathbf{b}) = |aj| - |\mathbf{b}j|.$$

Here $|aj| = \text{ind}(a)$ is the degree of the Reeb chord a (which takes only values $0, 1, 2$), and to the vector of local winding numbers $\mathbf{n} = (n_1, \dots, n_m)$ (where the n_j are

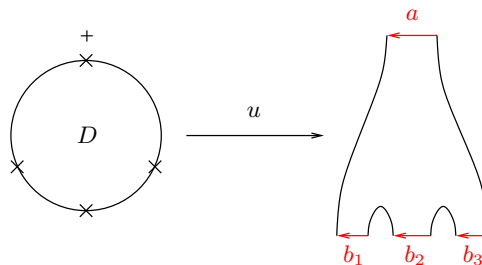


FIGURE 10.1. Disks $u : (D, \partial D) \rightarrow (R \setminus S \setminus Q, R \setminus \Lambda_K)$ in the symplectization.

positive half-integers or integers) we have associated the nonnegative integer

$$j\mathbf{n}j = \sum_{j=1}^m 2(n_j - \frac{1}{2}) > 0.$$

If either \mathbf{n} or \mathbf{b} is empty, the corresponding contribution to the index formula is 0. If a is a Reeb chord of $\Lambda_K \setminus S \setminus Q$, then $0 < j\mathbf{a}j < 2$. Since J_1 is R -invariant, 0-dimensional moduli spaces in the symplectization consists only of Reeb chord strips. Thus the only non-empty moduli spaces $\mathcal{M}^{SY}(a; \mathbf{b})$ of dimension d^{SY} are the following (write $\mathbf{b} = b_1 \dots b_m$), see Figure 10.1:

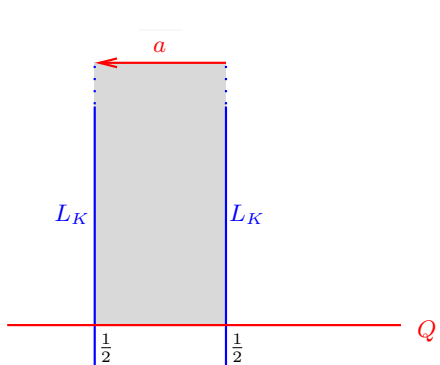
- $[2, 0]^{SY}$: If $j\mathbf{a}j = 2$ and $j\mathbf{b}j = 0$ (i.e., $j\mathbf{b}_j j = 0$ for all j) then $d^{SY} = 2$.
- $[2, 1]^{SY}$: If $j\mathbf{a}j = 2$ and $j\mathbf{b}j = 1$ (i.e., $j\mathbf{b}_j j = 0$ for all $j \notin s$ and $j\mathbf{b}_s j = 1$) then $d^{SY} = 1$.
- $[1, 0]^{SY}$: If $j\mathbf{a}j = 1$ and $j\mathbf{b}j = 0$ then $d^{SY} = 1$.

Similarly, the only non-empty moduli spaces $\mathcal{M}(a; \mathbf{n})$ of dimension d are the following (write $\mathbf{n} = n_1 \dots n_m$), see Figures 10.2, 10.3, 10.4, 10.5:

- $[2, 0]$: If $j\mathbf{a}j = 2$ and all $n_j = \frac{1}{2}$, then $j\mathbf{n}j = 0$ and $d = 2$.
- $[2, 1]$: If $j\mathbf{a}j = 2$ and $n_j = \frac{1}{2}$ for all $j \notin s$ and $n_s = 1$, then $j\mathbf{n}j = 1$ and $d = 1$.
- $[2, \frac{3}{2}]$: If $j\mathbf{a}j = 2$ and $n_j = \frac{1}{2}$ for all $j \notin s$ and $n_s = \frac{3}{2}$, then $j\mathbf{n}j = 2$ and $d = 0$.
- $[2, 2]$: If $j\mathbf{a}j = 2$ and $n_j = \frac{1}{2}$ for all $j \notin s, t$, and $n_s = n_t = 1$, then $j\mathbf{n}j = 2$ and $d = 0$.
- $[1, 0]$: If $j\mathbf{a}j = 1$ and all $n_j = \frac{1}{2}$, then $j\mathbf{n}j = 0$ and $d = 1$.
- $[1, 1]$: If $j\mathbf{a}j = 1$ and $n_j = \frac{1}{2}$ for all $j \notin s$ and $n_s = 1$, then $j\mathbf{n}j = 1$ and $d = 0$.
- $[0, 0]$: If $j\mathbf{a}j = 0$ and $n_j = \frac{1}{2}$ all j , then $j\mathbf{n}j = 0$ and $d = 0$.

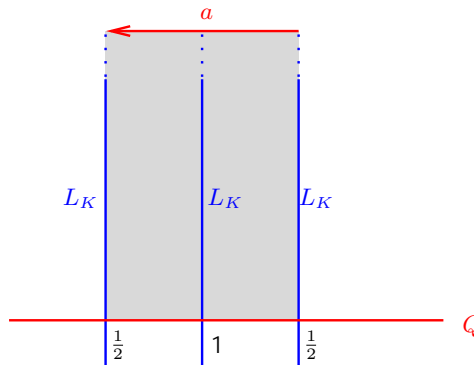
It follows from Theorem 8.12 and Lemma 9.5 (see also Section 9.4) that the 0-dimensional moduli spaces listed above are transversely cut out compact 0-manifolds. The corresponding structure theorems for moduli spaces of dimension one and two are the following.

Recall that R acts on holomorphic disks in the symplectization $R \setminus S \setminus Q$ by translation. Dividing out this action, we obtain moduli spaces of dimension zero and one in the symplectization which have the following structure.



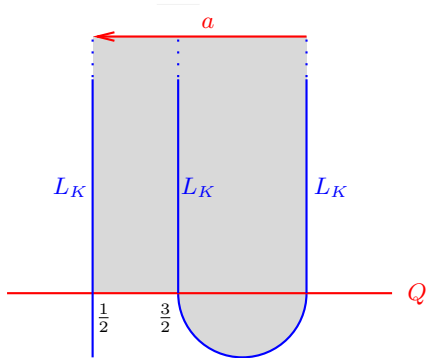
$[2, 0] : |ja| = 2, \dim = 2$
 $[1, 0] : |ja| = 1, \dim = 1$
 $[0, 0] : |ja| = 0, \dim = 0$

FIGURE 10.2. Curves with $|jn| = 0$.



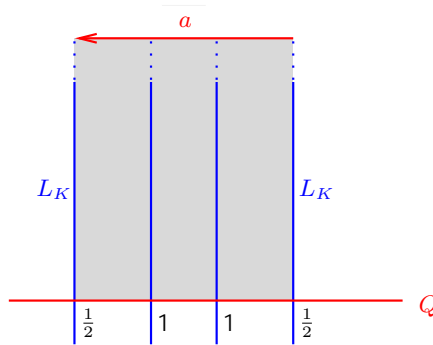
$[2, 1] : |ja| = 2, \dim = 1$
 $[1, 1] : |ja| = 1, \dim = 0$

FIGURE 10.3. Curves with $|jn| = 1$.



$[2, \frac{3}{2}] : |ja| = 2, \dim = 0$

FIGURE 10.4. Curves with $|jn| = 2$.



$[2, 2] : |ja| = 2, \dim = 0$

FIGURE 10.5. Curves with $|jn| = 2$.

THEOREM 10.1. — *Moduli spaces of holomorphic disks in the symplectization satisfy the following.*

(i) *If $M^{\text{sy}}(a; \mathbf{b})$ is a moduli space of type $[2, 1]^{\text{sy}}$ or of type $[1, 0]^{\text{sy}}$, then $M^{\text{sy}}(a; \mathbf{b})/\mathbb{R}$ is a compact 0-manifold.*

(ii) *If $M^{\text{sy}}(a; \mathbf{b})$ is a moduli space of type $[2, 0]^{\text{sy}}$, then $M^{\text{sy}}(a; \mathbf{b})/\mathbb{R}$ admits a natural compactification $\overline{M^{\text{sy}}}(a; \mathbf{b})$ which is a compact 1-manifold with boundary. Boundary points of $\overline{M^{\text{sy}}}(a; \mathbf{b})$ correspond to two-level disks \widehat{v} where the level one disk v^1 is of type $[2, 1]^{\text{sy}}$, and where exactly one level two disk $v^{2,s}$ is of type $[1, 0]^{\text{sy}}$ and all other level two disks $v^{2,j}, j \notin s$, are trivial Reeb chord strips.*

In the cotangent bundle we have moduli spaces of dimension zero, one, or two. We start with the 0-dimensional case.

THEOREM 10.2. — *Moduli spaces $\mathcal{M}(a; \mathbf{n})$ of holomorphic disks of types $[2, \frac{3}{2}]$, $[2, 2]$, $[1, 1]$, or $[0, 0]$ are compact 0-dimensional manifolds.*

In the 1-dimensional case we consider two cases separately. We first consider the case when $\sum \mathbf{n}_j = 0$.

THEOREM 10.3. — *Moduli spaces $\mathcal{M}(a; \mathbf{n})$ of disks of type $[1, 0]$ admit natural compactifications $\overline{\mathcal{M}}(a; \mathbf{n})$ which are 1-manifolds with boundary. Boundary points of $\overline{\mathcal{M}}(a; \mathbf{n})$ correspond to the following.*

- (a) *Two-level disks \dot{v} where the level one disk v^1 has type $[1, 1]$ and where the second level is a three punctured constant disk v^2 attached at the Lagrangian intersection puncture of v^1 where the asymptotic winding number equals 1.*
- (b) *Two-level disks \dot{v} where the top level disk v^1 is a symplectization disk of type $[1, 0]^{\text{sy}}$ and where all the second level disks $v^{2,j}$, $1 \leq j \leq k$ are of type $[0, 0]$.*
- (c) *If there are no entries in \mathbf{n} , then all points of the reduced moduli space $\mathcal{M}^{\text{sy}}(a; ?)/\mathbb{R}$ containing disks of type $[1, 0]^{\text{sy}}$ appear as boundary points.*

In the second 1-dimensional case $\sum \mathbf{n}_j = 1$ and we have the following.

THEOREM 10.4. — *Moduli spaces $\mathcal{M}(a; \mathbf{n})$ of disks of type $[2, 1]$ admit natural compactifications $\overline{\mathcal{M}}(a; \mathbf{n})$ which are 1-manifolds with boundary. Boundary points of $\overline{\mathcal{M}}(a; \mathbf{n})$ correspond to the following.*

- (a) *Two-level disks \dot{v} where the level one disk v^1 has type $[2, 2]$ and where the second level is a three punctured constant disk v^2 attached at the new-born Lagrangian intersection puncture of v^1 with winding number 1.*
- (b) *Two-level disks \dot{v} where the top level disk v^1 is a symplectization disk of type $[2, 1]^{\text{sy}}$ and where the second level consists of disks $v^{2,j}$, $1 \leq j \leq k$ such that for some s , $v^{2,s}$ has type $[1, 1]$ and $v^{2,j}$ has type $[0, 0]$ for $j \neq s$.*
- (c) *Two-level disks \dot{v} where the level one disk is of type $[2, \frac{3}{2}]$ and where the second level disk is a constant three punctured disk attached at the Lagrangian intersection puncture with winding number $\frac{3}{2}$. (Here the constant disk has winding number $\frac{3}{2}$ at its positive puncture, and 1 and $\frac{1}{2}$ at its negative punctures.)*

REMARK 10.5. — In order to parameterize a neighborhood of the boundary points in Theorem 10.3(a) and Theorem 10.4(a) one can use the local model (4.1) from Section 4.3. Here the location $\varepsilon > 0$ of the puncture on the real axis can be used as local coordinate for the moduli space. Furthermore, the maps in the moduli space differ from the map in (4.1) by terms of order $O(z^2)$, so they have a spike that vanishes as $\varepsilon \rightarrow 0$ as shown at the top of Figure 4.1. Similarly, in order to parameterize a neighborhood of the boundary points in Theorem 10.4 (c) one can use the local model (4.2) from Section 4.3 with $\delta = 0$. Here the location $\varepsilon > 0$ of the puncture on the real axis can be used as local coordinate for the moduli space. Furthermore, the maps in the moduli space differ from the map in (4.2) by terms of order $O(z^{5/2})$, so they have a spike that vanishes as $\varepsilon \rightarrow 0$ as shown at the bottom of Figure 4.1 (with $\delta = 0$). See Remark 10.16 for details.

In the 2-dimensional case we have the following description of the structure of the moduli space which is naturally more involved.

THEOREM 10.6. — *Moduli spaces $\mathcal{M}(a; \mathbf{n})$ of disks of type $[2, 0]$ admit natural compactifications $\overline{\mathcal{M}}(a; \mathbf{n})$ which are 2-manifolds with boundary with corners. The top-dimensional strata of the boundary have codimension 1 in $\overline{\mathcal{M}}(a; \mathbf{n})$ and correspond to the following.*

(a1) *Two-level disks \dot{v} where the top level disk v^1 has type $[2, 1]$ and where the second level is a three punctured constant disk v^2 attached at the Lagrangian intersection puncture of v^1 where the asymptotic winding number equals 1.*

(b1) *Two-level disks \dot{v} where the top level disk v^1 is a symplectization disk of type $[2, 0]^{\text{SY}}$ and where all the second level disks $v^{2,j}$, $1 \leq j \leq k$ are of type $[0, 0]$.*

(c1) *Two-level disks \dot{v} where the top level disk v^1 is a symplectization disk of type $[2, 1]^{\text{SY}}$ and where the second level consists of disks $v^{2,j}$, $1 \leq j \leq k$ such that for some s , $v^{2,s}$ has type $[1, 0]$ and $v^{2,j}$ has type $[0, 0]$ for $j \neq s$.*

The corner points on the boundary (i.e., the codimension two strata) of $\overline{\mathcal{M}}(a; \mathbf{n})$ correspond to the following.

(a2) *Two-level disks \dot{v} where the top level disk v^1 has type $[2, 2]$ and where the second level consists of two three punctured constant disks $v^{2,1}$ and $v^{2,2}$ attached at the Lagrangian intersection punctures of v^1 where the winding numbers are 1.*

(b2) *Three-level disks \dot{v} where the top level disk v^1 is a symplectization disk of type $[2, 1]^{\text{SY}}$, where the second level disk $v^{2,s}$ is of type $[1, 1]$ and all other second level disks $v^{2,j}$, $j \neq s$ are of type $[0, 0]$, and where the third level consists of a constant three punctured disk v^3 attached at the Lagrangian intersection puncture of $v^{2,s}$ with winding number 1.*

(c2) *Three-level disks \dot{v} where the top level disk v^1 is a symplectization disk of type $[2, 1]^{\text{SY}}$, where the second level disk $v^{2,s}$ is of type $[1, 0]^{\text{SY}}$ and all other second level disks $v^{2,j}$ are Reeb chord strips, and where the third level consists of disks $v^{3,j}$ all of type $[0, 0]$.*

(d2) *Two-level disks \dot{v} where the top level disk v^1 has type $[2, \frac{3}{2}]$ and where the second level consists of a 4-punctured constant disk v^2 attached at the Lagrangian intersection puncture of v^1 where the asymptotic winding number is $\frac{3}{2}$.*

REMARK 10.7. — In order to parameterize a neighborhood of the corner points in Theorem 10.6 (d2) one can use the local model (4.2) from Section 4.3 Here the locations (ε, δ) of the punctures on the real axis can be used as local coordinates for the moduli space. Furthermore, the maps in the moduli space differ from the map in (4.2) by terms of order $O(z^{5/2})$, so they have two spikes that vanish as $\varepsilon, \delta \rightarrow 0$ as shown at the bottom of Figure 4.1. See Remark 10.17 for details.

Some of the moduli spaces above admit natural maps into others by forgetting some Lagrangian intersection punctures. We next describe such maps. It is convenient to

write

$$\frac{1}{2}^s = \frac{1}{2}, \dots, \frac{1}{2}.$$

We consider first the case when the target is a one dimensional moduli space.

THEOREM 10.8. — Consider a moduli space $M(a; \frac{1}{2}^s, 1, \frac{1}{2}^t)$ of disks of type $[1, 1]$. Forgetting the $(s + 1)^{\text{th}}$ Lagrangian intersection puncture we get a map

$$M(a; \frac{1}{2}^s, 1, \frac{1}{2}^t) \rightarrow \overline{M}(a; \frac{1}{2}^{s+t})$$

into the compactified moduli space of disks of type $[1, 0]$. This map is an embedding of a 0-dimensional manifold into the interior of a 1-manifold.

Finally, we consider similar maps when the target space is two dimensional.

THEOREM 10.9. — Consider a compactified moduli space $\overline{M}(a; \frac{1}{2}^s, 1, \frac{1}{2}^t)$ of disks of type $[2, 1]$. Forgetting the $(s + 1)^{\text{th}}$ Lagrangian intersection puncture we get a map

$$\iota: \overline{M}(a; \frac{1}{2}^s, 1, \frac{1}{2}^t) \rightarrow \overline{M}(a; \frac{1}{2}^{s+t}) = \overline{M}$$

into the compactified moduli space of disks of type $[2, 0]$. This map is an immersion of a 1-dimensional manifold into a 2-manifold with boundary with corners. Let \overline{M}_{s+1} denote the image of this immersion. Then \overline{M}_{s+1} consists of those disks for which some point in the $(s+1)^{\text{th}}$ boundary arc hits K . Then \overline{M}_{s+1} and \overline{M}_{t+1} intersect (self-intersect if $s = t$) transversely at disks with two points hitting K (this corresponds to disks of type $[2, 2]$). The boundary of \overline{M}_s consists of points in the codimension one boundary of \overline{M} corresponding to disks as in Theorem 10.4 (a) and (b) as well as to interior points corresponding to disks of type $[2, \frac{3}{2}]$ as in Theorem 10.4 (c). Furthermore \overline{M}_{s+1} and \overline{M}_{s+2} with a common boundary point corresponding to a disk of type $[2, \frac{3}{2}]$ fit together smoothly at this point.

10.2. FLOER'S PICARD LEMMA. — In the following subsections we show that the broken disks in Theorems 10.3–10.4 can be glued in a unique way to give disks in the interior of the moduli space thus providing a standard neighborhood of the boundary of the moduli space inside the compactified moduli space. Our approach here is standard and starts from Floer's Picard lemma, see [26] for a proof.

LEMMA 10.10. — Let $f: B_1 \rightarrow B_2$ be a smooth map of Banach spaces which satisfies

$$f(v) = f(0) + df(0)v + N(v),$$

where $df(0)$ is Fredholm and has a right inverse Q satisfying

$$(10.1) \quad \|QN(u) - QN(v)\| \leq G(\|u\| + \|v\|)$$

for some constant G . Let $B(0, r)$ be the r -ball centered at $0 \in B_1$ and assume that

$$(10.2) \quad \|Qdf(0)\| \leq \frac{1}{8G}.$$

Then for $r < 1/4G$, the zero-set $f^{-1}(0) \setminus B(0, r)$ is a smooth submanifold of dimension $\dim(\ker(df(0)))$ diffeomorphic to the r -ball in $\ker(df(0))$.

We will apply this result as well as a parameterized version of it, see [18, Lem. 5.13]. In our case f will be the $\bar{\partial}_J$ -operator. To show existence of solutions near a broken solution we must thus establish three things: a sufficiently good approximate solution w near the broken solution corresponding to 0 in Lemma 10.10, a right inverse for the linearization of the $\bar{\partial}_J$ -operator at w , corresponding to Q in Lemma 10.10, and a quadratic estimate for the non-linear term in the Taylor expansion, corresponding to (10.1). Here the Banach space B_1 will be a product of a weighted Sobolev space and a certain finite dimensional space that will serve as a neighborhood of the broken configuration and the Banach space B_2 will be a space of fields of complex antilinear maps. In addition to verifying uniform invertibility of the differential and the non-linear estimate we must also check that the gluing construction captures all solutions near the broken solution and that the natural change of coordinates (from the Banach space around the broken solution to the standard charts in the interior of the moduli space) is smooth.

10.3. GLUING CONSTANT DISKS. — The boundary strata of the moduli spaces we study involve splitting off of constant disks and splitting off of disks in the symplectization. In this section we consider gluing constant disks.

We first consider a configuration v as in Theorem 10.3 (a), 10.4 (a), or 10.6 (a1). In all these cases the broken configuration is a two level disk where the second level consists of a constant 3-punctured disk v^2 that is attached to the first level disk at a Lagrangian intersection puncture with asymptotic winding number 1. After we have carried out the gluing argument in this case we will discuss modifications needed for the other cases of constant disk gluing.

Assume that the first level disk v^1 has m Lagrangian intersection punctures. We take the domain of v^1 to be the standard domain $\Delta^1 = \Delta_m$. (As explained in Section 9.4, we may assume that the domain is stable by adding extra marked points near the positive puncture.) Recall that we defined a functional analytic neighborhood $W(a; \mathbf{n})$ of v^1 , where $W(a; \mathbf{n})$ is a product of an infinite dimensional weighted Sobolev space $H(a; \delta, \mathbf{n})$ and a finite dimensional space which is an open neighborhood B of the origin in $\mathbb{R}^{m-2} \times \mathbb{R} \times \mathbb{R}^m$, see Section 9.2. Here the first \mathbb{R}^{m-2} -component of an element in B corresponds to variations of the conformal structure of Δ^1 , the second \mathbb{R} -factor to shifts of the map in the symplectization direction near the positive puncture, and the last \mathbb{R}^m -factor corresponds to shifts along the knot near the Lagrangian intersection punctures. Here we will write W^1 for this neighborhood $W(a; \mathbf{n})$ and think of it as a product

$$W^1 = W_0^1 \times B^1$$

where B^1 is an open subset of \mathbb{R} , as follows. Let q denote the negative puncture where the second level disk is attached. Then B^1 corresponds to shifts along the knot at q .

Consider the negative puncture q at which the constant three punctured disk $v^2: \Delta^2 \rightarrow K$, where $\Delta^2 = \Delta_3$ is a standard domain with three punctures, is attached and fix a half-strip neighborhood $Q = (-1, 0] \times [0, 1]$ of it such that $v^1(Q)$ lies entirely in the standard neighborhood of K with complex analytic coordinates.

For $\rho > 0$, define a standard domain $\Delta_\rho \subset \Delta_{m+1}$ as follows. Remove the neighborhood $(-\rho, \rho) \cap [0, 1]$ of q from Δ^1 and the neighborhood $(\rho, 1) \cap [0, 1]$ of the positive puncture in Δ^2 , getting domains Δ_ρ^1 and Δ_ρ^2 . The domain Δ_ρ is then obtained by identifying the boundary segments $[-\rho, \rho] \cap [0, 1] \subset \Delta_\rho^1$ and $[\rho, 1] \cap [0, 1] \subset \Delta_\rho^2$. Then Δ_ρ contains the strip $Q_\rho = [-\rho, \rho] \cap [0, 1]$:

$$Q_\rho = \Delta_\rho \cap (\Delta_0^1 \cup \Delta_0^2).$$

We next define a pre-gluing $w_\rho: \Delta_\rho \rightarrow T^*Q$ (i.e., an approximate solution close to the broken disk v) and a neighborhood of it in a suitably weighted space of maps. We start with the map. Fix complex analytic coordinates $\mathbb{C} \subset \mathbb{C}^2$ around $p \in K$ on T^*Q , where p is the point where the constant disk v^2 sits. Let $\phi: \Delta_\rho \rightarrow \mathbb{C}$ be a smooth function which equals 1 on $\Delta_{\rho/2}^1$, equals 0 on Δ_ρ^2 and is real-valued and holomorphic on the boundary. (Holomorphic on the boundary just means that the restriction of $\bar{\partial}$ to the boundary vanishes. For example, if $s + it \in \mathbb{R} \cap [0, 1]$ are coordinates on the strip and $\phi(s)$ is an ordinary real valued cut-off function then a corresponding complex valued cut-off function that is holomorphic on the boundary is $\phi(s) + i\psi(t)d\phi/ds(s)$, where $\psi(t)$ is a small function with support near $\partial[0, 1]$ such that $\psi(0) = \psi(1) = 0$ and $d\psi/dt(0) = d\psi/dt(1) = 1$.)

Define

$$w_\rho(z) = \begin{cases} v^1(z), & z \in \Delta_{\rho/2}^1, \\ \phi(z)v^1(z) & z \notin \Delta_{\rho/2}^1, \end{cases}$$

where the last expression refers to the analytic coordinates around q corresponding to 0 in the coordinate system. Note then that w_ρ takes the boundary $\partial\Delta_\rho$ to $L = L_K \cup Q$ and that $\bar{\partial}_J w_\rho$ is supported in Q_ρ . Furthermore, using the Fourier expansion of v^1 near q ,

$$v^1(z) = \sum_{k>1} c_k e^{-k\pi z},$$

we find that

$$\bar{\partial}_J w_\rho|_{C^1} = O(e^{-\pi\rho}).$$

Define a weight function $\lambda_\rho: \Delta_\rho \rightarrow \mathbb{R}$ as follows, where $\eta_\delta: \Delta_{m+1} \rightarrow \mathbb{R}$ denotes the weight function on Δ^1 ,

$$\lambda_\rho(z) = \begin{cases} \eta_\delta(z) & \text{for } z \in \Delta_0^1, \\ e^{\delta(\rho - j\tau_j)} & \text{for } z \in Q_\rho = [-\rho, \rho] \cap [0, 1], \\ 1 & \text{for } z \in \Delta_0^2. \end{cases}$$

Let $\| \cdot \|_{k, \rho}$ denote the Sobolev norm with k derivatives on Δ_ρ and weight function λ_ρ . From the above we then find

$$(10.3) \quad \| \bar{\partial}_J w_\rho \|_{k, \rho} \leq \| \bar{\partial}_J w_\rho \|_{C^1} \int_\rho^\infty e^{\delta(\rho - j\tau_j)} d\tau = O(e^{-(\pi - \delta)\rho}).$$

We next define configuration spaces of maps giving neighborhoods of the approximate solutions w_ρ . As in Section 9.2 this space is a direct sum of an infinite dimensional

space and two finite dimensional summands. We first discuss the infinite dimensional summand.

Define $H_{2,\rho}(w_\rho)$ as the Sobolev space of vector fields v along w_ρ (i.e., sections of $w_\rho T(T \setminus Q) \setminus \Delta_\rho$) which satisfies the following requirements.

- If $\zeta \in \partial\Delta_\rho$ maps to L_K (maps to Q) then $v(\zeta)$ is tangent to L_K (resp. to Q).
- $r v + J \lrcorner v \lrcorner i = 0$ along $\partial\Delta_\rho$.
- Fix an endpoint $\zeta_0 \in \partial\Delta_\rho$ of the vertical segment which separates the part of Δ_ρ which corresponds to Δ^1 from that corresponding to Δ^2 . We require that $v(\zeta_0) = 0$.

Here the first two requirements have counterparts in Section 8.4 and the third is connected to the addition of certain cut-off solutions in the gluing region. We endow $H_{2,\rho}(w_\rho)$ with the weighted Sobolev 2-norm $\|\cdot\|_{k, k_{2,\rho}}$.

Second, we discuss the finite dimensional factor $B_\rho = B_\rho^1 \times B_\rho^2$. Here B_ρ^1 is an open neighborhood of the origin in $\mathbb{R}^{m-2} \times \mathbb{R} \times \mathbb{R}^{m-1}$ and agrees with the finite dimensional factor of W_0^1 in the following sense. The first \mathbb{R}^{m-2} -factor corresponds to the conformal variations of Δ_ρ inherited from Δ^1 , the second \mathbb{R} -factor corresponds to shifts at the positive puncture, and the last \mathbb{R}^{m-1} -factor corresponds to shifts along the knot K at Lagrangian intersection punctures that are also punctures of Δ^1 . The second factor B_ρ^2 is an open neighborhood of the origin in $\mathbb{R}^3 \times \mathbb{R}^2 \times \mathbb{R}$, where the first \mathbb{R}^3 -factor corresponds to constant vector fields supported in Q_ρ along the Lagrangian in a neighborhood of p that are cut-off in finite regions near the ends of Q_ρ , where the weight function λ_ρ is uniformly bounded and where the second factor corresponds to the shifts along K supported at the Lagrangian intersection punctures that are also punctures of Δ^2 . Finally, the third \mathbb{R} -factor is a newborn conformal variation defined as follows.

Consider the domain of the constant disks as a strip $\mathbb{R} \times [0, 1]$ with positive puncture at $+1$, one negative puncture at 0 , and one at -1 . Let v be the constant vector field ∂_τ and note that its flow moves the puncture at 0 and that in the standard model of the 3-punctured disk this vector field looks like $c_1 + O(e^{-\pi j \tau})$ at the puncture at $+1$ and at one of the punctures at -1 , whereas it looks like $c_2 e^{-\pi \tau} + O(1)$ at the other puncture at -1 , where $c_j, j = 1, 2$ are real constants. We extend this vector field v holomorphically over the gluing region Q_ρ and then cut it off using a cut-off function β with derivative supported near the end of Q_ρ that comes from Δ^1 where the weight function λ_ρ is close to 1. The conformal variation is then the complex anti-linear map $i\bar{\partial}(\beta v)$.

Note that the conformal variation in Δ_ρ that is inherited from the conformal variation at q in Δ^1 can be identified with the linear combination of conformal variations as above for the two punctures from Δ^2 which looks like $0 + O(e^{-\pi j \tau})$ at the positive puncture. We take the \mathbb{R} -factor to correspond to this variation. (Note that this conformal variation is supported in Δ_ρ^1 and agrees with the conformal variation in Δ^1 corresponding to the negative puncture q where the constant disk is attached.)

REMARK 10.11. — We note that there is a complementary linear combination of the two newborn conformal variations with non-zero leading constant term at the positive puncture of the constant disk that corresponds to the gluing parameter ρ which, from the point of view of the domain, shifts the boundary maximum between the two new punctures, see Remark 9.4.

Let $E_{1,\rho}$ denote the space of complex anti-linear maps $T\Delta_{m+2}^\rho \rightarrow w_\rho T(T^*Q)$, again weighted by λ_ρ . The linearization of the $\bar{\partial}_J$ -operator at w_ρ is then an operator

$$L\bar{\partial}_J: H_{2,\rho}(w_\rho) \rightarrow B_\rho \rightarrow E_{1,\rho},$$

LEMMA 10.12. — *The operators $L\bar{\partial}_J$ admit right inverses which are uniformly bounded as $\rho \rightarrow 1$.*

Proof. — The argument here is standard. Let k_1, \dots, k_l be a basis of the kernel K of the linearized operator on v^1 . Fix a cut-off function β which equals 1 on the part of Δ_ρ corresponding to Δ^1 and with first and second derivatives supported in Q_ρ of size $O(\rho^{-1})$. (Such a cut-off function exists since the length of Q_ρ equals 2ρ .) We will establish an estimate

$$(10.4) \quad \|v\|_{2,\rho} \leq C \|L\bar{\partial}_J v\|_{1,\rho},$$

where $C > 0$ is a constant, for v in the L^2 -complement of the subspace \tilde{K} spanned by the cut-off solutions $\beta k_1, \dots, \beta k_l$. The lemma follows from this estimate.

We argue by contradiction: assume that the estimate does not hold. Then there is a sequence v_ρ in this L^2 -complement with

$$(10.5) \quad \|v_\rho\|_{2,\rho} = 1, \quad \text{and} \quad \|L\bar{\partial}_J v_\rho\|_{1,\rho} \rightarrow 0.$$

We write $v_\rho = u_\rho + b_\rho^1 + b_\rho^2$, where $u_\rho \in H_{2,\rho}(w_\rho)$, $b_\rho^1 \in B_\rho^1$, and $b_\rho^2 \in B_\rho^2$. Fix cut-off functions β_j on Δ_ρ , $j = 1, 2$ with the following properties. The function β_1 equals 1 on $\Delta_{\rho/2}^1 \subset \Delta_\rho$ and equals 0 on $\Delta_\rho^2 \setminus \Delta_\rho$. Furthermore, $\beta_1 u_\rho$ is holomorphic on the boundary, and $jD\beta_1 = O(\rho^{-1})$. The function β_2 has similar properties but with support in $\Delta_\rho^2 \subset \Delta_\rho$. We also let α be a similar cut-off function on Δ_ρ , equal to 1 on $Q_{\rho/2}$ and equal to 0 outside Q_ρ .

Since $\|j\bar{\partial}\beta_j^p\|_{1,\rho} \rightarrow 0$ as $\rho \rightarrow 0$ we then have

$$\begin{aligned} \left\| L\bar{\partial}_J(\beta_1 u_\rho + b_\rho^1 + b_\rho^2) \right\|_{1,\rho} &\leq \|j\bar{\partial}\beta_1^p\|_{1,\rho} \|u_\rho\|_{2,\rho} + \left\| L\bar{\partial}_J(u_\rho + b_\rho^1 + b_\rho^2) \right\|_{1,\rho} \\ &\leq \|j\bar{\partial}\beta_1^p\|_{1,\rho} \|u_\rho\|_{2,\rho} + \|L\bar{\partial}_J v_\rho\|_{1,\rho} \rightarrow 0, \end{aligned}$$

as $\rho \rightarrow 1$. We then conclude from transversality of v^1 (i.e., invertibility of the linearized operator on its kernel) that there exists a constant $M > 0$ such that

$$(10.6) \quad \left\| \beta_1 u_\rho + b_\rho^1 + b_\rho^2 \right\|_{2,\rho} \leq M \left\| L\bar{\partial}_J(\beta_1 u_\rho + b_\rho^1 + b_\rho^2) \right\|_{1,\rho} \rightarrow 0.$$

In particular the cut-off constant solution in the gluing region goes to 0.

Similarly we have

$$\left\| L\bar{\partial}_J(\beta_2 u_\rho + b_\rho^2) \right\|_{1,\rho} \rightarrow 0.$$

We conclude from the invertibility of the standard operator on the three punctured disk that

$$(10.7) \quad \left\| \beta_2 u_\rho + b_\rho^2 j \right\|_{2,\rho} \neq 0.$$

After dividing the weight function in the gluing region $Q_{\rho/2} \subset [\frac{\rho}{2}, \frac{\rho}{2}] \times [0, 1]$ by its maximum the problem on the gluing region converges to the $\bar{\partial}$ -problem on the strip with \mathbb{R}^3 -boundary condition and negative exponential weights at both ends (i.e., weight function $\rho(s + it) = e^{-\delta|s|}$). This problem has a three-dimensional kernel spanned by constant solutions in \mathbb{R}^3 . As mentioned above, the estimates (10.6) and (10.7) imply that the components along the constant solutions go to zero. This gives first that

$$\|L\bar{\partial}_J(\alpha u_\rho)\|_{1,\rho} \neq 0,$$

and then, by invertibility of $L\bar{\partial}_J$ on the complement of the kernel, also that $\|k\alpha u_\rho\|_{2,\rho} \neq 0$. Our assumption thus implies that $\|kv_\rho\|_{2,\rho} \neq 0$. This contradicts (10.5). The lemma follows.

The next thing to establish is the quadratic estimate for the non-linear term in the Taylor expansion of $\bar{\partial}_J$ around w_ρ , i.e., around the origin in $H_{2,\rho} \subset B_\rho$. We use the exponential map as in Section 9.2 to define the local coordinate system around w_ρ and the estimate for the non-linear term follows from a standard argument that uses the uniform bounds on the derivatives of the exponential map in our metric, see [18, Lem. A.18] and also [13, 15]. In fact the standard argument gives the corresponding unweighted estimate but then the case of positive weights follows since the left hand side of the inequality is linear in the weight whereas the right hand side is quadratic. So the inequality follows for weights bounded from below. Note also that variations along the cut-off solutions in B_ρ give contributions to the non-linear term only in the regions where the derivatives of the cut-off functions are supported and in such regions the weight functions have finite size.

REMARK 10.13. — It is essential here that the cut-off solutions are real solutions to the non-linear equation since a small error term would give a large norm contribution because of the large weight function in N_ρ , which in turn is key for the proof of the uniform invertibility of the differential in Lemma 10.12.

The final step is then to show surjectivity of the construction. More concretely, this means that we must show that any sequence of disks which converges in the sense of Section 8.5 to a broken disk eventually lies in a small k - $k_{2,\rho}$ -neighborhood of w_ρ . This follows once we show that any holomorphic disk in a C^0 -neighborhood of the approximate solution is also close in k - $k_{2,\rho}$ -norm. The proof of that fact follows from the knowledge of explicit solutions in the region where the weight is big. Here C^0 -control at the ends gives norm control, see [18, Proof of Th. A.21] or [9, Proof of Th. 1.3]. This finishes the gluing results needed in the cases when we glue one constant 3-punctured disks at a Lagrangian intersection puncture of winding number 1.

The remaining cases for gluing constant disks are proved by modifications of the above argument that we describe next. Consider first Theorem 10.6 (a2). Here we replace the gluing parameter ρ with two independent gluing parameters $(\rho_1, \rho_2) \in [0, 1]^2$, one for each constant disk. Likewise we have two copies of the new finite dimensional factors in the configuration space. The gluing argument is then a word by word repetition of the above.

Next consider broken disks as in Theorem 10.4(c). Here the exponential weight at the winding $\frac{3}{2}$ -puncture of v^1 is $\delta \in (\pi/2, \pi)$ and the boundary condition in the strip Q_ρ has different constant Lagrangians along the two boundary components. The cut-off solutions in B_ρ^2 change accordingly: instead of an \mathbb{R}^3 -factor of cut-off solutions we have an \mathbb{R}^5 -factor, $\mathbb{R}^5 = \mathbb{R} \times \mathbb{R}^2 \times \mathbb{R}^2$. The \mathbb{R} -factor is a constant solution in the direction of the knot. The first \mathbb{R}^2 -factor contains cut-off solutions near the positive puncture of Δ_ρ^2 of the form $ce^{\pi z/2}$ for c a vector in the appropriate Lagrangian 2-space perpendicular to the knot, the second \mathbb{R}^2 -factor consists of cut-off solutions of the form $ce^{-\pi z/2}$. Then in Lemma 10.12 we replace (10.7) with the estimate on the three punctured disk with boundary condition corresponding to the constant disk. I.e., in directions perpendicular to the knot the boundary condition are two perpendicular Lagrangian planes at the two boundary components near the positive puncture and one of these planes between the two negative punctures. There is a small positive exponential weight at the negative punctures, the weight δ and two cut-off solutions at the positive puncture. In the directions perpendicular to the knot the $\bar{\partial}$ -operator is then an isomorphism and the argument above proceeds as before.

REMARK 10.14. — In Theorem 10.3 (c) there are two different constant disks and the corresponding boundary points cancel out. Geometrically this corresponds to pushing a winding $\frac{1}{2}$ puncture through a winding 1 puncture.

Finally, we consider Theorem 10.6 (d2). The argument here is the same as that just described for Theorem 10.4 (c) with the only difference being that the 3-punctured constant disk should be replaced by a 4-punctured disk and that we invert the operator on the L^2 complement of the additional conformal variation in the 4-punctured disk. In fact, when the 4-punctured disk is broken into two levels it corresponds to the 3-level configuration with the two top levels as in Theorem 10.4 (c) and a third level constant disk attached at the winding 1 puncture of the second level constant disk.

10.4. SYMPLECTIZATION GLUING. — Consider a disk with two non-constant levels as in Theorem 10.6 (b) or (c), Theorem 10.3 (b) or (c), or Theorem 10.4 (b). The argument needed to glue such configurations is similar to the one in Section 10.3 and we only sketch the details. There are again four steps: define an approximate solution, prove uniform invertibility of the differential, establish a quadratic estimate for the non-linear term, and show surjectivity of the construction.

We consider first the case when we glue a symplectization disk to a disk in T^*Q and discuss modifications needed when the second level also lives in the symplectization later. Denote the top-level disk in the symplectization $v^1: \Delta^1 = \Delta_m \times \mathbb{R} \times S^1 \times Q$

and the m second level disks $v^{2,j}: \Delta_{m_j} \rightarrow T^*Q$, $j = 1, \dots, m$. Recall that by adding marked points we reduce to the case when all domains involved are stable, see Section 9.4.

Each symplectization disk lies in a natural \mathbb{R} -family. Let t denote a standard coordinate on the \mathbb{R} -factor. Fix the unique map v^1 in this family that takes the largest boundary maximum in Δ^1 to the slice $\hbar t = 0g$. By asymptotics at the negative punctures, for all $T > 0$ sufficiently large $(v^1)^{-1}(\hbar t \in Tg)$ consists of m half strip regions with one component around each negative puncture of v^1 . Furthermore, as $T \rightarrow \infty$ the inverse image of the slice $\hbar t = Tg$ converges to vertical segments at an exponential rate (since the map agrees with trivial Reeb chord strips up to exponential error). We fix such a slice and consider the vertical segments through its end point. Parameterize the neighborhoods of all the punctures cut at these vertical segments by $(-\tau, 0] \cap [0, 1]$. For $\rho > 0$, let $\Delta_\rho^1 \subset \Delta^1$ be the subset obtained by removing $(-\tau, -\rho) \cap [0, 1]$ from the neighborhood $(-\tau, 0] \cap [0, 1]$ of each negative puncture.

Fix neighborhoods $[0, \tau) \cap [0, 1]$ of the positive puncture in each $\Delta^{2,j}$, $j = 1, \dots, m$ in which the map is well approximated by the trivial strip at the positive puncture and let $\Delta_\rho^{2,j} \subset \Delta^{2,j}$ denote the subset obtained by removing $(\rho, \tau) \cap [0, 1]$ from this neighborhood. Let Δ_ρ denote the domain obtained by adjoining $\Delta_\rho^{2,j}$ to Δ_ρ^1 by identifying the vertical segment at the positive puncture of $\Delta_\rho^{2,j}$ with the vertical segment of the negative puncture in Δ_ρ^1 where $v^{2,j}$ is attached to v^1 . Then we get m strip regions $Q_\rho^j = [-\rho, \rho] \cap [0, 1] \cup \Delta_\rho$ around each vertical segment where the disks were joined.

By interpolating between the two maps joined at each negative puncture using the standard coordinates near the Reeb chords we find a pregluing

$$w_\rho: \Delta_\rho \rightarrow T^*Q$$

such that $\bar{\partial}_J w_\rho$ is supported only in the middle $[-1, 1] \cap [0, 1]$ of each Q_ρ^j and such that

$$\bar{\partial}_J w_\rho|_{C^1} = O(e^{-\alpha\rho}),$$

where $\alpha > 0$ depends on the angle between the Lagrangian subspaces of the contact hyperplane obtained by moving the tangent space of Λ_K at the Reeb chord start point to the tangent space of Λ_K at the Reeb chord end point by the linearized Reeb flow.

As in Section 10.3 we use a configuration space of maps in a neighborhood of w_ρ that is a product of an infinite and a finite dimensional space of functions. We first consider the infinite dimensional factor. Define weight functions $\lambda_\rho: \Delta_\rho \rightarrow \mathbb{R}$ by patching (suitably scaled) weight functions η_δ of the domains of the broken disks where we take $0 < \delta < \alpha$. In particular, we have $\lambda_\rho(\tau + it) = c_j e^{\delta j \tau}$ for $\tau + it \in Q_\rho^j$. Then, writing $\|\cdot\|_{k, \rho}$ for the Sobolev k -norm with this weight, we have

$$\|\bar{\partial}_J w_\rho\|_{k, \rho} = O(e^{(\delta - \alpha)\rho}).$$

We let $H_{2, \rho}(w_\rho)$ denote the λ_ρ -weighted Sobolev space of vector fields along w_ρ which are tangent to the Lagrangians, holomorphic on the boundary, and which satisfy the following vanishing condition. The map w_ρ maps the strip regions Q_ρ^j into

small neighborhoods of the Reeb chord strips where we have standard coordinates $\mathbb{R} \times (\varepsilon, L + \varepsilon) \subset \mathbb{C}^2$ and we require that the \mathbb{R} -component of the vector field vanishes at one of the endpoints of the vertical segments where the disks were joined. Thus there are in total m vanishing conditions.

Next we discuss the finite dimensional factor $B_\rho = B_\rho^0 \times B_\rho^1 \times B_\rho^2$. The second factor B_ρ^1 is an open subset of the origin in \mathbb{R} corresponding to the shift at the positive puncture of w_ρ . The third factor B_ρ^2 contains all the conformal variations and the shifts inherited from the negative punctures of the second level disks. Thus B_ρ^2 is a neighborhood of the origin in

$$\prod_{j=1}^m (\mathbb{R}^{m_j - 2} \times \mathbb{R}^{m_j}).$$

Finally, the first factor B_ρ^0 is an open subset of the origin in a codimension one subspace of

$$(\mathbb{R} \times \mathbb{R}^2)^m,$$

where each $(\mathbb{R} \times \mathbb{R}^2)$ -factor corresponds to a specific second level disk. The \mathbb{R} -component of the j^{th} puncture of v^1 corresponds to a cut-off shifting vector field a_j in the \mathbb{R} -direction of the symplectization supported in Q_ρ^j . The \mathbb{R}^2 -component corresponds to the two newborn conformal variations in $\Delta_\rho^{2,j}$. As before these conformal variations have the form $\gamma = \bar{\partial}V$ where V is a vector field along Δ_ρ . The first factor of \mathbb{R}^2 corresponds to a variation γ_1^j that agrees with the conformal variation at the negative puncture in Δ^1 where $v^{2,j}$ is attached. The second factor is spanned by $\gamma^{2,j} = \bar{\partial}V_2$ where V_2 is the vector field in $\Delta_\rho^{2,j} \setminus Q_\rho^j$ that corresponds to translations along the real axis that moves all the boundary maxima in $\Delta_\rho^{2,j}$ cut off near the end of Q_ρ in Δ_ρ^1 . The codimension one subspace is the orthogonal complement of the line given by the equation

$$\gamma^{2,1} = \gamma^{2,2} = \dots = \gamma^{2,m}.$$

Note that this later conformal variation corresponds to changing ρ .

REMARK 10.15. — The nature of the conformal variations $\gamma^{1,j}$ and $\gamma^{2,j}$ are easy to see using a different conformal model for the domain Δ_ρ as follows. Consider the domain of Δ^1 as the upper half plane H with positive puncture at 1 and negative punctures along the real axis. The conformal variations of this domain can be viewed as translating the negative punctures along the real axis. To construct the domain Δ_ρ we think also of the domains $\Delta^{2,j}$ as upper half planes. Cut out small half disks of radius $c_j e^{-\alpha\rho}$ near the negative punctures of Δ^1 and glue in the half disks in the domain $\Delta^{2,j}$ of radius $c_j e^{\alpha\rho}$ scaled by $e^{-2\alpha\rho}$. Now the conformal variation $\gamma^{1,j}$ corresponds to translating the whole half disk at the j^{th} negative puncture of Δ^1 rigidly in the real direction and the conformal variation $\gamma^{2,j}$ corresponds to keeping the small half disk fixed but scaling it so that its negative punctures move closer together.

We use the neighborhood $W_\rho = H_{2,\rho} \times B_\rho$ of w_ρ . In order to apply Lemma 10.10 we must first establish the counterpart of Lemma 10.12. Here we invert the linearized

operator on the L^2 -complement of the subspace spanned by cut-off kernel elements in Δ^1 and $\Delta^{2,j}$ defined as follows. The infinite dimensional components are indeed just a cut-off vector field. For the finite dimensional components we identify the conformal variation at the j^{th} negative puncture of Δ^1 with $\gamma^{1,j}$, the shift at this negative puncture with a_j , and the shift at the positive puncture of $\Delta^{2,j}$ with $\gamma^{2,j}$. To show uniform invertibility we then argue by contradiction as in the proof of Lemma 10.12. Using the above identifications of finite dimensional factors, the result follows in a straightforward way.

Finally, the two remaining steps, the quadratic estimate for the non-linear term and the surjectivity of the construction are completely analogous to their counterparts in Section 10.3 and will not be discussed further.

In the case that the second level disk lies in the symplectization as well we start as above by fixing a representative for v^1 and a slice $\tilde{t} = Tg$ after which this representative is well approximated by Reeb chord strips. We then fix representatives for all the non-trivial second level curves $v^{2,j}$ (of which there is only one in our case) that are translated sufficiently much so that they are well approximated by Reeb chord strips in the slice $\tilde{t} = Tg$ at their positive punctures. We then repeat the argument above.

10.5. POINT CONSTRAINTS ON THE KNOT. — An analogous construction allows us to express neighborhoods of disks with Lagrangian intersection punctures of winding number 1 inside the space of disks with these punctures removed. In the analytical $\mathbb{C} \times \mathbb{C}^2$ -coordinates around the knot a disk v with such a puncture looks like

$$v(z) = \sum_{n>0} c_n e^{-n\pi z}, \quad z \in [0, 1) \cup [0, 1], \quad c_n \in \mathbb{R}^3 \text{ or } c_n \in \mathbb{R} \times i\mathbb{R}^2$$

with $c_0 = (c_0^0, 0)$, whereas a general disk looks the same way but has unrestricted c_0 . We can thus construct a configuration space W for unrestricted disks in a neighborhood of v as

$$W = W^\theta \times \mathbb{R}^2,$$

where W^θ is the configuration space for disks in a neighborhood of v with Lagrangian intersection puncture of winding number 1 and \mathbb{R}^2 is spanned by two cut-off constant solutions in the Lagrangian perpendicular to K . The zero-set of the $\bar{\partial}_J$ -operator acting on W then gives a neighborhood of v in the space of unrestricted disks.

10.6. PROOFS OF THE STRUCTURE THEOREMS. — The proof of all the theorems on the structure of the compactified moduli spaces as manifolds with boundary with corners now follow the same pattern. Transversality and compactness results give the possible degenerations and gluing give neighborhoods of several level disks in the boundary. The manifold structure in the interior is a consequence of standard Fredholm theory, whereas charts near the boundary are obtained from the conformal structures of the domains.

Proof of Theorem 10.1. — Part (i) follows immediately from Lemma 9.5 and Theorem 8.12. Consider part (ii). Lemma 9.5 and Theorem 8.12 imply that the broken disks listed are the only possible configurations in the boundary of the compactified moduli space. It follows from (the parameterized version of) Lemma 10.10 that the gluing parameter gives a parameterization of the boundary of the reduced moduli space. Recall that we identified the gluing parameter with a certain conformal variation (that shifts all the boundary maxima in the second level disk) and we topologize a neighborhood of the broken configuration using the induced map to the compactified space of conformal structures. This establishes (ii).

Proof of Theorem 10.2. — The theorem follows immediately from Lemma 9.5 and Theorem 8.12.

Proof of Theorem 10.3. — The proof is analogous to the proof of Theorem 10.1 (ii) except for (c). Here a disk without Lagrangian intersection punctures moves out as a rigid disk in the symplectization into the \mathbb{R} -invariant region and the translations along \mathbb{R} give a neighborhood of the boundary.

Proof of Theorem 10.4. — The argument is analogous to the proofs above and we explain only how to parameterize the boundary in the cases that differ from the above. Consider (b). Recall that we identified the gluing parameter with the conformal variation that translates all the boundary maxima in the second level disks uniformly. As above we use this to parameterize a neighborhood of the boundary. Finally, consider (c). Here again the boundary can be parameterized by the gluing parameter which corresponds to a conformal variation. In particular, the boundary point corresponds to a three punctured disk splitting σ . As explained in Remark 10.14 there are two such disks and the corresponding boundaries of the moduli space naturally fit together to a smooth 1-manifold.

REMARK 10.16 (cf. Remark 10.5). — Consider a holomorphic disk near the codimension one boundary as in Theorem 10.4 (c). Remark 8.13 gives a local model (4.1) for the disk, parameterized by a half disk in the upper half plane near the two colliding corners with one puncture at 0 and one at $\varepsilon > 0$. The above proof shows that the newborn conformal variation which here is the length of the stretching strip can be used as local coordinate in the moduli space near the corner. A conformal map that takes a vertical segment in the stretching strip to the upper arc in the unit circle and the boundary of the domain in the disk splitting σ to the real line gives a smooth change of coordinates from this parameter to the coordinates given by ε . Thus the local model (4.1) used in the definition of the string operations is C^k -close to the actual moduli space, when both are viewed as parameterized by the coordinates ε . A similar discussion applies to Theorem 10.4 (c), using the local model (4.2) with $\delta = 0$.

Proof of Theorem 10.6. — Arguments for producing neighborhoods of codimension one boundary strata are similar to the above, so we discuss the codimension two parts.

Consider a broken disk as in (a2). The gluing result needed in this case is analogous to the argument in Section 10.3. Here however we attach two constant disks, producing approximate solutions w_{ρ_1, ρ_2} depending on two independent variables $\rho_1, \rho_2 \in \mathbb{R}$. In this case there are two independent newborn conformal variations and the linearized $\bar{\partial}_J$ -operator is inverted on the complement of their linear span. It follows as above that the projection of the moduli space is an embedding into the space of conformal structures and we induce the corner structure from there. Note that this is coherent with our treatment of nearby codimension one boundary disks.

The arguments in cases (b2), (c2), and (d2) follow the same lines. We produce approximate solutions depending on two independent variables. In case (b2) the linearized operator is inverted on the complement of the 2-dimensional spaces spanned by the cut-off shift of the symplectization disk and the newborn conformal structure of the constant disk. In case (c2) the linearized operator is inverted on the complement of the (independent) shifts of the first and second level disks, and in case (d2) on the complement of the newborn conformal structure and the additional conformal structure in the constant 4-punctured disk. In all cases, the corner structure is induced from the corresponding structure on the space of conformal structures and the construction is compatible with nearby strata of lower codimension.

REMARK 10.17 (cf. Remark 10.7). — Consider a holomorphic disk near the codimension two corner as in Theorem 10.6 (d2). Remark 8.13 gives a local model (4.2) for the disk, parameterized by a half disk in the upper half plane near the three colliding corners with one puncture at 0 and the two others at boundary points $\delta < 0$ and $\varepsilon > 0$. The above proof shows that the newborn conformal variation (which here is the length of the stretching strip) together with the difference between the boundary maxima in the 4-punctured disk splitting-off can be used as local coordinates in the moduli space near the corner. A conformal map that takes a vertical segment in the stretching strip to the upper arc in the unit circle and the boundary of the domain in the disk splitting-off to the real line gives a smooth change of coordinates from these two parameters to the coordinates given by (ε, δ) . Thus the local model (4.2) used in the definition of the string operations is C^k -close to the actual moduli space, when both are viewed as parameterized by the coordinates (ε, δ) .

Proof of Theorem 10.8. — The theorem follows from the discussion in Section 10.5.

Proof of Theorem 10.9. — The theorem follows from the discussion in Section 10.5 in combination with the argument in the proof of Theorem 10.4 (c).

REFERENCES

- [1] M. AGANAGIC, T. EKHOLM, L. NG & C. VAFA – “Topological strings, D-model, and knot contact homology”, *Adv. Theo. Math. Phys.* **18** (2014), no. 4, p. 827–956.
- [2] S. BASU, J. MCGIBBON, D. SULLIVAN & M. SULLIVAN – “Transverse string topology and the cord algebra”, *J. Symplectic Geom.* **13** (2015), no. 1, p. 1–16.
- [3] F. BOURGEOIS, Y. ELIASHBERG, H. HOFER, K. WYSOCKI & E. ZEHNDER – “Compactness results in symplectic field theory”, *Geom. Topol.* **7** (2003), p. 799–888.
- [4] M. CHAS & D. SULLIVAN – “String topology”, arXiv: math.GT/9911159, 1999.
- [5] K. CIELIEBAK, T. EKHOLM & J. LATSCHEV – “Compactness for holomorphic curves with switching Lagrangian boundary conditions”, *J. Symplectic Geom.* **8** (2010), no. 3, p. 267–298.
- [6] K. CIELIEBAK & J. LATSCHEV – “The role of string topology in symplectic field theory”, in *New perspectives and challenges in symplectic field theory*, CRM Proc. Lecture Notes, vol. 49, American Mathematical Society, Providence, RI, 2009, p. 113–146.
- [7] G. DIMITROGLOU RIZELL – “Lifting pseudo-holomorphic polygons to the symplectisation of $P \times \mathbb{R}$ and applications”, *Quantum Topol.* **7** (2016), no. 1, p. 29–105.
- [8] N. M. DUNFIELD & S. GAROUFALIDIS – “Non-triviality of the A -polynomial for knots in S^3 ”, *Algebraic Geom. Topol.* **4** (2004), p. 1145–1153.
- [9] T. EKHOLM – “Morse flow trees and Legendrian contact homology in 1-jet spaces”, *Geom. Topol.* **11** (2007), p. 1083–1224.
- [10] ———, “Rational symplectic field theory over \mathbb{Z}_2 for exact Lagrangian cobordisms”, *J. Eur. Math. Soc. (JEMS)* **10** (2008), no. 3, p. 641–704.
- [11] T. EKHOLM, J. ETNYRE, L. NG & M. SULLIVAN – “Filtrations on the knot contact homology of transverse knots”, *Math. Ann.* **355** (2013), no. 4, p. 1561–1591.
- [12] ———, “Knot contact homology”, *Geom. Topol.* **17** (2013), no. 2, p. 975–1112.
- [13] T. EKHOLM, J. ETNYRE & M. SULLIVAN – “The contact homology of Legendrian submanifolds in \mathbb{R}^{2n+1} ”, *J. Differential Geom.* **71** (2005), no. 2, p. 177–305.
- [14] ———, “Orientations in Legendrian contact homology and exact Lagrangian immersions”, *Internat. J. Math.* **16** (2005), no. 5, p. 453–532.
- [15] ———, “Legendrian contact homology in $P \times \mathbb{R}$ ”, *Trans. Amer. Math. Soc.* **359** (2007), no. 7, p. 3301–3335.
- [16] T. EKHOLM & T. KÁLMÁN – “Isotopies of Legendrian 1-knots and Legendrian 2-tori”, *J. Symplectic Geom.* **6** (2008), no. 4, p. 407–460.
- [17] T. EKHOLM, L. NG & V. SHENDE – “A complete knot invariant from contact homology”, arXiv: 1606.07050.
- [18] T. EKHOLM & I. SMITH – “Exact Lagrangian immersions with one double point revisited”, *Math. Ann.* **358** (2014), no. 1-2, p. 195–240.
- [19] ———, “Exact Lagrangian immersions with a single double point”, *J. Amer. Math. Soc.* **29** (2016), no. 1, p. 1–59.
- [20] Y. ELIASHBERG, A. GIVENTAL & H. HOFER – “Introduction to symplectic field theory”, *Geom. Funct. Anal.* (2000), p. 560–673, Special Volume, Part II, GAFA 2000 (Tel Aviv, 1999).
- [21] A. FLOER – “Morse theory for Lagrangian intersections”, *J. Differential Geom.* **28** (1988), no. 3, p. 513–547.
- [22] K. FUKAYA, Y.-G. OH, H. OHTA & K. ONO – *Lagrangian intersection Floer theory: anomaly and obstruction. Part I*, AMS/IP Studies in Advanced Mathematics, vol. 46, American Mathematical Society, Providence, RI, 2009.
- [23] C. GORDON & T. LIDMAN – “Knot contact homology detects cabled, composite, and torus knots”, arXiv: 1509.01642.
- [24] C. GORDON & J. LUECKE – “Knots are determined by their complements”, *J. Amer. Math. Soc.* **2** (1989), no. 2, p. 371–415.
- [25] M. W. HIRSCH – *Differential topology*, Graduate Texts in Math., vol. 33, Springer-Verlag, New York-Heidelberg, 1976.
- [26] H. HOFER, C. H. TAUBES, A. WEINSTEIN & E. ZEHNDER (eds.) – *The Floer memorial volume*, Progress in Mathematics, vol. 133, Birkhäuser Verlag, Basel, 1995.
- [27] H. HOFER, K. WYSOCKI & E. ZEHNDER – “A general Fredholm theory. I. A splicing-based differential geometry”, *J. Eur. Math. Soc. (JEMS)* **9** (2007), no. 4, p. 841–876.

- [28] P. B. KRONHEIMER & T. S. MROWKA – “Dehn surgery, the fundamental group and $SU(2)$ ”, *Math. Res. Lett.* **11** (2004), no. 5-6, p. 741–754.
- [29] R. H. LAGRANGE & A. H. RHEMTULLA – “A remark on the group rings of order preserving permutation groups.”, *Canad. Math. Bull.* **11** (1968), p. 679–680.
- [30] L. NG – “Knot and braid invariants from contact homology. I”, *Geom. Topol.* **9** (2005), p. 247–297.
- [31] ———, “Knot and braid invariants from contact homology. II”, *Geom. Topol.* **9** (2005), p. 1603–1637, With an appendix by the author and Siddhartha Gadgil.
- [32] ———, “Framed knot contact homology”, *Duke Math. J.* **141** (2008), no. 2, p. 365–406.
- [33] ———, “Combinatorial knot contact homology and transverse knots”, *Adv. Math.* **227** (2011), no. 6, p. 2189–2219.
- [34] ———, “A topological introduction to knot contact homology”, in *Contact and symplectic topology*, Bolyai Soc. Math. Stud., vol. 26, János Bolyai Math. Soc., Budapest, 2014, p. 485–530.
- [35] V. SHENDE – “The conormal torus is a complete knot invariant”, arXiv: 1604.03520.
- [36] D. SULLIVAN – “Open and closed string field theory interpreted in classical algebraic topology”, in *Topology, geometry and quantum field theory*, London Math. Soc. Lecture Note Ser., vol. 308, Cambridge Univ. Press, Cambridge, 2004, p. 344–357.
- [37] ———, “String topology background and present state”, in *Current developments in mathematics, 2005*, Int. Press, Somerville, MA, 2007, p. 41–88.

Manuscript received February 5, 2016

accepted May 23, 2017

KAI CIELIEBAK, Institut für Mathematik, Universität Augsburg
86135 Augsburg, Germany

E-mail : kai.cieliebak@math.uni-augsburg.de

Url : <https://www.math.uni-augsburg.de/prof/geo/mitarbeiter/cieliebak/>

TOBIAS EKHOLM, Department of Mathematics, Uppsala University
751 06 Uppsala, Sweden

and

Institut Mittag-Leffler

Aurav 17, 182 60 Djursholm, Sweden

E-mail : Tobias.Ekholm@math.uu.se

JANKO LATSCHEV, Universität Hamburg, Fachbereich Mathematik
Bundesstraße 55, 20146 Hamburg, Germany

E-mail : janko.latschev@uni-hamburg.de

Url : <https://www.math.uni-hamburg.de/home/latschev/>

LENHARD NG, Department of Mathematics, Duke University
Durham, NC 27708-0320, USA

E-mail : ng@math.duke.edu

Url : <http://alum.mit.edu/www/ng>