Low Dimensional Topology

Brandon Williams
mbw@math.sunysb.edu

November 25, 2008

Contents

1 Transversality and Intersection Theory

1.1 Immersions, Submersions and Transversality

Let $f : M \to N$ be a smooth map between smooth manifolds. A point $p \in M$ is said to be **regular** if $df_p : T_pM \to T_{f(p)}N$ is surjective, otherwise it is said to be **critical**. The image of a critical point is called a **critical value**, and a **regular value** in N is any point that is not a critical value. We say that f is an **immersion** at $p \in M$ if df_p is injective, and a **submersion** at $p \in M$ if df_p is surjective. Recall the inverse function theorem and its corollaries pertaining to immersions and submersions.

Theorem 1.1 (Inverse Function Theorem). *If $df_p : T_pM \to T_{f(p)}N$ is an isomorphism, then f is a local diffeomorphism at p, i.e. there is a neighborhood U of p and V of $f(p)$ such that $f|_U : U \to V$ is a diffeomorphism.*

Corollary 1.1. *If $f : M^m \to N^n$ is an immersion at $p \in M$, then there are coordinates $\phi : U \to \mathbb{R}^m$ around $p \in M$ and $\varphi : V \to \mathbb{R}^n$ around $f(p) \in N$ such that $\varphi \circ f \circ \phi^{-1}(x_1, \ldots, x_m) = (x_1, \ldots, x_m, 0, \ldots, 0)$.*

Proof. First let ϕ and φ be any charts such that $\phi(p) = \varphi(p) = 0$, and let $g = \varphi \circ f \circ \phi^{-1}$. Then $dg_0 : R^m \to R^n$ is injective since f is an immersion, so we can change the coordinates so that

$$dg_0 = \begin{pmatrix} I_m \\ 0 \end{pmatrix}$$

where I_m is the $m \times m$ identity matrix and 0 is the $(n - m) \times m$ zero matrix. We cannot apply the inverse function theorem to g since $m < n$, but if we define $G : \phi(U) \times R^{n-m} \to R^n$ by $G(x, z) = g(x) + (0, z)$, then $dG_0 = I_n$. So G is a local diffeomorphism, and $G \circ \varphi$ is another chart around $f(p)$ of N. In these coordinates we have that the local representation of f is just $(x_1, \ldots, x_m) \mapsto (x_1, \ldots, x_m, 0, \ldots, 0)$. $\square$

Corollary 1.2. *If $f : M^n \to N^m$ is an submersion at $p \in M$, then there are coordinates $\phi : U \to \mathbb{R}^n$ around $p \in M$ and $\varphi : V \to \mathbb{R}^m$ around $f(p) \in N$ such that $\varphi \circ f \circ \phi^{-1}(x_1, \ldots, x_n) = (x_1, \ldots, x_m)$.*

Proof. First let ϕ and φ be any charts such that $\phi(p) = \varphi(p) = 0$, and let $g = \varphi \circ f \circ \phi^{-1}$. Then $dg_0 : R^n \to R^m$ is injective since f is a submersion, so we can change the coordinates so that

$$dg_0 = \begin{pmatrix} I_m & 0 \end{pmatrix}$$

where I_m is the $m \times m$ identity matrix and 0 is the $m \times (n - m)$ zero matrix. We cannot apply the inverse function theorem to g since $m < n$, but if we define $G : R^n \to \varphi(V) \times R^{n-m}$ by $G(x_1, \ldots, x_n) = (g(x), x_{m+1}, \ldots, x_n)$, then $dG_0 = I_n$. So G is a local diffeomorphism, and $G^{-1} \circ \phi$ is another chart around p of M. In these coordinates we have that the local representation of f is just $(x_1, \ldots, x_n) \mapsto (x_1, \ldots, x_m)$. $\square$

Corollary 1.3. *If q is a regular value of $f : M^n \to N^m$, then $f^{-1}(q)$ is a submanifold of M with $\mathrm{codim}\, f^{-1}(q) = m$.*

Proof. Let $p \in f^{-1}(q)$, then f is a submersion at p, so by corollary 1.2 we can find coordinates $\phi : U \to \mathbb{R}^n$ around p and $\varphi : V \to \mathbb{R}^m$ around q such that $\varphi \circ f \circ \phi^{-1}(x_1, \ldots, x_n) = (x_1, \ldots, x_m)$ and $\varphi(q) = (0, \ldots, 0)$. Thus for $x \in U$ and $f(x) = q$ we have $\phi(x) = (0, \ldots, 0, x_{m+1}, \ldots, x_n)$ for some $x_{m+1}, \ldots, x_n$. So ϕ restricted to $U \cap f^{-1}(q)$ provides a chart of the relatively open set $U \cap f^{-1}(q)$ into $\mathbb{R}^{n-m}$, making $f^{-1}(q)$ into a submanifold of dimension $n - m$. $\square$

Consider the following simple application of corollary 1.3. For real-valued functions $f_1, \ldots, f_k :$ $M \to \mathbb{R}$, when is the set Z of common zeros a manifold? If we set $f = (f_1, \ldots, f_k) : M \to \mathbb{R}^k$, then it is clear that $f^{-1}(0)$ is a manifold if 0 is a regular value of f. Another way of looking at this is to see that $df_p = ((df_1)_p, \ldots, (df_k)_p) : T_pM \to T_{f(p)}\mathbb{R}^k$, and this is surjective if and only if $(df_1)_p, \ldots, (df_k)_p$ are linearly independent functionals. When this happens we say $f_1, \ldots, f_k$ are independent at p, and so we have the following corollary.

Corollary 1.4. *If $f_1, \ldots, f_k : M \to \mathbb{R}$ are independent at all points where they simultaneously vanish, then the set of common zeros Z is a submanifold with $\dim Z = \dim M - k$.*

In the case of the above corollary we say that the submanifold Z has been *cut out* by the functions $f_1, \ldots, f_k$. However, not every submanifold can be globally cut out by independent functions.

Now that we know the pre-image of regular values are submanifolds, we naturally wonder if the pre-image of *submanifolds* are submanifolds. Two smooth maps $f : M \to N$ and $g : L \to N$ are said to be **transversal** if at any point $p \in M$ and $q \in L$ for which $f(p) = g(q)$ we have

$$df_p(T_pM) + dg_q(T_qN) = T_{f(p)}N$$

and we denote this by $f \pitchfork g$. Note that the above sum does *not* have to be direct, and in fact most of the time will not be. If g is just the inclusion of a submanifold L of N, then we say f is transverse to L and write $f \pitchfork L$. If f is also an inclusion of a submanifold M of N, then we say the submanifolds M and L are transversal, and write $M \pitchfork L$.

Proposition 1.1. *If $f : M \to N$ is transverse to a submanifold L in N, then $f^{-1}(L)$ is a submanifold of M with $\operatorname{codim} f^{-1}(L) = \operatorname{codim} L$. In particular, if two submanifolds M and L of N are transversal, then $M \cap L$ is another submanifold of N and $\operatorname{codim}(M \cap L) = \operatorname{codim} M + \operatorname{codim} L$.*

An important feature of immersions, submersions, transversal intersections and embeddings is that they can be "perturbed" slightly, and they will remain in their respective class. To make this precise we need to define what we mean by perturbation. Fix a smooth map $f_0 : M \to N$, and let $F : M \times [-1, 1] \to N$ be a smooth map such that $F(-, 0) = f_0$. In fact, we will use the notation $f_t = F(-, t)$. Then any of the smooth maps $f_t : M \to N$ for $t \in [-1, 1]$ are called **smooth perturbations** of f_0; that is, we just smoothly homotoped f_0 a little bit. We say that a statement P defined on the space of smooth functions $C^\infty(M, N)$ is **stable** if for every $f \in C^\infty(M, N)$ such that $P(f)$ and every family of perturbations F of f, then there is a small enough ϵ such that $P(f_t)$ for all $|t| < \epsilon$.

Theorem 1.2. *Let M be compact. The following statements are stable as statements on $C^\infty(M, N)$.*

1. *$P(f) = f$ is an immersion.*

2. *$P(f) = f$ is an submersion.*

3. *$P(f) = f$ is a local diffeomorphism.*

4. *For a fixed submanifold $Z \subset N$, $P(f) = f$ is transversal to Z.*

5. *$P(f) = f$ is an embedding.*

6. *$P(f) = f$ is an diffeomorphism.*

Proof.

1. Suppose $f_0 : M^m \to N^n$ is an immersion, and take a family $F : M \times [-1, 1] \to N$ perturbations of f_0, i.e. $F(-, 0) = f_0$. The result will follow if we prove every point $(x, 0) \in M \times [-1, 1]$ has an open neighborhood U_x such that df_p is injective for all $p \in U$. For then we can take their union $U = \cup_x U_x$ and choose an $\epsilon > 0$ small enough so that $M \times (-\epsilon, \epsilon) \subseteq U$. Since this is now just a local question we might as well assume M is an open subset of $\mathbb{R}^m$ and N is an open subset of $\mathbb{R}^n$. Since $(df_0)_x$ is injective, we have that the Jacobian matrix $(\partial (f_0)_i / \partial x_j(x))_{ij}$ contains at least m linearly independent columns. We can of course assume that these columns are the first m columns of the Jacobian. If we consider the Jacobian as a function of $(x, t) \in X \times [-1, 1]$, then we see that the determinant of the first m columns is a smooth function that is not equal to zero at $(0, x)$, hence we can find an $\epsilon > 0$ small enough so that it is not zero for (t, x), $|t| < \epsilon$. Therefore all the f_t's are immersions for $|t| < \epsilon$.

2. The proof of this identical to 1 as now the Jacobian has n linearly independent rows.

3. This follows directly from 1 since a local diffeomorphism is just an immersion for when $\dim M = \dim N$.

4. The condition of being transverse has an equivalent reformulation in terms of local submersions, so this follows from 2.

5. Recall that an embedding is a proper, injective immersion $f_0 : M \to N$. Since M is compact, f_0 is automatically proper, so we only need to prove that if f_0 is injective, then f_t is injective for $|t|$ less than some small $\epsilon > 0$. We prove this by contradiction, so we can assume there is a sequence of numbers $t_i \to 0$ and distinct points $x_i \neq y_i$ such that $f_{t_i}(x_i) = f_{t_i}(y_i)$. Since M is compact we can pass to a subsequence such that $t_i \to 0$, $x_i \to x$, and $y_i \to y$. If we define the function $G : M \times I \to N \times I$, then we have $G(x_i, t_i) = G(y_i, t_i)$ (by definition), hence

$$G(x, 0) = \lim_{i \to \infty} G(x_i, t_i) = \lim_{i \to \infty} G(y_i, t_i) = G(y, 0)$$

Since f_0 is injective this implies $x = y$. The Jacobian of G at $(x, 0)$ is in block form

$$\begin{pmatrix} & & & * \\ & d(f_0)_x & & \vdots \\ & & & * \\ 0 & \cdots & 0 & 1 \end{pmatrix}$$

But, $d(f_0)_x$ is injective, hence $dG_{(x,0)}$ has at least $m + 1$ linearly independent columns and so G is an immersion at $(x, 0)$. This will remain true in a small neighborhood of $(x, 0)$, which must contain the points (x_i, t_i) and (y_i, t_i) for large enough i. This contradicts that $G(x_i, t_i) = G(y_i, t_i)$, therefore no such sequence can exist, and so f_t is injective for $|t|$ small enough.

6.

$\square$

Now let us extend the ideas of transversality to manifolds with boundary. For a manifold M with boundary, and a smooth map $f : M \to N$, the restriction of f to the boundary ∂M will be denoted by ∂f.

UNFINISHED

1.2 Intersection Theory

Recall the following classification of 1-dimensional manifolds and one of its immediate corollaries.

Proposition 1.2. *Every compact 1-dimensional manifold is homeomorphic to a disjoint union of circles and closed intervals.*

Corollary 1.5. *The number of boundary components of any compact 1-dimensional manifold is even.*

Consider a smooth map $f : M \to N$, with M compact, and submanifold $Z \subseteq M$. By slightly perturbing f, if necessary, we can assume $f \pitchfork Z$ and so $f^{-1}(Z)$ is a submanifold of M. If $\operatorname{codim} Z = \dim M$, then $\operatorname{codim} f^{-1}(Z) = \operatorname{codim} Z = \dim M$, hence $f^{-1}(Z)$ is a 0-dimensional manifold. Since M is compact, $f^{-1}(Z)$ consists of finitely many points, so we define the **mod-2 intersection number** of f and Z by $I_2(f, Z) = \# f^{-1}(Z) \bmod 2$. This is well-defined by the following proposition.

Proposition 1.3. *If $f, g : M \to N$ are homotopic, then $I_2(f, Z) = I_2(g, Z) \bmod 2$.*

Proof. Let $F : M \times I \to N$ be a homotopy between f and g. We can perturb F slightly, while keeping F_0 and F_1 fixed, to make $F \pitchfork Z$. Then $F^{-1}(Z)$ is a submanifold of $M \times I$ with $\operatorname{codim} F^{-1}(Z) = \operatorname{codim} Z = \dim M$, hence $\dim F^{-1}(Z) = 1$. The boundary of $F^{-1}(Z)$ consists of

$$F^{-1}(Z) \cap \partial(M \times I) = F^{-1}(Z) \cap (M \times \{0\} \cup M \times \{1\}) = f^{-1}(Z) \times \{0\} \cup g^{-1}(Z) \times \{1\}$$

Since $\#\partial F^{-1}(Z)$ is even, we must have that $\# f^{-1}(Z)$ and $\# g^{-1}(Z)$ are both even or both odd, hence $I_2(f, Z) = I_2(g, Z) \bmod 2$. $\qquad\square$

We can define the mod-2 intersection number of two submanifolds $W, Z \subseteq M$ as $I_2(W, Z) := I_2(i, Z)$ where $i : W \hookrightarrow M$ is the inclusion. Although we can slightly perturb W and Z to create more intersection points (or remove intersection points), the above proposition says that points must be added in pairs.
UNFINISHED

2   Knot Theory

One of the reasons that low dimensional topology is so much harder to study than higher dimensions is the appearance of knots. In high dimensions one can unravel all knots to become "trivial," but this does not happen in 3 dimensions. Fortunately we can get a combinatorial presentation of 3-manifolds from knots, but before that we have to cover some basics of knot theory.

2.1   Basic Definitions

Classically, a **knot** k is just a smoothly embedded copy of S^1 into S^3. We could also embed our knots into $\mathbb{R}^3$, the theories would turn out to be equivalent, but S^3 is easier to think about since it is a nice compact 3-manifold. The complement of a knot in its surrounding manifold is called the **knot complement**, and its first homology group is quite simple.

Proposition 2.1. *If k is a knot in S^3, then $H_1(S^3 \backslash k) = \mathbb{Z}$ and is generated by any meridian.*

Proof. Let U be an open tubular neighborhood of k in S^3, and let V be the complement of the closure of a *smaller* tubular neighborhood. Then U is homotopy equivalent to S^1, V is homotopy equivalent to the knot complement $S^3 \backslash k$, and $U \cap V$ is homotopy equivalent to a torus $S^1 \times S^1$. The Mayer-Vietoris sequences gives us the following exact sequence

$$0 \longrightarrow H_1(U \cap V) \longrightarrow H_1(U) \oplus H_1(V) \longrightarrow 0$$

which is just

$$0 \longrightarrow \mathbb{Z} \oplus \mathbb{Z} \longrightarrow \mathbb{Z} \oplus H_1(S^3 \backslash k) \longrightarrow 0$$

This implies that $H_1(S^3 \backslash k) = \mathbb{Z}$. Further, note that this map is of the form $(i_*, -j_*)$, where i and j are the inclusions of $U \cap V$ into U and V, respectively. The group $H_1(U \cap V) = \mathbb{Z} \oplus \mathbb{Z}$ is generated by a meridian $[\mu]$ and longitude $[\ell]$, and we can choose this longitude such that $j_*[\ell] = 0$. Since $[\ell]$ is mapped to an element generating the first factor of $\mathbb{Z} \oplus H_1(S^3 \backslash k)$, we have that $j_*[\mu]$ generates the second factor. $\qquad \square$

Given two knots k_1, k_2 in a manifold S^3, we say that they are **equivalent**, or **isotopic**, if they are ambiently isotopic. This means that there is a smooth 1-parameter family of maps $\varphi_t : S^3 \to S^3$ such that $\varphi_0 = \mathrm{id}$, $\varphi_1(k_1) = k_2$, and φ_t is a diffeomorphism for all time t. A knot that is isotopic to the standardly embedded S^1 in S^3 is called the **unknot** or the **trivial knot**.

There is a nice way to visualize a knot k in S^3. First, move the knot a little if needed so that it does not pass through the point at infinity. Now k is embedded in $\mathbb{R}^3$. For a generic plane Π in $\mathbb{R}^3$ not intersecting k, the projection D of k onto Π is a closed curve with finitely many double points, and no other types of singularities. We can decorate these double points as an over or under crossing depending on whether one strand is above another relative to the plane. Such a projection D is called a **knot diagram** of k. Conversely, given a closed curve D in the plane which is embedded on the complement of finitely many double points and whose double points are decorated with under or over crossings, we can find a knot k such that D is a projection of k.

Clearly a single knot k can admit many different knot diagrams, and so we hope there is a way to determine when two diagrams correspond to the same isotopy class of knot. It turns out there are 3 moves one can apply to a knot diagram, called the **Reidemeister moves**, such that two knot diagrams D_1 and D_2 are the diagrams of two isotopic knots if and only if one can get from one diagram to the other by the Reidemeister moves (see figure 2.1). We will use R_1, R_2 and R_3
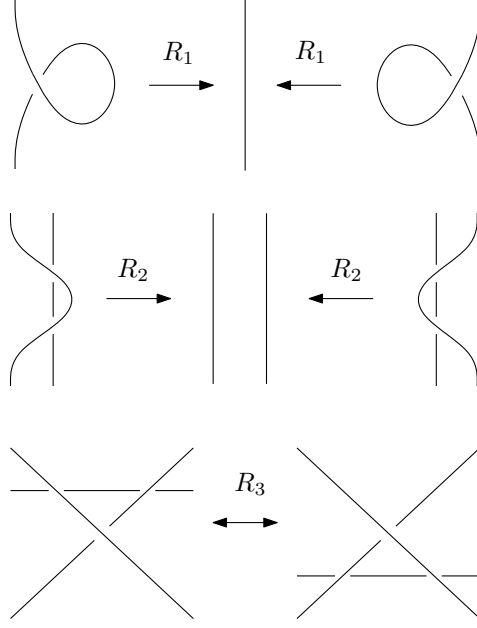
Figure 2.1: Reidemeister Moves

$+1$ $\qquad\qquad\qquad\qquad -1$
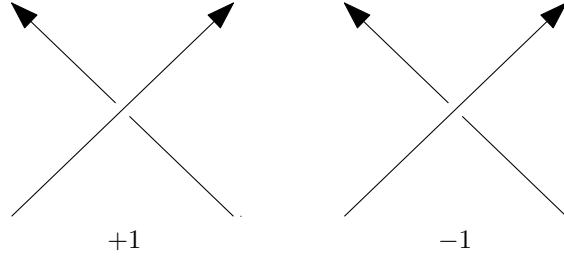
Figure 2.2: Crossing Signs

to denote these moves. This reduces a lot of knot theory to combinatorial pictures, from which we can define a lot of knot invariants.

A slight generalization of a knot is a link. An n-**component link** is a smooth embedding of n copies of S^1. All of the above definitions and discussion has a direct generalization to links. Clearly if two links L_1 and L_2 are isotopic, then their number of components must be equal.

One can put an orientation on a knot k by simply choosing a non-vanishing tangent vector field on k, or equivalently placing an arrow on any knot diagram of k. A knot has precisely 2 orientations and an n-component link has 2^n orientations. Given a knot diagram D of an oriented knot k, we can define the **sign** of each crossing of D. We adopt the right hand rule for sign conventions, which means signs are determined as in figure 2.2. Note that the sign of a crossing is information we get once we have chosen a knot diagram, not from just a knot. The sum of all the signs of the crossing of a knot diagram D is called the **writhe** of the diagram, and is denoted by $w(D)$. The writhe of a knot diagram is *not* a knot invariant because it is not invariant under R_1. It is, however, a ribbon invariant, but we will not discuss that here.

Let k be a knot in S^3 and D a diagram of k. A **resolution of a crossing**, or a **smoothing**, is the process of removing the double point and reconnecting the strands so that there is no crossing. There are two ways of doing this, called the 0-resolution and 1-resolution, shown in figure 2.3. Each

0-resolution   crossing   1-resolution

Figure 2.3: Ways to resolve a crossing

time one resolves a crossing the number of components of the diagram either stays the same or increases by one. A **resolution of a diagram** is the diagram resulting from resolving all crossings, which is just a finite collection of disjoint circles. Clearly if there are n crossings in the diagram D, then there are 2^n resolutions, which are parameterized by $\{0,1\}^n$. If the knot k has an orientati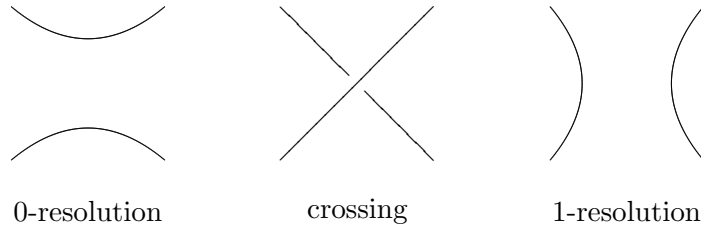on, then there is a canonical resolution of D, called the **oriented resolution**, where we resolve all the crossings of sign $+1$ by a 1-resolution and all crossings of sign -1 by a 0-resolution.

A **Seifert surface** for a link L in S^3 is a smoothly embedded, compact, connected and orientable surface F in S^3 with boundary such that $\partial F = L$.

Proposition 2.2 (Seifert's Algorithm). *Every link L in S^3 has a Seifert surface.*

Proof. We can construct the surface explicitly. Choose any orientation on L and let D be a diagram for L. Let D' denote the the oriented resolution of D. Fill each component of D' so that there is a disc for each circle, and connect the discs a twisting band for each crossing. The result is a surface with boundary equal to exactly L. □

Any link has many Seifert surfaces, and Seifert's algorithm almost never produces the surface of minimal genus. We define the **genus** of a link L to be the minimum of genera of Seifert surfaces of L, and is denoted by $g(L)$. The **four-ball genus** of L is the minimum of genera of compact, connected, oriented and smoothly embedded surfaces in D^4 whose boundary is $L \subset \partial D^4 = S^3$, and is denoted by $g^*(L)$.

Lemma 2.1. *Let k_1, k_2 be two knots embedded in S^3. Then the genus satisfies $g(k_1 \# k_2) = g(k_1) + g(k_2)$.*

Proof. We can think of k_1 and k_2 embedded in the same sphere, but situated far apart. Let F_1 and F_2 be Seifert surfaces of minimal genus for k_1 and k_2, respectively. Then the boundary sum $F_1 \natural F_2$ is a Seifert surface of $k_1 \# k_2$, and so we have proved

$$g(k_1 \# k_2) \leq g(k_1) + g(k_2)$$

To prove the other inequality we fix a Seifert surface F of minimal genus for $k_1 \# k_2$. We can fix a 2-sphere Σ in S^3 so that $\Sigma \cap k_1 \# k_2$ consists of two points. Then, if the two arcs of $k_1 \# k_2$ between these two points are labelled α_1 and α_2, and we take a small arc β in Σ between these two points, then the curve $\alpha_1 + \beta$ is k_1 and the curve $\alpha_2 + \beta$ is k_2. By perturbing Σ slightly we can guarantee that $\Sigma \cap F$ is just a collection of closed curves, as well as the curve β. We will perform surgery on these closed curves one by one to remove these intersections, so that the only intersection points left is the arc β.

9

Let C be one of these closed curves with the additional property that the disc it bounds in Σ does not contain any of the other closed curves. These ensures that if we push the interior of this disc a little off of Σ (in either direction) it does not intersect F. Let $\hat{F}$ be the surface obtained from F by cutting out an open annulus around C (so now there are two more boundary components) and gluing in two discs, each parallel to the disc in Σ bounded by C, and on either side of Σ. The result is another Seifert surface $\hat{F}$ of $k_1 \# k_2$. If the curve C does not separate F, then $\hat{F}$ will have genus one less that F (we just cut open a handle!), which is not possible since F is assumed to have minimal genus. So, we must have that C separates F, hence $\hat{F}$ consists of a disjoint union of a 2-sphere and a Seifert surface of $k_1 \# k_2$ that has one less intersection with Σ. Repeating this for all the closed curves in $\Sigma \cap F$ shows that we can assume that Σ and F intersect only at the arc β.

Now we can take F_1 to be the closure of the part of F outside of the sphere Σ, and F_2 the closure of the part of F inside the sphere. Then F_1 and F_2 are Seifert surfaces of k_1 and k_2 respectively, hence

$$g(k_1) + g(k_2) \leq g(k_1 \# k_2)$$

$\square$

Given disjoint, oriented knots k_1 and k_2, let D be a common projection. We can assign signs to the double points of D just as before, except now the two strands might belong to different knots. The **linking number** $\text{lk}(k_1, k_2)$ of k_1 and k_2 is defined to be the sum of the signs of all crossings where k_1 passes under k_2. In order to show that this definition is well-defined we need to see that it does not depend on the diagram chosen for $k_1 \cup k_2$. It is an easy matter to show that $\text{lk}(k_1, k_2)$ is invariant under the Reidemeister moves, hence $\text{lk}(k_1, k_2)$ is an invariant of the isotopy class of $k_1 \cup k_2$. It is also easy to check that $\text{lk}(k_1, k_2) = \text{lk}(k_2, k_1)$ and $\text{lk}(-k_1, k_2) = -\text{lk}(k_1, k_2)$, where $-k_1$ is the knot k_1 with opposite orientation.

There are two other ways of describing the linking number that will be useful later. Let $L = k_1 \cup k_2$ be the link with components our knots, and let F be a Seifert surface for L. At each point on the boundary k_1 in F we have a tangent vector τ that determines the orientation of k_1 and a vector ν pointing into the surface. Let n be a vector normal to F such that the ordered basis (τ, ν, n) matches with the orientation of S^3. The normal vector n can be extended to the rest of the surface F, and it determines an orientation. We can perturb k_2 slightly if needed so that k_2 intersects L transversely in finitely many points. If we assign these points the sign $+1$ when k_2's tangent vector points in the same direct as n and -1 when it points in the opposite direction, then $\text{lk}(k_1, k_2)$ is equal to the sum of these numbers.

The second way is to notice that the first homology of the knot complement $H_1(M \backslash k_1) = \mathbb{Z}$ is generated by the homology class $[\mu]$ of any meridian μ of k_1. The other knot k_2 determines a homology class $[k_2] \in H_1(M \backslash k_1)$, which must be an integer multiple of $[\mu]$. The linking number $\text{lk}(k_1, k_2)$ is precisely this number, i.e. $[k_2] = \text{lk}(k_1, k_2) \cdot [\mu]$.

A **framing** for a knot k in S^3 is a choice of a normal vector field along k. If X is a framing of k, then we can push k a little bit in the direction of X to get a new knot k', and the integer $\text{lk}(k, k')$ is called the **framing number** or **framing coefficient** of X. Conversely, given an integer n we can find a parallel k' of k such that $\text{lk}(k, k') = n$, and this curve is unique up to isotopy in $S^3 \backslash k$. So, we can think of framings as either normal vector fields, parallel knots, or just integers assigned to knots. A knot with a framing is called a framed knot, and a framed link is defined similarly. For an oriented, framed link L with components $L_1, \ldots, L_n$ we define its **linking matrix** $[L]$ to be the symmetric matrix with (i, j)-entry equal to $\text{lk}(L_i, L_j)$, where we use the convention that $\text{lk}(L_i, L_i)$ is just the framing of the i-th component of L.
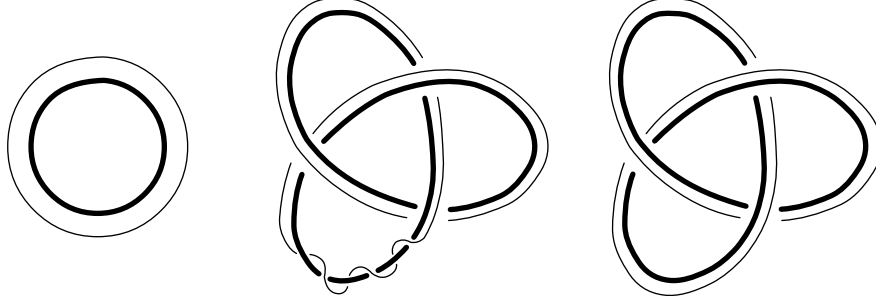
Figure 2.4: Examples of the canonical longitude and a non-canonical longitude

Given a knot k in S^3 the meridian μ of $\partial N(k)$ chosen above is called the **canonical meridian**, and is always defined. However, a longitude ℓ of $\partial N(k)$ is not so clear how to define since it can wrap around k in complicated ways. The ways in which ℓ can wrap around k are parameterized by linking numbers, so we define the **canonical longitude** to be the curve ℓ in $\partial N(k)$ such that $\mathrm{lk}(k, \ell) = 0$. This curve is unique up to isotopy, and will be useful in our discussion on knot surgeries. Given a diagram D of a knot k one can draw the canonical longitude as a parallel of k that wraps $-w(D)$ full times around k, where $w(D)$ is the writhe of the diagram. Take note that we are using the linking number to define the canonical longitude even though our knot need not be oriented. However, we can given k any orientation and point ℓ in the same direction, and then the linking number will not depend on the orientation we gave k.

In figure 2.4 we have three knots drawn (the bolded knots) with longitudes (not bolded). The first two longitudes are canonical, as one can easily calculate their linking numbers to be 0. However the third longitude, which seems like a more obvious choice for a longitude than the second, is *not* canonical since its linking number is -3.

The fundamental group of the knot complement $\pi_1(S^3 \backslash k)$ is called the **knot group**, and clearly is an invariant of the knot. There is a simple algorithm for computing a presentation of $\pi_1(S^3 \backslash k)$ given a diagram of D. For a diagram D of k, let $n+1$ be the number of crossings. Let $\alpha_0, \ldots, \alpha_n$ be the arcs in D that start at a crossing, travel along the under-strand, and end at the next crossing that the strand goes under such that α_i starts at where α_{i-1} ends (we always take the subscripts of the α's to be mod n). Also orient these arcs such that α_i points towards the starting point of α_{i+1}. For convenience, suppose that k does not pass through $(0, 0, 1)$ (the point at infinity in $\mathbb{R}^3$), and let us use this as the base point in $S^3 \backslash k$. In terms of the knot diagram, we can think of the base point as sitting right where we (the viewers) are, and so loops based at ∞ leave from us, travel through the knot diagram, and then return to us. Hence, a loop in the knot diagram can be represented by just a curve segment (*not* a loop), and we just assume the ends are connected to the point at infinity in the obvious way.

Generators of $\pi_1(S^3 \backslash k)$ are just short arc segments $x_0, \ldots, x_n$ in the diagram such that x_i passes under α_i, and oriented so that it crosses under α_i from the right to the left. Each crossing of the knot diagram gives a relation r_i on the generators of the fundamental group. Suppose the crossing is where the arcs α_i, α_j and α_{j+1} come together for some indices i and j (hence α_i is the over-strand). If the sign of the crossing is $+1$, then we add the relation $x_i x_j = x_{j+1} x_i$, and if the sign is -1, then we add the relation $x_j x_i = x_i x_{j+1}$. Note that we are computing the sign of these crossings with the orientation on k induced by the orientations on the α arcs, which were chosen for convenience. If we changed those orientations we would get an equivalent presentation of the fundamental group. Let r_i denote the relation obtained at the crossing where α_i and α_{i+1} meet.
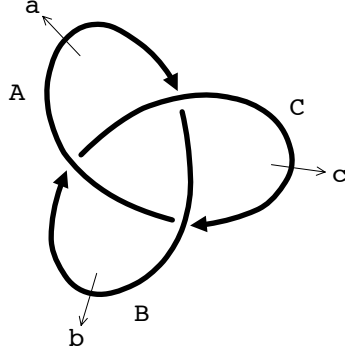
Figure 2.5: Computing the knot group of the trefoil

Proposition 2.3. *The fundamental group of $S^3 \backslash k$ is isomorphic to $\langle x_0, \ldots, x_n : r_0, \ldots, r_n \rangle$.*

Proof. First we isotope k so that it lies in $\{z = 0\} \subset \mathbb{R}^3$, except at a crossing where the understand dips to the $\{z = -\epsilon\}$ plane for a short segment, which we call β_i for the crossing corresponding to where α_i and α_{i+1} meet. We will decompose $S^3 \backslash k$ into a bunch of pieces and apply van Kampen's theorem. Let $A = \{z \geq -\epsilon\} \backslash k$, and let B_i $(i = 0, \ldots, n)$ be a small open rectangular box around the segment β_i, but with β_i remove and with a thin tube connected the box to ∞. Finally, let C be the complement of the closure of $A \cup B_0 \cup \cdots \cup B_n$.

First we claim that $\pi_1(A)$ is a free group generated by $x_0, \ldots, x_n$. To see this let $\overline{\alpha}_i$ be the arc α_i whose endpoints dive straight down to connect to the plane $\{z = -\epsilon\}$, and let $\overline{k}$ be the union of these arcs. Then $\pi_1(S^3 \backslash k)$ is isomorphic to $\pi_1(S^3 \backslash \overline{k})$, so we compute the later group. If we take a curtain below the arc $\overline{\alpha}_i$ to the plane $\{z = -\epsilon\}$ and then thicken it we get a ball B_i. Removing the B_i's from $\{z \geq -\epsilon\}$ yields a simply connected space, so we determine what happens to the fundamental group when we add the pieces $B_i \backslash \overline{\alpha}_i$ back. The piece $B_i \backslash \overline{\alpha}_i$ is just a punctured disc, so its fundamental group is generated by a loop going around this hole once. Also the piece $B_i \backslash \overline{\alpha}_i$ intersects the space where we removed all the B_j's from $\{z \geq -\epsilon\}$ is just a disc, hence we added one generator by adding the piece $B_i \backslash \overline{\alpha}_i$. Continuing this we get the claim.

UNFINISHED □

This presentation of the knot group is called the **Wirtinger presentation**.

Example 2.1. Consider the diagram of the trefoil knot k shown in figure 2.5 (we label our arcs by A, B, C and generators of π_1 by a, b, c). The Wirtinger representation of the knot group is

$$\pi_1(S^3 \backslash k) = \langle a, b, c : ba = ac, ac = cb, cb = ba \rangle$$

We can eliminate a generator from this presentation by seeing that $c = bab^{-1}$, hence another presentation of the fundamental group is given by

$$\pi_1(S^3 \backslash k) = \langle a, b : bab = aba \rangle$$

If we set $x = aba$ and $y = ba$, then we see that $x^2 = abaaba = bababa = y^3$, hence yet another presentation is given by

$$\pi_1(S^3 \backslash k) = \langle x, y : x^2 = y^3 \rangle$$

2.2 The Alexander Polynomial

We will now construct some classical invariants of knots and links in S^3. With some work these invariants can be extended to links in integral homology spheres, as we will see later. These combinatorial invariants have very interesting relationships with "high-powered" invariants, like Seiberg-Witten invariants and Heegaard Floer homology.

Fix a diagram D of a knot L in S^3 with n components, and give the link any orientation. Suppose this diagram has k double points and the oriented resolution of this diagram has c connected components. Let us compute the Euler characteristic and genus of the Seifert surface F constructed from Seifert's algorithm on D (see proposition 2.2). The Euler characteristic of c disjoint discs is c, and attaching a band to these discs decreases the Euler characteristic by one. Therefore $\chi(F) = c - k$. On the other hand, an orientable surface of genus g with n boundary components has Euler characteristic $2 - 2g - n$. Setting these two Euler characteristics equal to each other and solving for g shows that the genus of F is

$$g = 1 + \frac{k - c - n}{2} \tag{2.1}$$

One can compute that the first homology of a surface of genus g with $n \geq 1$ boundary components is the free abelian group of rank $2g + n - 1$ (just look at the handlebody decomposition of such a surface), therefore, by (2.1), $H_1(F)$ is free abelian with rank $1 + k - c$.

Now, for a general Seifert surface F of a link L in S^3, let $\alpha = (x_1, \ldots, x_\ell)$ be an ordered collection of non-intersecting, oriented, closed curves in F that form an ordered basis for $H_1(F)$. Choose an orientation of F by specifying a nowhere zero, normal vector field n on F. For each curve x_i, let x_i^+ be the curve in S^3 that is x_i pushed off the surface F slightly in the direction of n. For each pair of curves x_i, x_j^+ we can compute their linking numbers (they do not intersect), and so we define the **Seifert matrix** of F with orientation n to be the matrix $[F] = (\mathrm{lk}(x_i, x_j^+))_{i,j=1}^\ell$. Of course this matrix depends on our choice of orientation of F and basis for $H_1(F)$, so to include this dependence in the notation we may sometimes write $[F]_{\alpha,n}$. Suppose we have fixed our basis $x_1, \ldots, x_\ell$ with pushed off curves $x_1^+, \ldots, x_\ell^+$ relative to n. If we change the orientation of F to $-n$, then we can translate F a little bit so that $x_1^+, \ldots, x_\ell^+$ lie in F, in which case the curves $x_1, \ldots, x_\ell$ are now the pushed off curves relative to $-n$. So, the Seifert matrix with respect to this orientation is $(\mathrm{lk}(x_i^+, x_j))$. This is equal to $(\mathrm{lk}(x_j, x_i^+))$ by the symmetry of linking numbers, which is the transpose of the Seifert matrix computed with respect to n. So, reversing orientations transposed the Seifert matrix. If we fix the orientation, then any two bases of $H_1(F)$ are related by an invertible matrix U (invertible over $\mathbb{Z}$ means $\det U = \pm 1$), so if α and β are two bases related by U, then $[F]_\beta = U^T [F]_\alpha U$.

However, even if we consider two Seifert matrices A and B equivalent if they are related by $A = U^T B U$, for some invertible U, we do not get a knot invariant. We could just analyze how Seifert matrices transform under the Reidemeister moves, but a nicer way is to find a set of moves that can be performed on Seifert surfaces such that any two Seifert surfaces of a knot are related by these moves. One such move is called stabilization and it increases the genus of a Seifert surface by one. Let F be a Seifert surface of k embedded in S^3 and T' a torus with a small disc cut out (so its boundary is S^1) also embedded in S^3 such that $F \cap T' = \emptyset$. Taking the boundary sum of these two surfaces $F \sharp T'$ gives another Seifert surface of k, and is called the **stabilization** of F. We say two Seifert surfaces are **stably isotopic** if they become isotopic after a finite number of stabilizations.

Proposition 2.4. *Any two Seifert surfaces of a fixed knot are stably isotopic.*

13

We have already remarked that a change of basis of $H_1(F)$ corresponds to changing the Seifert matrix of F by $U^T S U$, for some integral matrix with $\det U = \pm 1$. Let us see how stabilization of a Seifert surface changes the Seifert matrix. Let F be a Seifert surface of k with generators $x_1, \ldots, x_{2g}$ for $H_1(F)$. By forming $F' = F \sharp T'$ we are adding two generators to the first homology, one curve going around the longitude of T' and one curve going around the meridian, so let y and z denote these curves. Since the torus T' can knot and link with F in complicated ways, we have that column and row of $[F']$ corresponding to y can consist of arbitrary integers, except that either $\mathrm{lk}(y, z^+) = \pm 1$ and $\mathrm{lk}(z, y^+) = 0$, or the opposite. Therefore the Seifert matrix $[F']$ is given by

$$
\begin{pmatrix}
 & & & * & 0 \\
 & [F] & & \vdots & \vdots \\
 & & & * & 0 \\
* & \cdots & * & * & \pm 1 \\
0 & \cdots & 0 & 0 & 0
\end{pmatrix}
\quad \text{or} \quad
\begin{pmatrix}
 & & & * & 0 \\
 & [F] & & \vdots & \vdots \\
 & & & * & 0 \\
* & \cdots & * & * & 0 \\
0 & \cdots & 0 & \pm 1 & 0
\end{pmatrix}
$$

This leads us to define two operations we can perform on integral matrices. Given an integral, $n \times n$ matrix A (not necessarily invertible), performing an S_1 move is to change the matrix to $U^T A U$ for some invertible integral matrix U. Performing an S_2 move is to change the matrix to the $(n+2) \times (n+2)$ matrix given by

$$
\begin{pmatrix}
 & & & * & 0 \\
 & S & & \vdots & \vdots \\
 & & & * & 0 \\
* & \cdots & * & * & \pm 1 \\
0 & \cdots & 0 & 0 & 0
\end{pmatrix}
\tag{2.2}
$$

where the $*$'s can be any numbers. We say that two integral matrices are S-**related** one can be brought to the other with a finite sequence of S_1 and S_2 moves. An immediate consequence of proposition 2.4 is

Corollary 2.1. *Any two Seifert matrices of a knot are S-equivalent.*

Therefore the S-equivalence class of any Seifert matrix of a knot is an invariant of the knot. There are simpler or more computable invariants we can derive from the Seifert matrix. Given a knot k, let S be any Seifert matrix for k. The Laurent polynomial

$$
\Delta_k(t) = \det\left(t^{1/2} S - t^{-1/2} S^T \right)
$$

in $t^{1/2}$ is called the **Alexander polynomial** of k. Technically we should call this the Conway normalization of Alexander's polynomial, as Alexander's original definition used the infinite cyclic covering of the knot complement, and was defined only up to multiplication by a unit. Since the dimension of S is $2g$ (where g is the genus of the surface associated to S) we have $\Delta_k(t)$ is actually a Laurent polynomial in t, and a direct substitution shows $\Delta_k(t) = \Delta_k(t^{-1})$. If S is of genus g, then we can easily see that $\Delta_k(t) = t^{-g} \det(S - tS^T)$.

Proposition 2.5. *The Alexander polynomial does not depend on the Seifert matrix chosen, hence is a knot invariant.*

Proof. If S' is another Seifert matrix for k, then S and S' are related by a finite sequence of S_1 and S_2 moves by corollary 2.1. Let $\Delta_k(t)$ denote the Alexander polynomial with respect to S, and $\Delta'_k(t)$ the Alexander polynomial with respect to S'. If $S' = U^T S U$, then we see that

$$\Delta'_k(t) = \det\left(t^{1/2} U^T S U - t^{-1/2} U^T S U\right)$$
$$= (\det U)^2 \det\left(t^{1/2} S - t^{-1/2} S^T\right)$$
$$= \Delta_k(t)$$

where the last line follows since $\det U = \pm 1$. On the other hand, if S' is the result of performing the S_2 move on S (see (2.2)), then S'^T can be written as

$$\begin{pmatrix} & & & * & 0 \\ & S^T & & \vdots & \vdots \\ & & & * & 0 \\ * & \cdots & * & * & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}$$

and the matrix $t^{1/2} S' - t^{-1/2} S'^T$ can be written as

$$\begin{pmatrix} & & & & * & 0 \\ & t^{1/2} S - t^{-1/2} S^T & & & \vdots & \vdots \\ & & & & * & 0 \\ * & & \cdots & & * & * & t^{1/2} \\ 0 & & \cdots & & 0 & -t^{-1/2} & 0 \end{pmatrix}$$

When computing the determinant of this matrix we can perform a cofactor expansion along the last column, and then another cofactor expansion along the last row of the resulting matrix, in which case we see that the determinant of the above is precisely

$$-t^{-1/2}\left(-t^{1/2} \det\left(t^{1/2} S - t^{-1/2} S^T\right)\right) = \Delta_k(t)$$

hence the Alexander polynomial is invariant under the S_2 moves. $\qquad\square$

We can extend the definition of Alexander polynomial to links in the obvious way, but it will no longer be independent of the Seifert surface. The problem is that a link L with n components has 2^n orientations. A Seifert surface F for L can induce two of these orientations on L (and they are opposite), and conversely we say that a Seifert surface is compatible with an oriented link if the induced orientation agrees with the link's orientation. While it is true that any two Seifert surfaces compatible with an oriented link are stably equivalent, it is *not* true that any two Seifert surfaces of a link are stably equivalent. So, the Alexander polynomial is an invariant of oriented links with compatible Seifert matrices.

As mentioned above, the Alexander polynomial is not sensitive to the orientation of a knot, i.e. $\Delta_k(t) = \Delta_{-k}(t)$. Unfortunately, the Alexander polynomial is also not sensitive to changing a knot to its mirror. The **mirror** of a knot k is the knot $\bar{k}$ obtained by reflecting k through any plane in $\mathbb{R}^3$. This is equivalent to changing all over crossings to under crossings (and vice versa) in a diagram of k.

Proposition 2.6. *For any knot k, $\Delta_k(t) = \Delta_{\bar{k}}(t)$.*

Proof. If S is a Seifert matrix of k, then $-S$ is a Seifert matrix of $\bar{k}$. Let $2m$ be the order of S. Then

$$\Delta_{\bar{k}}(t) = \det\left(-t^{1/2}S + t^{-1/2}S^T\right)$$
$$= \det\left(t^{-1/2}S^T - t^{1/2}S\right)$$
$$= \det\left(t^{-1/2}S - t^{1/2}S^T\right)$$
$$= \Delta_k(t^{-1}) = \Delta_k(t)$$

$\square$

There are also scalar invariants of knots that one can derive from Seifert matrices.

Proposition 2.7. *If S is the Seifert matrix of some Seifert surface of a knot k, then $|\det(S+S^T)|$ is a knot invariant.*

Proof. We just have to check that $|\det(S + S^T)|$ does not change after applying S_1 and S_2 moves to S. If we change S by $U^T S U$, then clearly

$$|\det(U^T S U + U^T S^T U)| = |(\det U)^2 \det(S + S^T)| = |\det(S + S^T)|$$

Let S' be the matrix resulting from applying S_1 to S (see (2.2)). Then $S' + S'^T$ is of the form

$$\begin{pmatrix} & & & * & 0 \\ & S+S^T & & \vdots & \vdots \\ & & & * & 0 \\ * & \cdots & * & * & \pm 1 \\ 0 & \cdots & 0 & \pm 1 & 0 \end{pmatrix}$$

By adding multiples of the last column to the other columns, and multiples of the last row to the other rows we can bring this matrix into the form

$$\begin{pmatrix} & & & 0 & 0 \\ & S+S^T & & \vdots & \vdots \\ & & & 0 & 0 \\ 0 & \cdots & 0 & 0 & \pm 1 \\ 0 & \cdots & 0 & \pm 1 & 0 \end{pmatrix} \qquad (2.3)$$

and the determinant of this matrix is $-\det(S + S^T)$. Therefore the absolute value of this determinant remains invariant under S_1 and S_2 moves. $\square$

We call $|\det(S+S^T)|$ the **determinant** of the knot k, and denote it by $\det(k)$. The relationship between the determinant of k and the Alexander polynomial is clear.

Proposition 2.8. $|\Delta_k(-1)| = \det(k)$

Proof. Let S be any Seifert matrix of k and let $2m$ be the order of S (the order must be even for a knot). Then

$$|\Delta_k(-1)| = \left|\det\left((-1)^{1/2}S - (-1)^{-1/2}S^T\right)\right|$$
$$= \left|\det\left((-1)^{1/2}S + (-1)^{1/2}S^T\right)\right|$$
$$= \left|(-1)^m \det(S + S^T)\right|$$

$\square$

The Alexander polynomial satisfies a skein relation, which is closely related to another polynomial invariant that is essentially the same as the Alexander polynomial. For an oriented link L and a diagram D of L, choose any crossing in this diagram and let L_+ denote the diagram resulting from switching this crossing to a positive crossing, L_- the diagram resulting from switching to a negative crossing, and L_0 the diagram resulting from resolving the crossing in the oriented way (see ??).

Theorem 2.1. *The Alexander polynomial of an oriented link L satisfies the skein relation*

$$\Delta_{L_+}(t) - \Delta_{L_-}(t) = (t^{1/2} - t^{-1/2})\Delta_{L_0}(t)$$

Further, this relation, along with the normalization $\Delta_{unknot}(t) = 1$, uniquely determines the Alexander polynomial as a function from oriented links to $\mathbb{Z}[t^{1/2}, t^{-1/2}]$.

Proof. Let F_0 be the Seifert surface for L_0 constructed from Seifert's algorithm (see proposition 2.2). We can construct Seifert surfaces $F_\pm$ for $L_\pm$ by gluing in a twisted band. Take a basis $\{x_1, \ldots, x_{k-1}\}$ for $H_1(F_0)$, then we can get a basis for $H_1(F_\pm)$ by adding one extra curve x_k that goes through the twisted band once. Let S_0 be the Seifert matrix of F_0 computed with respect to $\{x_1, \ldots, x_{k-1}\}$. The Seifert matrices $S_\pm$ of $F_\pm$ are of the form

$$S_+ = \begin{pmatrix} & & & b_1 \\ & S_0 & & \vdots \\ & & & b_{k-1} \\ a_1 & \cdots & a_{k-1} & N \end{pmatrix} \qquad S_- = \begin{pmatrix} & & & b_1 \\ & S_0 & & \vdots \\ & & & b_{k-1} \\ a_1 & \cdots & a_{k-1} & N-1 \end{pmatrix}$$

for some integers $a_1, \ldots, a_k, b_1, \ldots, b_k, N$. Writing out the entries of $t^{1/2}S_\pm - t^{-1/2}S_\pm^T$ we see that the last column and row of these matrices are equal, except for the bottom-right corner entry where one matrix has $(t^{1/2} - t^{-1/2})N$ and the other has $(t^{1/2} - t^{-1/2})(N-1)$. So, if we compute $\Delta_{L_+}(t) - \Delta_{L_-}(t)$ by performing a cofactor expansion along the last columns, then the terms from $\Delta_{L_+}(t)$ will cancel those from $\Delta_{L_-}(t)$ except for the last term, which is equal to

$$(-1)^{2k}(t^{1/2} - t^{-1/2})N\Delta_{L_0}(t) - (-1)^{2k}(t^{1/2} - t^{-1/2})(N-1)\Delta_{L_0}(t) = (t^{1/2} - t^{-1/2})\Delta_{L_0}(t)$$

Therefore $\Delta_L(t)$ satisfies the skein relation. $\square$

Although the Alexander polynomial satisfies a skein relation, this relation does not quite look like the relation for the Jones polynomial. To make things more uniform we introduce the Conway polynomial $\nabla_L(z)$. This is a function $\nabla : \{\text{oriented links}\} \to \mathbb{Z}[z]$ that is uniquely determined by the normalization $\nabla_{\text{unknot}}(z) = 1$ and the skein relation

$$\nabla_{L_+}(z) - \nabla_{L_-}(z) = z\nabla_{L_0}(z)$$

The relationship between the Alexander and Conway polynomial is given by $\Delta_L(t) = \nabla_L\left(t^{1/2} - t^{-1/2}\right)$.

2.3 Knot Signatures and 4-Ball Genus

A very important invariant of knots is the signature.

Proposition 2.9. *Let ω be a unit complex number not equal to 1, and let S be a Seifert matrix of the knot k. Then the signature of the Hermitian matrix $(1-\omega)S + (1-\overline{\omega})S^T$ over $\mathbb{C}$ is an invariant of the knot.*

Proof. We just have to check that the signature of $(1 - \omega)S + (1 - \overline{\omega})S^T$ does not change after applying S_1 and S_2 moves to S. Clearly the signature is invariant under S_1 moves. Let S' be the Seifert matrix resulting in applying an S_2 move, then we have $(1 - \omega)S + (1 - \overline{\omega})S^T$ is congruent to

$$
\begin{pmatrix}
 & & & & 0 & 0 \\
 & (1 - \omega)S + (1 - \overline{\omega})S^T & & & \vdots & \vdots \\
 & & & & 0 & 0 \\
0 & \cdots & & 0 & 0 & 1 - \omega \\
0 & \cdots & & 0 & 1 - \overline{\omega} & 0
\end{pmatrix}
$$

The signature of this block matrix is the sum of the signatures of the blocks. Since the signature of the bottom-right block is zero, we have that the signature of the above matrix is the signature of $(1 - \omega)S + (1 - \overline{\omega})S^T$, hence the signature is invariant under S_2 moves. $\qquad\square$

We call the signature of $(1 - \omega)S + (1 - \overline{\omega})S^T$ the ω-**signature** of the knot, and denote it by $\sigma_\omega(k)$. If $\omega = -1$, then we just call this *the* signature of k, and denote it by $\sigma(k)$.

For a knot k in S^3, consider the 4-ball D^4 bounded by S^3. Let $g^*(k)$ denote the minimal genus of a surface F that can be smooth embedded in D^4 with $\partial F = k \subset S^3$. This knot invariant is called the **smooth 4-ball genus**, or sometimes the **smooth slice genus**. A knot is said to be **slice** if $g^*(k) = 0$. The relationship between signature and slice knots is the following.

Theorem 2.2. *If k is a slice knot, then $\sigma(k) = 0$.*

We will outline the proof of this theorem in a series of lemmas. Suppose k is a slice knot with slicing disc $D \subset D^4$. If F is a Seifert surface for k in S^3, then we can glue together $F \cup D$ to obtain a closed surface in D^4.

Lemma 2.2. *The surface $F \cup D$ bounds a smoothly embedded, orientable 3-manifold M in D^4 with $M \cap S^3 = F$.*

Lemma 2.3. *Let M be a compact, orientable 3-manifold such that ∂M is a surface of genus g. Then the kernel of the map $i_* : H_1(\partial M; \mathbb{Q}) \to H_1(M; \mathbb{Q})$ induced by the inclusion is of dimension g.*

Proof. We can put the homology and cohomology long exact sequences of the pair $(M, \partial M)$ into the following commutative diagram

$$
\begin{array}{ccccc}
H_2(M, \partial M; \mathbb{Q}) & \longrightarrow & H_1(\partial M; \mathbb{Q}) & \xrightarrow{\ i_*\ } & H_1(M; \mathbb{Q}) \\
\downarrow & & \downarrow & & \downarrow \\
H^1(M; \mathbb{Q}) & \xrightarrow{\ i^*\ } & H^1(\partial M; \mathbb{Q}) & \xrightarrow{\ \delta\ } & H^2(M, \partial M; \mathbb{Q})
\end{array}
$$

where the vertical arrows are the Poincaré and Poincaré-Lefschetz duality isomorphisms. By the universal coefficient theorem we have that $H^1(\partial M; \mathbb{Q})$ is dual to $H_1(\partial M; \mathbb{Q})$, $H_1(M; \mathbb{Q})$ is dual to $H^1(M; \mathbb{Q})$, and i^* is dual to i_*. By the rank-nullity theorem we have

$$2g = \dim \operatorname{im} i_* + \dim \ker i_* = \dim \operatorname{im} i_* + \dim \ker \delta = \dim \operatorname{im} i_* + \dim \operatorname{im} i^* = 2 \dim \operatorname{im} i_*$$

hence $\dim \operatorname{im} i_* = g$, and so $\dim \ker i_* = g$. $\qquad\square$

Lemma 2.4. *There is a basis $\{x_i\}$ over $\mathbb{Z}$ for $H_1(\partial M)$ such that $i_*(x_j) = 0$ for $1 \le j \le g$.*

Proof. We think of $H_1(\partial M) = \mathbb{Z}^{2g}$ as sitting inside $H_1(\partial M; \mathbb{Q}) = \mathbb{Q}^{2g}$. Let U be the g dimensional subspace of $\mathbb{Q}^{2g}$ which is the kernel of the map $\mathbb{Q}^{2g} \to H_1(M; \mathbb{Q})$ induced by the inclusion. We can take a basis $\{x_1, \ldots, x_g\}$ for this vector space such that $x_i \in \mathbb{Z}^{2g}$, and so let $\widetilde{U}$ be the $\mathbb{Z}$-span of these vectors.

As a $\mathbb{Z}$-module we can write $\mathbb{Z}^{2g}/\widetilde{U}$ as $A/\widetilde{U} \oplus B/\widetilde{U}$, where A and B are submodules of $\mathbb{Z}^{2g}$ such that $A/\widetilde{U}$ is free and $B/\widetilde{U}$ is torsion. The fact that this latter module is torsion means that for $b \in B$ there is an integer n such that $nb \in \widetilde{U}$, hence $nb \in U$, or simply $b \in U$. Thus a $\mathbb{Z}$-basis for B also serves as a $\mathbb{Q}$-basis for U. So, the $\mathbb{Z}$-basis we take for $H_1(\partial M)$ is any extension of a basis for B to a full basis for $H_1(\partial M)$. $\qquad\square$

Lemma 2.5. *Let F be a genus g Seifert surface for a slice knot k. Then there is a basis for $H_1(F)$ such that the associated Seifert matrix is of the form*

$$S = \begin{pmatrix} 0 & P \\ Q & R \end{pmatrix}$$

for some $g \times g$ matrices P, Q, R.

Proof. Let D be a slicing disc for k in D^4 so that there is a 3-manifold M with $\partial M = F \cup D$. Let $\{x_i\}$ be a $\mathbb{Z}$-basis for $H_1(\partial M)$ as in the previous lemma so that $x_i = 0$ in $H_1(M; \mathbb{Q})$ for $1 \le i \le g$. We can assume that the homology classes x_i are represented by oriented, closed curves in ∂M, and we will also write x_i to denote these curves. For each $1 \le i \le g$ we can find a non-zero integer n_i such that $n_i x_i = 0$ in $H_1(M)$. The class $n_i x_i$ can be represented by a closed curve in ∂M, which we also denote by $n_i x_i$, and it bounds a 2-chain in M. This means we can find maps $f_i : F_i \to D^4$, where F_i is a surface with boundary, such that $f_i(\partial F_i) = n_i x_i$. If we push the curve $n_i x_i$ off of ∂M in the normal direction, then we can also pushing the mapping f_i in this direction. Therefore, for $1 \le i, j, \le g$ the curves $n_i f_i$ and $(n_j f_j)^+$ bound disjoint surfaces in D^4, and hence

$$0 = \mathrm{lk}(n_i f_i, (n_j f_j)^+) = n_i n_j \, \mathrm{lk}(f_i, f_j^+)$$

Therefore the upper-left $g \times g$ block of the Seifert matrix of F with respect to the basis $\{x_i\}$ is zero. $\qquad\square$

Lemma 2.6. *If Q is a non-degenerate quadratic form on an even dimensional space that vanishes on a subspace of half dimension, then $\sigma(Q) = 0$.*

Proof. Suppose the space V that Q is defined on is of dimension $2g$, and let U be a subspace of dimension g such that $Q(u) = 0$ for all $u \in U$. If $V^\pm$ denote the maximal positive/negative-definite subspaces of V with respect to Q, then $V = V^+ \oplus V^-$ since Q is non-degenerate. However, we also have $U \cap V^+ = \{0\}$ and $U \cap V^- = \{0\}$, hence $\dim V^+ \le g$ and $\dim V^- \le g$. Since we need $\dim V^+ + \dim V^- = 2g$ we are forced to have $g = \dim V^+ = \dim V^-$, which implies $\sigma(Q) = 0$. $\qquad\square$

Proof of Theorem 2.2. Clearly the combination of the previous lemmas proves the theorem. $\qquad\square$

We say that two knots k_1, k_2 in S^3 are **concordant** if there is a smooth embedding of a cylinder $S^1 \times I$ in $S^3 \times I$ such that $S^1 \times \{i\}$ maps to $k_i \times \{i\}$ ($i = 0, 1$). Concordance defines an equivalence relation on the set of knots in S^3. If a knot k is concordant to the unknot, then we can cap off the unknot with a disc and obtain a smoothly embedded disc in D^4 that is bounded by k, hence k is slice. Therefore a knot is concordant to the unknot if and only if it is slice.

Let $\mathcal{C}_1$ denote the concordance classes of knots in S^3. If k_1, k_1' and k_2, k_2' are concordant pairs of knots, then $k_1 \# k_2$ and $k_1' \# k_2$ are concordant. The cobordism can be constructed by splicing

together the cylinders connected k_1 to k_1' and k_2 to k_2'. Next we claim that for any knot k, the knot $k\#\overline{k}$ is slice. Define the following subsets of $\mathbb{R}^4$:

$$\mathbb{R}^3 = \big\{(x,y,z,0) \in \mathbb{R}^4\big\}$$

$$\mathbb{R}^3_- = \big\{(x,y,z,0) \in \mathbb{R}^4 : z \leq 0\big\}$$
$$\mathbb{R}^3_+ = \big\{(x,y,z,0) \in \mathbb{R}^4 : z \geq 0\big\}$$
$$\mathbb{R}^2 = \big\{(x,y,0,0) \in \mathbb{R}^4\big\}$$

Arrange the knot $k\#\overline{k}$ in $\mathbb{R}^3$ so that its intersection with the plane $\mathbb{R}^2$ in $\mathbb{R}^4$ consists of two points on the band used in the connect sum. Let k^+ denote the intersection of $k\#\overline{k}$ with $\mathbb{R}^3_+$, which is just an arc, so that if k_- is the reflection of k_+ through $\mathbb{R}^2$, then $k_+ \cup k_-$ is isotopic to $k\#\overline{k}$. Consider the set of points

$$D = \big\{(x,y,z\cos\theta, z\sin\theta) \in \mathbb{R}^4 : (x,y,z,0) \in k_+, 0 \leq \theta \leq \pi\big\}$$

That is, we spin all the points on k_+ about the plane $\mathbb{R}^2$ in $\mathbb{R}^4$ through all angles from 0 to π. Clearly D is just a disc (being a half spin of an arc) with boundary $k\#\overline{k}$, and this provides us with a slicing disc for $k\#\overline{k}$.

We can now conclude that $\mathcal{C}_1$ forms a group under connect sum with [unknot] serving as the identity and the mirror of a knot serving as its inverse. Further, if k_1 and k_2 are concordant, then $0 = \sigma(k_1\#\overline{k}_2) = \sigma(k_1) - \sigma(k_2)$, hence σ descends to a well-defined group homomorphism $\sigma : \mathcal{C}_1 \to \mathbb{Z}$.

2.4   Cyclic and Branched Covers

Now let us discuss Alexander's original definition of his polynomial using the infinite cyclic cover of the complement of a link. To construct this covering space we fix an oriented link L, set $X = S^3 \backslash L$, and fix a Seifert surface F (compatible with the orientation of L). Let Y be the space obtained from cutting X along F so that ∂Y consists of two disjoint copies of F, denoted by F_+ and F_-; that is, Y is homeomorphic to the complement of an open tubular neighborhood of F in S^3. There is a natural identification $\phi : F_- \to F_+$. Take the countable collection $Y \times \{i\}$ $(i \in \mathbb{Z})$ of copies of Y, and let $h_i : Y \to Y_i$ be the obvious homeomorphism. Form the space X_∞ by gluing Y_i to Y_{i+1} via the homeomorphism $h_{i+1} \circ \phi \circ h_i^{-1}$; that is, we glue Y_i to Y_{i+1} by gluing $F_- \times \{i\}$ to $F_+ \times \{i+1\}$ via the map $h_{i+1} \circ \phi \circ h_i^{-1}$. The obvious projection $\pi : X_\infty \to X$, where $\pi|_{Y_i} : Y_i \to X$ is given by $\pi(x,i) = x$, is an infinite cyclic covering space of X.

Let t denote the covering transformation of X_∞ whose restriction $t|_{Y_i} : Y_i \to X_\infty$ is given by $t(x,i) = (x, i+1)$. This map induces an isomorphism (which we also denote by t) $t : H_1(X_\infty) \to H_1(X_\infty)$, turning $H_1(X_\infty)$ into a $\mathbb{Z}[t,t^{-1}]$-module, called the Alexander module of L. Of course, we need to show that this module is independent of the various choices in its construction

Proposition 2.10. *If X_∞ and X_∞' are the infinite cyclic coverings of the link complement X of a link L constructed from two different Seifert surfaces, then X_∞ and X_∞' are isomorphic has covering spaces, and $H_1(X_\infty)$ and $H_1(X_\infty')$ are isomorphic as $\mathbb{Z}[t,t^{-1}]$ modules.*

Proof. Recall that covering spaces are determined up to equivalence by the image of the fundamental group of the total space under the covering projection. That subgroup of the fundamental group of the base space consists of those loops in the base that lift to loops in the total space. So, we show that the notion of lifting loops to loops does not depend on the Seifert surface F used in the construction of X_∞.

20

Let $\alpha : [0,1] \to X$ be any path in X and consider a lift $\tilde{\alpha} : [0,1] \to X_\infty$. This path is a loop in X_∞ if and only if $\tilde{\alpha}(0)$ and $\tilde{\alpha}(1)$ are both in X_∞, which means the path intersects the chambers $F_\pm \times \{i\}$ with algebraic multiplicity zero. Down in X this means that that α has zero algebraic intersection number with F, which we know is equivalent to α and L having zero linking number (which is the sum of the linking numbers with each component of L). This statement is of course independent of F, hence our notion of path lifting is independent of F, and so these covering spaces are equivalent.

Next we have to check that any two coverings are equivalent via a t-equivariant homeomorphism in order to ensure that the first homology $\mathbb{Z}[t, t^{-1}]$-modules are isomorphic. Suppose $p' : X'_\infty \to X$ is another covering space constructed with a different Seifert surface F'. Then there is a homeomorphism $h : X_\infty \to X'_\infty$ such that $p' \circ h = p$. Let α be a path in X_∞ from a point a to the translate $t \cdot a$. In the base, the path $p \circ \alpha$ is a loop with linking number 1 with L. Then $p' \circ h \circ \alpha = p \circ \alpha$, so $h \circ \alpha$ is a path in X'_∞ from a point a' to its translate $t \cdot a'$, hence $t \cdot h(a) = h(t \cdot a)$ and so h is t-equivariant. $\qquad\square$

Since the covering space $X_\infty \to X$ is well-defined given a link L we have that the $\mathbb{Z}[t, t^{-1}]$-module $H_1(X_\infty)$ is an invariant of the link, called the **Alexander module**. Let us connect this invariant with the Alexander polynomial defined earlier. Given an R-module M, a presentation of M is an exact sequence

$$F \xrightarrow{\alpha} E \xrightarrow{\phi} M \to 0$$

where F and E are free R-modules. If we fix bases for F and E then we get a matrix representation of α, called the presentation matrix of M. The r-th elementary ideal of M is the ideal of R generated by all $(m - r + 1) \times (m - r + 1)$ minors of a presentation matrix of M. In particular, the first elementary ideal of M is simply the determinant of any presentation matrix of M. The relationship between the Alexander polynomial and Alexander module can be stated as the following.

Theorem 2.3. *If S is a Seifert matrix for a knot k in S^3, then $tS - S^T$ is a presentation matrix for $H_1(X_\infty)$ as a $\mathbb{Z}[t, t^{-1}]$-module.*

Proof. In the notation we used when constructing X_∞, let us write $X_\infty = U \cup V$, where $U = \cup_i Y_{2i}$ and $V = \cup_i Y_{2i+1}$. We have $H_0(U \cap V) = \oplus_i H_0(F_i)$, and the action of t on this group shifts the summands of $\mathbb{Z}$ by one, so we can identify $H_0(U \cap V)$ with $\mathbb{Z}[t, t^{-1}] \otimes H_0(F)$ generated by $1 \otimes 1$. Similarly, $H_0(U) \oplus H_0(V) = \oplus_i H_0(Y_i)$ and so we can identify this with $\mathbb{Z}[t, t^{-1}] \otimes H_0(Y)$ generated by $1 \otimes 1$. With these identifications we see that the map $H_0(U \cap V) \to H_0(U) \oplus H_0(V)$ in the Mayer-Vietoris sequence is given by $1 \otimes 1 \mapsto -1 \otimes 1 + t \otimes 1$, and so is injective. Therefore the map $H_1(U) \oplus H_1(V) \to H_(X_\infty)$ is the Mayer-Vietoris sequence is surjective.

Let is now fix a basis $\{x_i\}$ for $H_1(F)$ and a dual basis $\{y_i\}$ with respect to the bilinear form $\beta : H_1(S^3 \backslash F) \times H_1(F) \to \mathbb{Z}$. Let $S = (s_{ij})$ be the Seifert matrix of F with respect to this basis. Similar to the above arguments we have an identification of $H_1(U \cap V)$ with $\mathbb{Z}[t, t^{-1}] \otimes H_1(F)$, the latter being generated by $\{1 \otimes x_i\}$, and an identification of $H_1(U) \oplus H_1(V)$ with $\mathbb{Z}[t, t^{-1}] \otimes H_1(Y)$, the latter being generated by $\{1 \otimes y_i\}$. With these identifications, the map $H_1(U \cap V) \to H_1(U) \oplus H_1(V)$ in the Mayer-Vietoris sequence is given by $1 \otimes x_i = -1 \otimes x_i^- + t \otimes x_i^+$. However, as elements in $H_1(S^3 \backslash F)$, the cycles x_i^+ and x_i^- can be written as

$$x_i^+ = \sum_j s_{ji} y_j$$

$$x_i^- = \sum_j s_{ij} y_j$$

21

Therefore

$$-1 \otimes x_i^- + t \otimes x_i^+ = \sum_j (-1 \otimes s_{ij} y_j + t \otimes s_{ji} y_j)$$

With, with respect to our bases we have that the matrix of the map in the Mayer-Vietoris sequence is given by $tS - S^T$. □

We can change the construction of the infinite cyclic cover of the link exterior slightly to obtain a finite cyclic cover. For an integer $k > 0$, take k copies of Y (the link exterior cut open along a Seifert surface) by setting $Y_i = Y \times \{i\}$ ($i = 0, 1, \ldots, k-1$). For the space X_k by gluing Y_i to Y_{i+1} via the homeomorphism $h_{i+1} \circ \phi \circ h_i^{-1}$ (where all induces are taken modulo k). Then the obvious projection $\pi_k : X_k \to X$ is a k-fold cyclic covering of the link exterior. Note that we have an infinite cyclic covering $X_\infty \to X_k$ that maps the points in $Y_i \subset X_\infty$ identically onto $Y_{i \bmod k}$. In fact, the space X_k is simply the quotient $X_\infty / \langle t^k \rangle$.

This collection of covering spaces depends only on the link L and not on the various choices used to define it.

This collection of covering spaces can now be extending to cyclic branched covers $\tilde{X}_k \to S^3$ of the 3-sphere, branched over the link L. A loop in X lifts to a loop in X_k if and only if it has linking number 0 (modulo k) with L. For a component L_i from the link L, let νL_i be a tubular neighborhood. We can identify $\partial \nu L_i$ with $S^1 \times S^1$ by choosing a meridian μ_i and longitude ℓ_i of $\partial \nu L_i$ such that ℓ_i links zero times with L_i. Then μ_i^k (the k-th power of μ_i) and ℓ_i lift to a loops in X_k which identify a torus in X_k. The restriction of the covering map to this torus gives a k-fold covering of one torus over another which is equivalent to $(z, w) \mapsto (z, w^k)$. We can extend this to a branched covering $S^1 \times D^2 \to S^1 \times D^2$ given by $(z, w) = (z, w^k)$, where this is branched over the core circle $S^1 \times 0$. So, if we glue solid tori into the boundary components of X_k we obtain a space $\tilde{X}_k$ with a natural projection $\tilde{X}_k \to S^3$ that is a k-fold covering branched over L.

2.5 Applications

Seifert matrices are closely related to the intersection form on a compact, orientable surface. Intersection forms are defined on any even dimensional manifold, and we give a fuller treatment for 4-manifolds in section 6.1, so we will just briefly state the facts. For a closed, oriented surface M, Poincaré duality gives us a non-degenerate, antisymmetric, bilinear form $Q_M : H^1(M) \otimes H^1(M) \to \mathbb{Z}$ defined by $Q_M(\alpha, \beta) = (\alpha \smile \beta, [M])$. It follows that the matrix representation of Q_M in any basis of $H^1(M)$ has determinant ± 1. However, we have the following lemma concerning these types of matrices.

Lemma 2.7. *If A is the matrix representation of a non-degenerate, antisymmetric bilinear form, then there is a matrix U such that $A = U^T J U$, where J is a block diagonal matrix of copies of*

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

Therefore the matrix of Q_M with respect to any basis must have determinant $+1$. There is a more geometric definition of the intersection form on M. A homology class $a \in H_1(M)$ can be represented by a closed, smoothly embedded 1-dimensional submanifold of M, call it C_a. If $a, b \in H_1(M)$ with representative submanifolds C_a and C_b, then perturb them a little so that they intersect transversely, and let $a \cdot b$ be the signed sum of their points of intersection, called the intersection product. Then one can show that $Q_M(\alpha, \beta) = PD(\alpha) \cdot PD(\beta)$.

If M has one boundary component (i.e. M is a Seifert surface of some knot), then Q_M is still a non-degenerate, antisymmetric, bilinear form, and it can still be expressed in terms of intersections of homology classes. If M has more than one boundary component, then Q_M is an antisymmetric, bilinear form, and can be defined in terms of intersections, but it is now necessarily degenerate, and so any matrix representation has zero determinant.

Proposition 2.11. *Let F be a Seifert surface of a knot k with Seifert matrix S in some basis of $H_1(F)$. Then the intersection form of F in this basis is precisely $S - S^T$.*

Proof. At a positive crossing of x_i and x_j we see that x_j^+ is above x_i. So we have $\mathrm{lk}(x_i, x_j^+)$ contributes $+1$ at this crossing and $\mathrm{lk}(x_j, x_i^+)$ does not contribute anything (since we only count crossings where the first strand passes under). On the other hand, at a negative crossing of x_i and x_j we have that x_j^+ passes below x_i. Hence $\mathrm{lk}(x_i, x_j^+)$ contributes nothing at this crossing and $\mathrm{lk}(x_j, x_i^+)$ contributes $+1$. Therefore $x_i \cdot x_j = \mathrm{lk}(x_i, x_j^+) - \mathrm{lk}(x_j, x_i^+)$. $\qquad\square$

Corollary 2.2. *If k is a knot, then $\Delta_k(1) = 1$, and if L is a link, then $\Delta_L(1) = 0$.*

Proof. Let F_k and F_L be Seifert surfaces for k and L. By proposition 2.11 we have $\Delta_k(1) = \det Q_M = 1$ and $\Delta_L(1) = \det Q_L = 0$. $\qquad\square$

The Alexander polynomial of the connected sum of two knots has a very simple expression.

Proposition 2.12. *If k_1 and k_2 are knots, then $\Delta_{k_1 \# k_2}(t) = \Delta_{k_1}(t) \cdot \Delta_{k_2}(t)$.*

Proof. We can assume that k_1 and k_2 are situated in S^3 far apart, and that F_1 and F_2 are Seifert surfaces for k_1 and k_2 respectively that do not intersect. Then by taking the boundary sum $F_1 \natural F_2$ we get a Seifert surface for $k_1 \# k_2$. If S_1 and S_2 are the Seifert matrices for F_1 and F_2 respectively, then the Seifert matrix of $F_1 \natural F_2$ is

$$\begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix}$$

We clearly now have

$$\Delta_{k_1 \# k_2}(t) = \det(t^{1/2} S_1 - t^{-1/2} S_1^T) \cdot \det(t^{1/2} S_2 - t^{-1/2} S_2^T) = \Delta_{k_1}(t) \cdot \Delta_{k_2}(t)$$

$\qquad\square$

The Alexander polynomial has a clear relationship with the genus of a knot.

Proposition 2.13. *For a knot k we have $\deg \Delta_k(t) \leq g(k)$.*

Proof. If S is a Seifert matrix of order $2m$ for k, then $\deg \Delta_k(t) \leq m$, hence $\deg \Delta_k(t) \leq g(k)$. $\qquad\square$

Proposition 2.14. *If $\det S \neq 0$ for some Seifert matrix S of k, then $\deg \Delta_k(t) = g(k)$.*

Proof. If S is a Seifert matrix associated to a Seifert surface of genus g, then we can write $\Delta_k(t) = t^{-g} \det(S - tS^T)$. The constant term of the polynomial $\det(S - tS^T)$ is just $\det S$, which is non-zero by assumption. Therefore the coefficient of t^{-g} in $\Delta_k(t)$ is non-zero, and so $\deg \Delta_k(t) = g$. $\qquad\square$

A link L is said to be a **split link** if one can find an embedded 2-sphere in S^3 that separates some components of L from the other components of L.

Proposition 2.15. *If L is an oriented, split link L, then $\Delta_L(t) = 0$.*

Proof. Suppose $L = k_1 \cup k_2$ such that k_1 can be separated from k_2 by an embedded 2-sphere. Let F_1 and F_2 be Seifert surfaces for k_1 and k_2 respectively, and let g_1 and g_2 be the genera of F_1 and F_2 respectively. We can constructed a Seifert surface for L by taking the connected sum $F_1 \# F_2$; that is, cut small, open discs out of F_1 and F_2 and glue in a tube $S^1 \times I$. Recall that a surface of genus g with n boundary components has first homology of rank $2g + n - 1$, hence $H_1(F_1) \cong \mathbb{Z}^{2g_1}$ and $H_2(F_2) \cong \mathbb{Z}^{2g_2}$. However, $H_1(F_1 \# F_2) \cong \mathbb{Z}^{2g_1 + 2g_2 + 1}$, so this homology group has one extra generator in addition to the generators coming from F_1 and F_2. This generator can be represented by the meridian μ of the tube $S^1 \times I$. This generator does not link with the push offs of any of the generators from F_1 and F_2, hence there is an entire row and column of zeros in the Seifert matrix of $F_1 \# F_2$, and so $\Delta_L(t) = 0$. $\qquad \square$

2.6   Knots in other 3-Manifolds

Much of what we have said so far can be generalized to 3-manifolds other than S^3. For example, a knot k in a 3-manifold M is just a smoothly embedded copy of S^1 into M. Unfortunately, there is no analogous concept of diagram of knots in a general manifold, but any term we defined before without the use of a diagram can be defined for knots in a manifold. We will work towards extending other concepts and results to knots in a manifold, but almost always we are going to need to require that the manifold is an integral homology sphere so that we have the following property.

Proposition 2.16. *If M is a homology sphere and k an embedded knot, then $M \backslash k$ is a homology circle, that is $H_*(M \backslash k) = H_*(S^1)$.*

Proof. The proof is the same as in proposition 2.1. $\qquad \square$

Our first definition of linking number of two oriented, disjoint knots needed a knot diagram, so we cannot use this definition in a manifold. However, if M is an integral homology sphere, and k_1 and k_2 knots in M, proposition 2.16 says that $H_1(M \backslash N(k_1)) = \mathbb{Z}$ is generated by $[\mu]$, where μ is any meridian of $N(k)$ (which is just the boundary of any normal cross-section). So, as we did before, we define the linking number $\mathrm{lk}(k_1, k_2)$ to be the integer such that $[k_2] = \mathrm{lk}(k_1, k_2) \cdot [\mu]$. We can also define the canonical longitude ℓ of a knot k in M to be the closed curve in $\partial(M \backslash N(k))$ such that $\mathrm{lk}(k, \ell) = 0$.

Let k be a knot in M with canonical longitude ℓ, and orient k and ℓ such that the basis (k, ℓ, n) is the same as the orientation on M (M is orientable since it is an integral homology sphere), where n is the inward point normal vector. The choice of meridian μ and longitude ℓ gives an identification $\varphi : \partial(M \backslash N(k)) \to S^1 \times S^1$. Let $p_0 : \partial(M \backslash N(k)) \to S^1$ be $p_0 = \pi_{S^1} \circ \varphi$, where π_{S^1} is projection onto the second factor.

Proposition 2.17. *Let $K = M \backslash N(k)$ be the knot complement. The projection $p_0 : \partial K \to S^1$ extends to a map $p : K \to S^1$.*

Proof. Let $i : \partial K \to K$ be the inclusion. We claim that p_0 extends to p if and only if the homotopy class $[p_0]$ is in the image of the induced map $i^* : [K, S^1] \to [\partial K, S^1]$. Clearly if such an extension exists then $[p_0] = i^*[p]$. Conversely, suppose there is some homotopy class of maps $[p]$ such that $[p_0] = i^*[p]$. This means that p_0 and $p \circ i$ are homotopic on ∂K, but we need them to be *equal.* Let H be a homotopy between $p \circ i$ and p_0 and let $U \subseteq M$ be a collar of ∂K in K, i.e. U is diffeomorphic to $\partial K \times [0, 1]$ such that $\partial K \times \{0\}$ is mapped identically to $\partial K \subset U$. If we define $p' : K \to S^1$ to be p outside U and $H(x, t)$ on the part of U corresponding to $(x, t) \in \partial K \times I$, then we see that p' is an extension of p_0.

In appendix C we saw that we have a natural isomorphism of groups

$$[\partial K, S^1] \cong H^1(\partial K)$$

since S^1 is a $K(\mathbb{Z}, 1)$ space. The inclusion i induces a map on cohomology, and by naturality of the above isomorphism, and naturality of the Poincaré isomorphisms we get the following commutative diagram from the long exact sequence of the pair $(K, \partial K)$

$$
\begin{array}{ccc}
H_1(\partial K) & \xrightarrow{\ i_* \ } & H_1(K) \\
\Big\uparrow{\scriptstyle PD} & & \Big\uparrow{\scriptstyle PD} \\
H^1(\partial K) & \xrightarrow{\ \delta \ } & H^2(K, \partial K)
\end{array}
$$

where δ is the connecting homomorphism. We have that $[\ell]$ is the Poincaré dual of $[p_0]$, so by commutativity of the above square we have $\delta[p_0] = PD^{-1}i_*[\ell] = 0$, since $[\ell]$ is the canonical longitude, hence $[p_0] \in \ker \delta$. But, by exactness we have $[p_0] \in \ker \delta = \operatorname{im} i^*$, hence p_0 extends to a map p by our previous claim. $\qquad \square$

This proposition can easily be extended to links. One might wonder if the map p_0 is a fibration of the knot complement over S^1. This is not always true, so it leads to the notion of a "fibered" knot, and will be discussed later. We can use proposition 2.17 to show that Seifert surfaces exist for links in integral homology spheres.

Proposition 2.18. *Given a link L in an integral homology sphere M there is a smoothly embedded, compact, orientable and connected surface F such that $\partial F = L$.*

Proof. Let $p : K \to \partial D^2$ be the map constructed in proposition 2.17, where K is the link complement in M. We can slightly homotopy p such that it is transverse to a point $* \in \partial D^2$, in which case $F' = p^{-1}(*)$ is a smoothly embedded submanifold of K of codimension 1 (by proposition 1.1) with boundary contained in ∂K, hence F is a surface. We can connect the boundary of F' to L via annuli contained in the tubular neighborhoods of the link components to get a Seifert surface F. $\qquad \square$

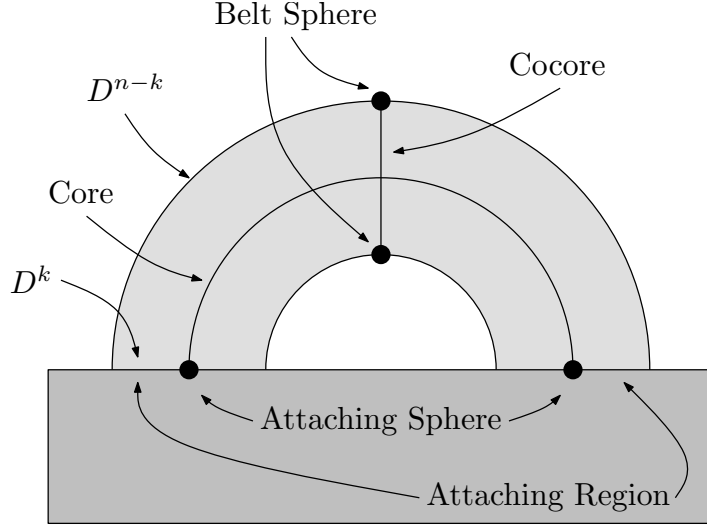| Handle Anatomy	
Attaching map	$\varphi : \partial D^k \times D^n \to \partial X$
Attaching region	$\partial D^k \times D^{n-k}$
Attaching sphere	$\partial D^k \times 0$
Belt sphere	$0 \times \partial D^{n-k}$
Core	$D^k \times 0$
Cocore	$0 \times D^{n-k}$

Figure 3.1: Anatomy of an n-dimensional k-handle

3 Handlebodies

3.1 Handlebody Decompositions

Let X be a smooth n-dimensional manifold with boundary. The process of attaching an n-dimensional k-handle to X is to form the quotient space $X \cup_\varphi h$, where h is a copy of $D^k \times D^{n-k}$, and where $\varphi : \partial D^k \times D^{n-k} \to \partial X$ is a smooth embedding. Notice that $\partial D^k \times D^{n-k}$ is only part of the boundary of $D^k \times D^{n-k}$. Also, the manifold $X \cup_\varphi h$ is not smooth, as it has corners. However, there is a canonical way to smooth the corners, so we may assume the quotient space is a smooth n-manifold with boundary. If two attaching maps $\varphi, \varphi' : \partial D^k \times D^{n-k} \to \partial X$ are isotopic, then the resulting manifolds $X \cup_\varphi h$ and $X \cup_{\varphi'} h$ are diffeomorphic. Also note that there is an obvious deformation retract of $D^k \times D^{n-k}$ onto $D^k \times 0$, hence $X \cup_\varphi h$ deformation retracts onto $X \cup_\varphi D^k \times 0$ (where φ is suitably restricted in the second space), hence attaching a k-handle is the same as attaching a k-cell.

The anatomy of a handle is simple and the pieces have intuitive names. The map φ is called the **attaching map**, while $\partial D^k \times D^{n-k}$ is called the **attaching region** (and sometimes the image of this set under φ). The subset $D^k \times 0$ is called the core of the handle, and $0 \times D^{n-k}$ the **cocore**. The region $\partial D^k \times 0$ the **attaching sphere**, and $0 \times \partial D^{n-k}$ the **belt sphere**. Finally, the number k is called the **index** of the handle. See figure 3.1.

Unfortunately this picture is a little misleading since attaching a k-handle does not always look like attaching a ball along two disconnected pieces of its boundary. For example, attaching

a 4-dimensional 2-handle means attaching a copy of $D^2 \times D^2$ along the solid torus $S^1 \times D^2$. The boundary of $D^2 \times D^2$ is a union of solid tori $S^1 \times D^2 \cup D^2 \times S^1$ that intersect at their boundaries $S^1 \times S^1$, and this decomposition of S^3 into two solid tori is clearly the genus 1 Heegaard splitting.

An embedding $\varphi : \partial D^k \times D^{n-k} \to \partial X$ is the same as embedding the trivial normal bundle of the attaching sphere $\varphi(\partial D^k \times 0)$ into ∂X. Therefore φ determines, and is determined by, an embedding $\varphi_0 : \partial D^k \times 0 \to \partial X$ of the attaching sphere and an equivalence f of the normal bundle $\nu\varphi_0(\partial D^k)$ with the trivial vector bundle $\partial D^k \times \mathbb{R}^{n-k}$; this is intuitively plausible, but is also the content of the Tubular Neighborhood Theorem. The embedding φ_0 is just a **knot** in ∂X, and the vector bundle equivalence f is called a **(normal) framing** of $\partial D^k = S^{k-1}$. If the data (φ_0, f) and (φ_0', f') are smoothly isotopic, then the resulting manifolds $X \cup_\varphi h$ and $X \cup_{\varphi'} h$ are isotopic.

Fix a framing f_0 for S^{k-1}, i.e. a vector bundle equivalence

$$
\begin{array}{ccc}
S^{k-1} \times \mathbb{R}^{n-k} & \xrightarrow{\ f_0\ } & \nu\varphi_0(S^{k-1}) \\
\downarrow & & \downarrow \\
S^{k-1} & \xrightarrow[\ \varphi_0\ ]{} & \varphi_0(S^{k-1})
\end{array}
$$

If f is another framing of S^{k-1}, then $f^{-1} \circ f_0 : S^{k-1} \times \mathbb{R}^k \to S^{k-1} \times \mathbb{R}^{n-k}$ must map (p, v) to $(p, g(p) \cdot v)$, where $g : S^{k-1} \to GL(n-k)$ is smooth. Without loss of generality we can assume that $g(p) = I$ for a fixed base point $p \in S^{k-1}$. Therefore each framing determines an element of the group $\pi_{k-1}(GL(n-k))$, which is isomorphic to $\pi_{k-1}(O(n-k))$ by the deformation retract of $GL(n-k)$ onto $O(n-k)$, and it is even true that the set of framings are in one-to-one correspondence with elements of $\pi_{k-1}(O(n-k))$. However, the set of framings is not canonically isomorphic to $\pi_{k-1}(O(n-k))$, since the identification depends on f_0, but it is a principal homogeneous space for $\pi_{k-1}(O(n-k))$. Sometimes there is a canonical choice of f_0, and this will be useful later.

We can get a lot of information about attaching very low index handles and very high index handles. The attaching sphere for a 0-handle $D^0 \times D^n$ is empty, so attaching a 0-handle to X is the same as simply taking the disjoint union of X with D^n. The attaching sphere of a 1-handle is a disjoint pair of points, and so if ∂X is connected and non-empty there is a unique isotopy class of attaching maps (well, except in the case $n = 2$ and ∂X non-compact, in which case we can also interchange the two components of the attaching sphere). Further, since $\pi_0(O(n-1)) = \mathbb{Z}/2$ for $n \geq 2$ there are exactly two framings on S^0, hence there are exactly two manifolds one can obtain by attaching a 1-handle to a 0-handle (these are the n-dimensional analogs of annulus and Möbius band).

Similarly, for $(n-1)$-handles, with $n \neq 2$, there is a unique framing of S^{n-2} since $\pi_{n-2}(O(1)) = 0$. In the case $n = 2$ we see that $(n-1)$-handles are just 1-handles, which were discussed above. Attaching an n-handle is the same as gluing a copy of D^n to ∂X along $\partial D^n = S^{n-1}$. Since ∂X is $(n-1)$-dimensional we see that we can only attach D^n to boundary components of X that are diffeomorphic to S^{n-1}. However, due to issues of exotic n-spheres and exotic diffeomorphisms of n-spheres we can only say that attaching an n-handle is unique in dimensions ≤ 6.

The other combinations of k and n, where k is neither small nor close to n, are harder to understand. If $n - k > 2$, then there is only one isotopy class of embeddings S^{k-1} into $(n-1)$-dimensional manifolds, hence all embeddings can be "unknotted." So, the only thing that controls what we get from attaching a k-handle to a 0-handle is the framing of an unknotted sphere in S^{n-1}, which are of course parameterized by $\pi_{k-1}(O(n-k))$. Suppose we are attaching a k-handle $h_2 = D^k \times D^{n-k}$ to a 0-handle $h_1 = D^n$ such that the attaching sphere is the standardly embedded, unknotted sphere $S^{k-1} \subset S^{n-1}$. Without loss of generality we can assume that we have

an identification $h_1 = D^k \times D^{n-k}$ and that the attaching map of h_1 is of the form $\varphi : \partial D^k \times D^{n-k} \to \partial D^k \times D^{n-k} \subset D^n$ such that $\pi_{\partial D^k} \circ \varphi$ is the identity, and the restriction of φ to any slice $\{*\} \times D^{n-k}$ is an orthogonal linear map. Let X be the manifold $h_1 \cup_\varphi h_2$. We claim that there is a canonical map $\pi : X \to S^k$ making X into a D^{n-k}-bundle. Decompose S^k as a union of two discs, the upper and lower hemispheres. Then π projects the handle h_1 onto its its first factor, which is then identified with the upper hemisphere, and π projects h_2 onto its first factor, which is then identified with the lower hemisphere. It is clear that this makes X into a D^{n-k}-bundle, and in fact all D^{n-k}-bundles are of this form. Further, consider the framing f_0 corresponding to $\mathrm{id}_{\partial D^k \times D^{n-k}}$. This clearly gives the trivial bundle $X = S^k \times D^{n-k}$, and so this is a canonical element to associate to the identity in $O(n-k)$.

If we apply the above discussion to the case $k = 2$, then we have that D^{n-2}-bundles over S^2 are classified by $\mathbb{Z}$ if $n = 4$ and by $\mathbb{Z}/2$ if $n > 4$. In this case we can get a canonical identification of these bundles with $\mathbb{Z}$ ($n = 4$) and $\mathbb{Z}/2$ ($n > 4$) by associating the trivial bundle with 0. For $n = 4$, the integer associated to a D^2 bundle over S^2 is called the Euler number of the bundle. If $n > 4$, then we denote the non-trivial D^{n-2} bundle over S^2 by $D^{n-2} \tilde{\times} S^2$, and we also define the associated S^{n-3} bundle over S^2 by $S^{n-3} \tilde{\times} S^2 = \partial(D^{n-2} \tilde{\times} S^2)$.

Let X be a compact n-manifold with boundary written as $\partial_+ X \sqcup \partial_- X$. If X is oriented give $\partial_+ X$ the induced orientation and $\partial_- X$ the opposite of the induced orientation. A **handlebody decomposition** of X is an identification of X with an n-manifold obtained by attaching handles to $\partial_- X \times I$ such that $\partial_- X$ is identified with $\partial_- X \times 0$. A manifold with such a decomposition is called a **relative handlebody** built on $\partial_- X$, or if $\partial_- X = \emptyset$ then X is simply called a **handlebody**.

The theory of handlebodies and Morse theory are closely related, so it is hard to discuss one without the other. Unfortunately we have now come to a point where we need to assume the reader has some knowledge of Morse theory. One can read this section concurrently with section 4.

According to the results in section 4 we have that every smooth, compact n-manifold has a handlebody decomposition using Morse theory, and this can even be extended to non-compact manifolds with compact boundary. Amazingly every topological n-manifold with $n \neq 4$ also admits a handlebody decomposition, but this is not true in dimension $n = 4$. For example, there is a 4-manifold with intersection form E_8 which does not admit a smooth structure or a handlebody decomposition.

Proposition 3.1. *Any handlebody decomposition of a pair $(X, \partial_- X)$ can be modified (by isotoping attaching maps) so that handles are attached in order of increasing index, and handles of the same index can be attached in any order or simultaneously.*

Proof. Suppose we attach a k-handle h_1 and then an ℓ-handle h_2 to a manifold X, with $\ell \leq k$. First we will perturb the attaching map of h_2 slightly so that its attaching sphere does not intersect h_1's belt sphere (if needed). Note that h_1's belt sphere, $0 \times \partial D^{n-k}$, is of dimension $n - k - 1$ and h_2's attaching sphere, $\partial D^\ell \times 0$, is of dimension $\ell - 1$. The codimension of the intersection of generically positioned submanifolds is the sum of the submanifolds' codimension, so the codimension of h_1's belt sphere intersected with h_2's attaching sphere, in ∂X, is

$$(n-1) - (n-k-1) + (n-1) - (\ell-1) = n - 1 - n + k + 1 + n - 1 - \ell + 1 = n + k - \ell \geq n$$

Hence we can perturb the attaching sphere of h_2 slightly so that the spheres become entirely disjoint. Now take a smooth vector field on $X \cup h_1$ emanating from the cocore of h_1 radially, and flow the attaching sphere of h_2 along this vector field. This will push the attaching sphere of h_2 to ∂X, so we can first attach h_2 and then h_1. $\qquad \square$

We will always assume that our handlebodies are ordered so that handles are attached with increasing index, and we will write X_k to denote $\partial_- X \times I$ union all the handles of index $\leq k$.

A handlebody decomposition of a pair $(X, \partial_- X)$ naturally leads to a dual handlebody decomposition of $(X, \overline{\partial_+ X})$, where k-handles in the first decomposition become $(n-k)$-handles in the second decomposition. When we attach a k-handle h to X, the points in $\partial D^k \times D^{n-k}$ become interior points of $X \cup h$ and the points in $D^k \times \partial D^{n-k}$ become boundary points in $X \cup h$. Other handles will be attached to the part $D^k \times \partial D^{n-k}$ of h, so we could reverse this and instead think of h has an $(n-k)$-handle being attached to those other handles. Intuitively, we are turning the handlebody upside down. If the handlebody decomposition is given by a Morse function f, then the dual handlebody is just the handlebody given by the Morse function $1 - f$.

Example 3.1. We will show how $\mathbb{C}P^n$, complex projective n-space, can be decomposed into a handlebody with one $2k$-handle for each $0 \leq k \leq n$. Recall that $\mathbb{C}P^n$ is the quotient of $\mathbb{C}^{n+1} - 0$ by $\mathbb{C}^*$ acting by scalar multiplication. The equivalence class of $(z_0, \ldots, z_n)$ is denoted by $[z_0 : \cdots : z_n]$. Complex projective space is an n-dimensional complex manifold (or $2n$-dimensional real manifold) that can be covered with $n + 1$ charts $\varphi_i : \mathbb{C}^n \to \mathbb{C}P^n$ defined by

$$\varphi_i(z_0, \ldots, z_{n-1}) = [z_0 : \cdots : z_{i-1} : 1 : z_i : \cdots : z_{n-1}]$$

The inverses $\phi_i : U_i \to \mathbb{C}^n$ of these charts are given by

$$\phi_i([z_0, \ldots, z_n]) = (z_0/z_i, \ldots, z_{i-1}/z_i, z_{i+1}/z_i, \ldots, z_n/z_i)$$

where U_i consists of all points $[z_0, \ldots, z_n]$ such that $z_i \neq 0$.

Let D be the unit disc in $\mathbb{C}$, and and let $B_i = \varphi_i(D \times \cdots \times D)$ (an embedded complex n-ball in $\mathbb{C}P^n$). Every point $p \in \mathbb{C}P^n$ has a unique representative $(z_0, \ldots, z_n)$ such that $|z_i| = 1$ for some $0 \leq i \leq n$ and $|z_j| \leq 1$ for all j (just divide by the coordinate with largest norm), which we will call the canonical representative of p. It follows that $\{B_i\}$ covers $\mathbb{C}P^n$. We also see that $p \in B_i$ if and only its canonical representative has $|z_i| = 1$, and $p \in \partial B_i$ if and only if $|z_j| = 1$ for some other $j \neq i$. From this it follows that two balls B_i and B_j can intersect only at points in the boundaries of each ball: $p \in B_i \cap B_j$ if and only if its canonical representative has $|z_i| = |z_j| = 1$ and $|z_k| \leq 1$ for all k. In particular, then ball B_k intersects the union $\cup_{i<k} B_i$ at all points whose canonical representative has $|z_k| = 1$, $|z_j| = 1$ for some $0 \leq j < k$ and $|z_\ell| \leq 1$ for all ℓ. This means that the restriction of φ_k to $\partial(D \times \cdots \times D) \times D \times \cdots \times D \subset D^k \times D^{n-k}$ is an embedding into $\cup_{i<k} B_i$, therefore $\cup_{i\leq k} B_i$ can be obtained from $\cup_{i<k} B_i$ by attaching a $2k$-handle via φ_k.

By replacing $\mathbb{R}$ for $\mathbb{C}$ in the above construction we get a decomposition of $\mathbb{R}P^n$ into a handlebody with one k-handle for each $0 \leq k \leq n$. We are going to explicitly compute the attaching maps for $\mathbb{R}P^1, \mathbb{R}P^2, \mathbb{C}P^1$ and $\mathbb{C}P^2$. From the description above we have that $\mathbb{R}P^1$ is constructed by gluing $B_1 = \{[x, 1] : -1 \leq x \leq 1\}$ to $B_0 = \{[1, x] : -1 \leq x \leq 1\}$ via the map φ_1 restricted to $\partial D \subset \mathbb{R}$. In this case $\varphi_1 : \{-1, +1\} \to B_0$ and $\varphi_1(-1) = [-1, 1]$ and $\varphi_1(1) = [1, 1]$. This is the attaching map, and it attaches one interval to another by identifying endpoints, hence $\mathbb{R}P^1$ is just a circle.

The real projective plane $\mathbb{R}P^2$ can be written as a union of a 0-, 1- and 2-handle. From the construction we described we have the following balls

$$B_0 = \{[1, x, y] : -1 \leq x, y \leq 1\}$$
$$B_1 = \{[x, 1, y] : -1 \leq x, y \leq 1\}$$
$$B_2 = \{[x, y, 1] : -1 \leq x, y \leq 1\}$$

We attach a 1-handle to B_0 via the map $\varphi_1|_{\partial D \times D}$, which maps $(\pm 1, x)$ to $[\pm 1, 1, x] = [1, \pm 1, \pm x] \in B_1$. Explicitly, this means that $(1, x)$ is mapped to $[1, 1, x]$ and $(-1, x)$ maps to $[1, -1, -x]$, so
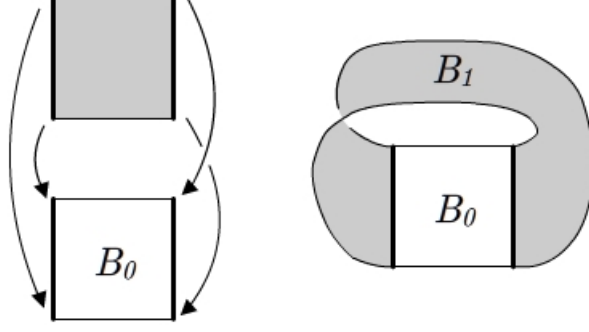
Figure 3.2: Attaching the 1-handle in $\mathbb{R}P^2$
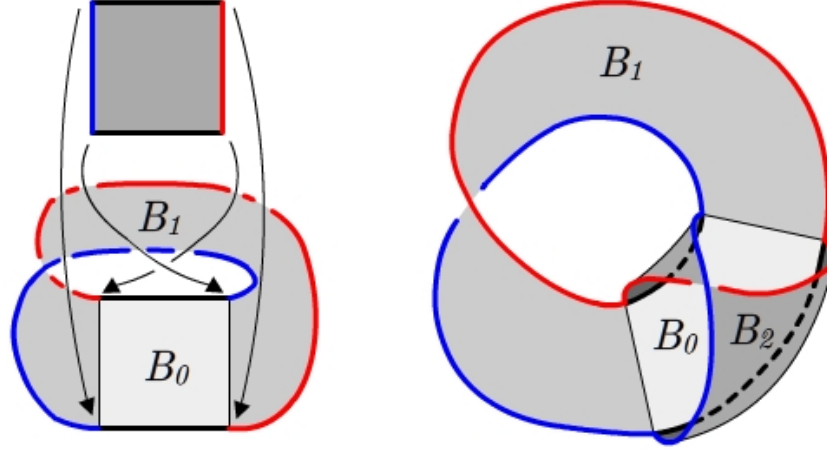
Figure 3.3: Attaching the 2-handle in $\mathbb{R}P^2$

this is identifying edges of two squares with a twist (see figure 3.2). We attach a 2-handle to $B_0 \cup B_1$ via $\varphi_2|_{\partial(D \times D)}$ which maps $(\pm 1, x)$ to $[\pm 1, x, 1] = [1, \pm x, \pm 1] \in B_0$ and maps $(x, \pm 1)$ to $[x, \pm 1, 1] = [\pm x, 1, \pm 1] \in B_1$. This glues a disc into $B_0 \cup B_1$ as in figure 3.3, where have not yet identified the final 2 sides of the square since that would make it difficult to draw $B_0 \cup B_1 \cup B_2$.

Now consider $\mathbb{C}P^1$, which is the union of a 0-handle and 2-handle. By definition we have

$$B_0 = \{[1, z] : |z| \leq 1\}$$
$$B_1 = \{[z, 1] : |z| \leq 1\}$$

We attach a 2-handle to B_0 via φ_1 which maps z to $[z, 1] = [1, 1/z] = [1, \bar{z}] \in B_0$. This means we attach two discs along the map $z \mapsto \bar{z}$, so $\mathbb{C}P^1$ is just the 2-sphere.

Finally, consider $\mathbb{C}P^2$, which is the union of a 0-handle, 1-handle and 2-handle. We have

$$B_0 = \{[1, z, w] : |z|, |w| \leq 1\}$$
$$B_1 = \{[z, 1, w] : |z|, |w| \leq 1\}$$
$$B_2 = \{[z, w, 1] : |z|, |w| \leq 1\}$$

We attach a 2-handle to B_0 via φ_1 restricted to $\partial D \times D$, which maps (z, w) to $[z, 1, w] = [1, 1/z, w/z] = [1, \bar{z}, \bar{z}w] \in B_0$. The 4-handle is attached to $B_0 \cup B_1$ via the restriction of φ_2 to $\partial(D \times D)$, which maps (z, w) to $[1, \bar{z}w, \bar{z}]$ on the $\partial D \times D$ part, and maps (z, w) to $[z\bar{w}, 1, \bar{w}]$ on the $D \times \partial D$ part.

There is a natural handlebody decomposition on the product of two handlebodies. Consider the simplest case

$$X = D^{n_1} \cup_{\varphi_1} h_1 \qquad h_1 = D^{k_1} \times D^{n_1-k_1} \qquad \varphi_1 : \partial D^{k_1} \times D^{n_1-k_1} \to \partial D^{n_1}$$

$$Y = D^{n_2} \cup_{\varphi_2} h_2 \qquad h_2 = D^{k_2} \times D^{n_2-k_2} \qquad \varphi_2 : \partial D^{k_2} \times D^{n_2-k_2} \to \partial D^{n_2}$$

Then $X \times Y$ has the following handlebody decomposition

$$X \times Y = D^{n_1} \times D^{n_2} \cup_{\varphi_1 \times \mathrm{id}} h_1 \times D^{n_2} \cup_{\mathrm{id} \times \varphi_2} D^{n_1} \times h_2 \cup h_1 \times h_2$$

Here we are thinking of $h_1 \times D^{n_2}$ as an $(n_1 + n_2)$-dimensional k_1-handle since it is just $D^{k_1} \times D^{n_1-k_1} \times D^{n_2}$ (the second two factors are like a $(n_1 + n_2)$-ball), and $D^{n_1} \times h_2$ is an $(n_1 + n_2)$-dimensional k_2-handle since it is just $D^{n_1} \times D^{k_2} \times D^{n_2-k_2}$ (here the factors are out of order, but that does not matter). We also think of $h_1 \times h_2$ as an $(n_1 + n_2)$-dimensional $(k_1 + k_2)$-handle since it is just $D^{k_1} \times D^{n_1-k_1} \times D^{k_2} \times D^{n_2-k_2}$, and the attaching region of this handle is

$$\partial D^{k_1} \times D^{n_1-k_1} \times D^{k_2} \times D^{n_2-k_2} \cup \partial D^{k_2} \times D^{n_2-k_2} \times D^{k_1} \times D^{n_1-k_1}$$

so the attaching map is $\varphi_1 \times \mathrm{id} \cup \varphi_2 \times \mathrm{id}$.

Note that if the handlebody decompositions of X and Y are induced by Morse functions f and g, respectively, then the product handlebody decomposition of $X \times Y$ is the same as the decomposition induced by the Morse function $h : X \times Y \to \mathbb{R}$ defined by $h(x,y) = f(x) + g(y)$. The critical points $(x,y) \in X \times Y$ of h are precisely where $dh_{(x,y)} = df_x + dg_y = 0$, hence $df_x = 0$ and $dg_y = 0$ and so x must be a critical point of f and y a critical point of g. If the critical point x is of index k and y of index l, then (x,y) is of index $k + l$, for the Hessian is of the form

$$H_h(x,y) = \begin{pmatrix} \left(\frac{\partial^2 f}{\partial x^i \partial x^j} \right) & 0 \\ 0 & \left(\frac{\partial^2 g}{\partial y^i \partial y^j} \right) \end{pmatrix} = \begin{pmatrix} H_f(x) & 0 \\ 0 & H_g(y) \end{pmatrix}$$

3.2   Handle Moves

Any manifold admits many "different" handlebody decompositions, so we want to find a set of moves that we could perform on handles in order to bring one handlebody presentation into another. First we show how to add a $(k-1)$-handle and k in a specific way such that the result manifold does not change; that is, the handles cancelled each other. Let X be an n-manifold with boundary. Write ∂D^k as a union of upper and lower hemispheres, $S^{k-1} = D_+^{k-1} \cup_{S^{k-2}} D_-^{k-1}$. Attach $D^k \times D^{n-k}$ to ∂X via an embedding of $D_-^{k-1} \times D^{n-k}$ in ∂X (this is *not* attaching a k-handle). This does not change the diffeomorphism type of X since it is just take the boundary sum of X with an n ball. We are going to show how to decompose this attached copy of $D^k \times D^{n-k}$ into a $(k-1)$-handle and k-handle. Let J be a small neighborhood of D_+^{k-1} in D^k, so that $J = D_+^{k-1} \times D^1$, and let D_0^k be the disc left from slicing J off of D^k. Then $J \times D^{n-k}$ is a $(k-1)$-handle attached to ∂X, and $D_0^k \times D^{n-k}$ is a k-handle attached to $\partial(X \cup J \times D^{n-k})$. This process of creating a pair of canceling handles is called **handle creation/cancellation**.

Sometimes it is possible to find a $(k-1)$-handle and k-handle that are situated in X so that they cancel like the above.

Proposition 3.2. *A $(k-1)$-handle h and a k-handle h' can be cancelled if the attaching sphere of h' intersects the belt sphere of h transversely in a single point.*

Proof. □

The process of finding two handles that cancel each other and canceling them from the handlebody is called **handle cancellation**. We can use handle cancellations to show that connected handlebodies can be made to have only one 0-handle and one n-handle.

Proposition 3.3. *If X is a compact and connected n-manifold, then X admits a handlebody decomposition with exactly one 0-handle and one n-handle.*

Proof. First we eliminate all 0-handles except for one. The 0-handles X_0 of X is just a disjoint union of n-balls, and since X is connected and since 1-handles are the only handles with disconnected attaching spheres, we have that each ball of X must be connected to some other ball via a 1-handle. The attaching sphere of a 1-handle is a pair of points, and the belt sphere of a 0-handle is S^{n-1}, so it is clear by proposition 3.2 that if a 1-handle connects two 0-handles then we can cancel one of the 0-handles and the 1-handle. Repeating this for all 1-handles connecting two 0-handles we can cancel all 0-handles except for one.

Now that we have eliminated all 0-handles except for one, take the dual handlebody decomposition. This is a handlebody decomposition with a unique n-handle, but possibly many 0-handles. Apply the procedure from before to eliminate all 0-handles except for one and we are left with a handlebody with a unique 0- and n-handle. □

Let (h_1, φ_1) and (h_2, φ_2) be two k-handles with attaching maps, and suppose they are attached to X simultaneously: $X \cup_{\varphi \cup \varphi_2} (h_1 \cup h_2)$. Let A be the image of the attaching sphere of h_1 in ∂X and B the image of the belt sphere in ∂X. We can isotope the attaching map of h_1, which can also be thought of as isotoping the attaching sphere and the framing of the attaching sphere, and the resulting manifold will not change. By isotoping the attaching sphere we are, intuitively, "sliding" the handle h_1 around $\partial(X \cup h_2)$. Now, slide the attaching sphere of h_1 over the handle h_2 until the attaching sphere of h_1 intersects the belt sphere of h_2. Since the attaching sphere of h_1 is of dimension $k-1$ and the belt sphere of h_2 is of dimension $n-k-1$, we have that they will intersect at a point p, and $T_p A \oplus T_p B$ is a codimension 1 subspace of $T_p \partial(X \cup h_2)$. This means there are two directions we can move A past B (corresponding to the two components of the complement of the hyperplane $T_p A \oplus T_p B$). One direction will slide h_1 back to where we started. If we push A the other way, and continue sliding h_1 until the attaching sphere lies entirely in ∂X, then we get a new handlebody which is the result of **sliding the handle h_1 over h_2**. Without loss of generality we can also assume that as we slide h_1 over h_2, the attaching sphere of h_1 stays inside some disc $D^k \times \{\text{pt}\}$ of h_2.

We have not change the manifold $X \cup h_1 \cup h_2$ by sliding the handle h_1 over h_2 (since we just isotoped the attaching sphere around), but the framing of h_1 might have changed in the process. For example, figure 3.4 shows that sliding a trivially attached 1-handle across a twisted 1-handle results in another twisting.

It turns out these two moves on handlebody decompositions can always determine when two handlebody decompositions give the same manifold.

Proposition 3.4. *For any two relative handlebody decompositions of a compact pair $(X, \partial_- X)$ (with handles attaching in increasing order), it is possible to bring one decomposition into the other via a finite sequence of creation/cancellation of handle pairs, handle slides, and isotopies of attaching maps.*
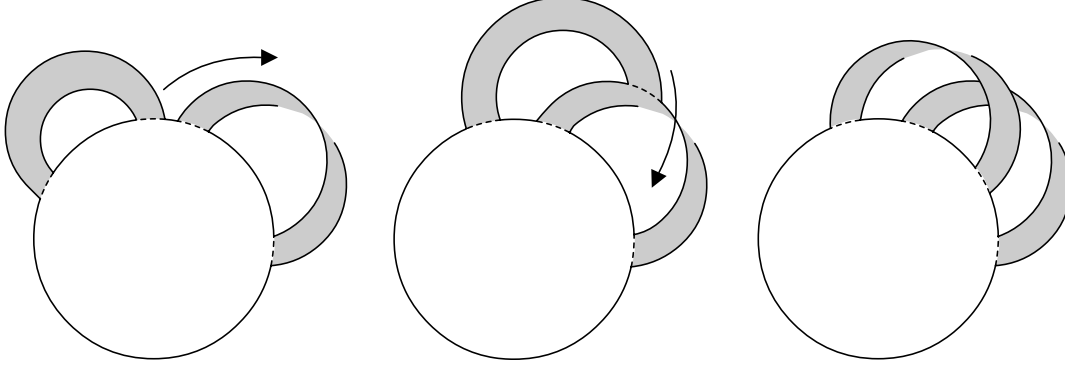
Figure 3.4: A handle slide that changes the framing of a handle

3.3   The Fundamental Group and Homology of a Handlebody

The combinatorial presentation of a manifold as a handlebody immediately gives a presentation of the fundamental group of the manifold. Suppose X is a handlebody with unique 0-handle, and give an orientation to the core of each 1-handle. We can make the oriented cores $D^1 \times 0$ into oriented loops in X_1 by connecting the end points with an oriented, straight line segment to the origin in $X_0 = D^n$. These loops are generators of the free group $\pi_1(X_1, x_0)$. As we attach 2-handles we will be adding relations to $\pi_1(X_1, x_0)$ to get $\pi_1(X_2, x_0)$. In particular, express the attaching circle $\partial D^2 \times 0$ of a handle in terms of the loops in X_1 to get the relation that attaching this 2-handle adds. Handles of higher index do not add any other relations since their attaching spheres are simply connected, so in particular $\pi_1(X_2, x_0) = \pi_1(X, x_0)$. Note that the above description of $\pi_1(X)$ is analogous to the presentation of $\pi_1(X)$ in terms of a CW-decomposition of a space.

Next we develop a homology theory in terms of handlebodies, which is similar to cellular homology derived from a CW-complex. For a relative handlebody $(X, \partial_- X)$, let the k-chains of our homology theory be $C_k(X, \partial_- X) = H_k(X_k, X_{k-1})$, where $H_k(-, -)$ is just regular singular homology, hence $C_k(X, \partial_- X)$ can be thought of as formal linear combinations of k-handles (once the cores of the handles are given a fixed orientation). The boundary operator $\partial_k : C_k(X, \partial_- X) \to C_{k-1}(X, \partial_- X)$ is defined by the long exact sequence of the triple (X_k, X_{k-1}, X_{k-2}). In particular, we have the following pieces of long exact sequences

$$\cdots \longrightarrow H_k(X_k, X_{k-1}) \xrightarrow{\Delta_k} H_{k-1}(X_{k-1}) \longrightarrow \cdots$$

$$\cdots \longrightarrow H_{k-1}(X_{k-1}) \xrightarrow{j_*} H_{k-1}(X_{k-1}, X_{k-2}) \longrightarrow \cdots$$

where Δ_k is the connecting morphism and j_* is induced by inclusion of the pairs $(X_{k-1}, \emptyset) \hookrightarrow (X_{k-1}, X_{k-2})$. Then we define $\partial_k = j_* \circ \Delta_k$, and clearly $\partial^2 = 0$ since $\Delta_k \circ j_* = 0$. If X is oriented, then there is a more geometric definition of ∂. Let h be a k-handle with attaching sphere A, and let B_i be the belt sphere of each $(k-1)$-handle h_i. Then

$$\partial_k h = \sum_i (A \cdot B_i) h_i \tag{3.1}$$

where $A \cdot B_i$ is the intersection number computed in $\partial_+ X_{k-1}$. Since $\ker \partial \subset \operatorname{im} \partial$ we define $H_k(X) = \ker \partial_k / \operatorname{im} \partial_{k+1}$, and call this the **$k$-th handlebody homology group** of X (this term is not standard). We can get a homology theory $H_k(X, \partial_- X; G)$ with coefficients in any abelian group G by defining $C_k(X, \partial_- X; G) = C_k(X, \partial_- X) \otimes_{\mathbb{Z}} G$ with boundary operator $\partial \otimes \operatorname{id}$. We can

also dualize this construction to get handlebody cohomology. In particular, set $C^k(X, \partial_- X; G) = \text{Hom}(C_k(X, \partial_- X), G)$ and let $\delta^k : C^k(X, \partial_- X; G) \to C^{k+1}(X, \partial_- X; G)$ be the induced differential, and define $H^k(X, \partial_- X; G) = \ker \delta_k / \operatorname{im} \delta_{k-1}$. This is called the k-**th handlebody cohomology group** of X. If there is risk of confusing the handlebody chain and homology groups with other homology theories (such as singular homology), we may sometimes write them as $C_k^{HB}(X)$ and $H_k^{HB}(X)$. Further, if $\partial_- X = \emptyset$, then we will write these groups as $C_k(X; G), H_k(X; G)$, and so on.

Since handle moves do not change the diffeomorphism type of X we have that the homology and cohomology groups will be the same. However, handle moves will add extra relations or change our canonical generators. Suppose a k-handle h and $(k-1)$-handle h_i form a canceling pair so that $A \cdot B_i = \pm 1$, where A is the attaching sphere of h and B_i is the belt sphere of h_i. Then $\partial h = \pm h_i + \sum_{j \neq i} a_j h_j$ for some integers a_j. When we cancel these handles we will add the relations $h = 0$ and $h_i = \mp \sum_{j \neq i} a_j h_j$ to the chain groups. On the other hand, sliding a handle h across a handle h' changes the canonical generators of $C_k(X)$ by replacing the basis element h with $h \pm h'$.

Example 3.2. Since S^n has a handlebody decomposition 0-handle $\cup$ 1-handle, we can easily see that its handlebody homology is $H_k(S^n) = \mathbb{Z}$ for $k = 0, n$ and 0 for $0 < k < n$. A more interesting example is writing S^3 as a union of a 0, 1, 2 and 3-handle, $S^3 = h_0 \cup h_1 \cup h_2 \cup h_3$. The standardly embedded solid torus $X_1 = S^1 \times D^2$ in S^3 is a 0-handle $\cup$ 1-handle. The attaching circle of the 2-handle we attach to X_1 is just the longitude $S^1 \times \{*\}$, where $*$ is any point on the boundary of D^2, and let X_2 be the result of this attachment. Then ∂X_2 is just S^2, and so we attach the 3-handle in the only way possible, and the result is $X_3 = S^3$.

We have $C_i(X) = \mathbb{Z} \langle h_i \rangle$. Since there are no (-1)-handles, we have $\partial_0 = 0$. The attaching sphere of h_1 intersects the belt sphere of h_0 at two points, but different signs, so $\partial_1(h_1) = 0$. The attaching sphere of h_2 intersects the belt sphere of h_1 at one point, which we can arrange to have intersection number $+1$, hence $\partial_2(h_2) = h_1$. Finally, the attaching sphere of h_3 intersects the belt sphere of h_2 at two points, but with different signs, so $\partial_3(h_3) = 0$. Taking homology we see again that $H_0(S^3) = H_3(S^3) = \mathbb{Z}$ and $H_1(S^3) = H_2(S^3) = 0$.

Example 3.3. We saw in example 3.1 that $\mathbb{C}P^n$ admits a handlebody decomposition with one k-handle for each even integer k between 0 and $2n$. So $C_k(\mathbb{C}P^n) = \mathbb{Z}$ for all even $0 \leq k \leq 2n$, and 0 for all other k. This implies that all the differentials are zero, hence $H_k(\mathbb{C}P^n) = C_k(\mathbb{C}P^n)$.

3.4   2-Dimensional Handlebodies

Two dimensional handlebodies have simple descriptions since there are so few ways to attach 2-dimensional k-handles ($k = 0, 1, 2$). The 0- and 2-handles are just 2-discs, and a 1-handle is a band $I \times I$. The attaching region of this band consists of two disjoint intervals $I \sqcup I$, and there are two ways to attach the band to a 0-handle: without a twist, or with a twist. Once the 1-handles are attached, there is a unique way to attach the 2-handles so that we get a closed surface (just fill the holes!), so we can draw diagrams of 2-dimensional handlebodies by just drawing the 0-handle $\cup$ 1-handles. Some examples of these decompositions are shown in figure 3.5, where the first image is $\mathbb{R}P^2$, the second is $T^2 \# \mathbb{R}P^2$, and the third is T^2.

3.5   Heegaard Diagrams

Heegaard splittings are an old method for constructing 3-manifolds, and we will see that it is equivalent to 3 dimensional handlebody. Although we get nice combinatorial pictures for 3 dimensional handlebodies, called Heegaard diagrams, they were still difficult to use. However, much later,
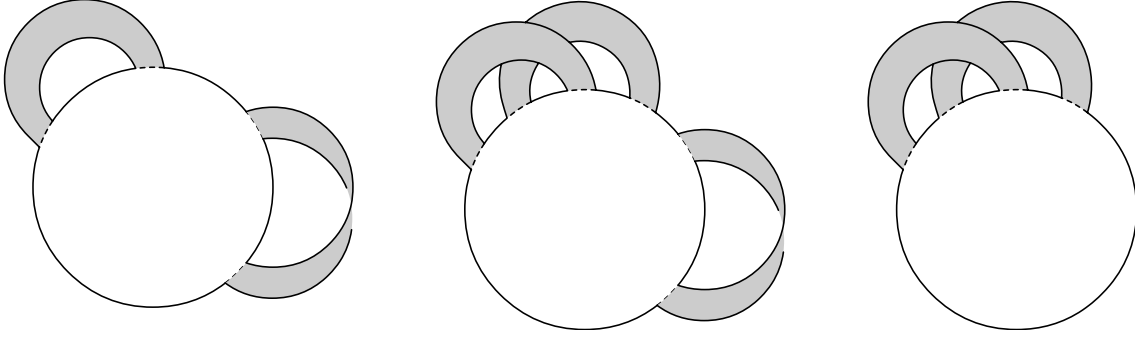
Figure 3.5: Examples of 2-dimensional handlebody diagrams

many invariants were built from Heegaard diagrams, including a recent variation of Lagrangian Floer homology, which provides powerful invariants of 3- and 4-manifolds.

We define a **handlebody of genus** g to be a 0-handle $\cup$ 1-handles such that the result is orientable, i.e. the interior (in the Jordan sense) of an orientable, genus g surface. If H_1 and H_2 are handlebodies of genus g, then for any orientation-reversing homeomorphism $\varphi : \partial H_1 \to \partial H_2$ of their boundaries, we can form the orientable, compact 3-manifold $H_1 \cup_\varphi H_2$. Conversely, given a 3-manifold M, it is always possible to find a pair of embedded submanifolds H_1, H_2, both homeomorphic to a handlebody of genus g, such that $\partial H_1 = \partial H_2$ (this is proved below). In either case we call this a **Heegaard splitting**, or **Heegaard decomposition**, of the 3-manifold.

Theorem 3.1. *Every compact, orientable 3-manifold admits a Heegaard splitting.*

Proof. We will give two different proofs.

- The first proof is more geometric, using the fact that every 3-manifold has a triangulation. Let K be a triangular of M, and let K' and K'' denote the first and second barycentric subdivision of K respectively. Let us recall how the barycentric subdivision is defined. Take a 2-face σ in the triangulation and connect, with a new edge, the barycenter of σ with each vertex of σ and the midpoint of each edge of σ. This divides σ into 6 new 2-faces. Then we can divide each 3-face ρ into 24 new 3-faces by coning the barycenter of ρ over each subdivided face of ρ. Performing this operation on all faces of K gives us a new triangulation of M, called the first barycentric subdivision K'. Performing this operation again gives us the second barycentric subdivision K''.

  Let H_1 be the union of the 3-faces in K'' that do not intersect the 1-faces of K, and let H_2 be the closure of the complement of H_1. Then H_1 and H_2 are handlebodies which intersect precisely at their boundaries. This gives a Heegaard splitting of M.

- The second proof is Morse theoretic, and is the beginning of more versatile methods in low dimensional topology via Morse theory. As we will see later (theorem 4.6), any smooth manifold admits a self-indexing Morse function $f : M \to [0, n]$ with unique index 0 and index n critical points. Let $f : M \to [0, 3]$ be such a Morse function for our 3-manifold, and let $H_1 = f^{-1}[0, 3/2]$ and $H_2 = f^{-1}[3/2, 3]$. If f has g index 1 critical points, then H_1 is a 0-handle with g 1-handles attached, hence a genus g handlebody. If we dualize the handlebody decomposition we see that H_2 is a 0-handle with h 1-handles attached, where h is the number of index 2 critical points. Since $\partial H_2 = \partial H_1 = $ a surface of genus g, we have that $g = h$, and
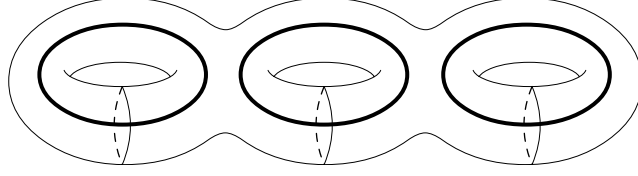
Figure 3.6: Standardly embedded handlebody

so H_2 is also a handlebody of genus g. Therefore M has been expressed as a union of two handlebodies along their boundary.

$\square$

The proof of theorem 3.1 is a little discouraging since any manifold admits many triangulations, and hence many splittings. But, there is a nice combinatorial picture to help us understand the structure of a Heegaard splitting. Let H_1 and H_2 be two copies of the standardly embedded handlebody (see figure 3.6), and let Σ be the standardly embedded surface of genus g. Take an orientation-preserving homeomorphism $\varphi_1 : \partial H_1 \to \Sigma$ and an orientation-reversing homeomorphism $\varphi_2 : \partial H_2 \to \Sigma$, and form the compact, orientable 3-manifold $M = H_1 \cup_{\varphi_2^{-1} \circ \varphi_1} H_2$. Let $\alpha_1, \ldots, \alpha_g$ be the image of the standard circles in Σ (the thin circles in figure 3.6) under φ_1, and $\beta_1, \ldots, \beta_g$ the image of the standard circles in Σ under φ_2. The curves $\alpha_1, \ldots, \alpha_g$ are collectively known as the α-curves, and are mutually disjoint, and the curves $\beta_1, \ldots, \beta_g$ are collectively known as the β-curves, and are also disjoint.

We now have a picture associated to any Heegaard decomposition: its a surface with $2g$ curves drawn on it. Many times we will take one of the collections of curves to be the standard curves, which would mean that φ_1 or φ_2 is just the identity, in which case we will not even draw them. The data of the surface Σ with these $2g$ embedded curves is called the **Heegaard diagram** of the splitting. We would like to know that these embedded curves somehow uniquely determine the 3-manifold obtained from gluing the handlebodies together, but first we state a small technical lemma that will be used repeatedly in these notes.

Lemma 3.1. *Let X and Y be two n-manifolds with homeomorphic boundaries, and let $Z = X \cup_f Y$ be the space formed by gluing X and Y together along a homeomorphism $f : \partial X \to \partial Y$. If $\tau : \partial X \to \partial X$ is any homeomorphism that extends to a homeomorphism $\widetilde{\tau} : X \to X$, then Z is homeomorphism to $Z' = X \cup_{f \circ \tau} Y$.*

Proof. There are embeddings of X and Y into Z by sending a point to its equivalence class in Z. If $x \in X, y \in Y$ are in the interior, then the corresponding points in Z are $[x] = \{x\}$ and $[y] = \{y\}$, whereas if $x \in X, y \in Y$ are boundary points, then $[x] = \{x, f(x)\}$ and $[y] = \{y, f^{-1}(y)\}$. Let $\widetilde{X}$ and $\widetilde{Y}$ denote the images of X and Y under these embeddings so that $\widetilde{X} \cap \widetilde{Y}$ is just the set of equivalence classes with representative in ∂X or ∂Y. We define a map $g : Z \to Z'$ on a point z in X or Y by the following

$$g(z) = \begin{cases} [\widetilde{\tau}^{-1}(x)] & z \in \widetilde{X} \\ [y] & z \in \widetilde{Y} \end{cases}$$

where $x \in X$ and $y \in Y$ are representatives of z. This map is clearly well-defined. In order for it to be continuous we need $[\widetilde{\tau}^{-1}(x)] = [y]$ in Z' for any $z = \{x, y\}$ in $\widetilde{X} \cap \widetilde{Y}$ (i.e. $y = f(x)$). This is easy to see for on one side we have

$$[\widetilde{\tau}^{-1}(x)] = [\tau^{-1}(x)] = \left\{ \tau^{-1}(x), f(\tau(\tau^{-1}(x))) \right\} = \left\{ \tau^{-1}(x), f(x) \right\}$$

36

and on the other side we have

$$[y] = [f(x)] = \left\{\tau^{-1}(x), f(x)\right\}$$

Therefore the map is continuous, and now it is clear that it is even a homeomorphism. $\qquad\square$

Theorem 3.2. *The homeomorphism type of the manifold resulting from gluing two handlebodies together depends only on the image of the standard meridians under the gluing map, hence equal Heegaard diagrams describe homeomorphic manifolds.*

Proof. Suppose we are gluing genus g handlebodies H_1 and H_2 together via a homeomorphism $f : \partial H_1 \to \partial H_2$. We can glue these handlebodies together in a two step process. First, let $\alpha_1, \ldots, \alpha_g$ denote a small tubular neighborhood of each standard meridian in H_1 (hence these are annuli), and let $D_1, \ldots, D_g$ denote the thickened discs these annuli bound in H_1. Then we can first attach the discs D_i to H_2 via f to obtain the intermedia space M'. Now we are left with attaching $H_1 \backslash (D_1 \cup \cdots \cup D_g)$ to M', but this is topologically a ball, and so any homeomorphism from the boundary of this ball to $\partial M'$ (which is just a sphere) can be extended to the interior of the ball, and hence any gluing map will give a homeomorphic space by the above gluing lemma. Therefore the homeomorphism type of $H_1 \cup_f H_2$ depends only on the images of the α_i curves. $\qquad\square$

We can also define Heegaard diagrams axiomatically. Let us temporarily say that an *H-diagram* is a surface Σ of genus g and embedded curves $\alpha_1, \ldots, \alpha_g, \beta_1, \ldots, \beta_g$ such that the α-curves are disjoint and the complement of their union is connected, and similarly for the β-curves.

Theorem 3.3. *An H-diagram $(\Sigma, \alpha_1, \ldots, \alpha_g, \beta_1, \ldots, \beta_g)$ corresponds to the Heegaard diagram of some splitting.*

Proof. Let α_i, β_i be curves with the above properties. First we claim that cutting Σ along the α-curves yields a 2-sphere with $2g$ holes. Suppose that Σ with the α-curves removed is a surface of genus h with $2g$ holes. Then its Euler characteristic would be $2 - 2h - 2g$. On the other hand, removing circles from surfaces that do not separate the surface does not change the Euler characteristic, so we should have $2 - 2g = 2 - 2h - 2g$, hence $h = 0$, and so $\Sigma \backslash \alpha_1, \ldots, \alpha_g$ is a sphere with $2g$ holes.

Since Σ with the removed α-curves is just a sphere with $2g$ holes, we can homeomorphically deform it so that it is in the standard position of a handlebody of genus g with its standard circles cut out. Now it is easy to find a handlebody of genus g with homeomorphism of its boundary onto Σ taking its standard circles to the α-circles. This can be done for the β curves in the same way. $\qquad\square$

Therefore H-diagrams and Heegaard diagrams are equivalent notions, so we will refer to both as Heegaard diagrams from now on. The condition that the complement of the union of the α-curves in Σ is connected is equivalent to saying that the homology classes of $\alpha_1, \ldots, \alpha_g$ span a rank g submodule of $H_1(\Sigma) \cong \mathbb{Z}^{2g}$.

There is also a Morse theoretic way of thinking of Heegaard diagrams. Let $f : M \to [0, 3]$ be a self-indexing Morse function with unique maximum and minimum, and let $H_1 = f^{-1}[0, 3/2]$, $H_2 = f^{-1}[3/2, 3]$ and $\Sigma = f^{-1}(3/2)$. By theorem 3.1 we know that H_1 and H_2 are handlebodies of the same genus, say g. Let $P_1, \ldots, P_g$ be the index 1 critical points of f, and $Q_1, \ldots, Q_g$ be the index 2 critical points. Let α_i consist of the points that flow to P_i under $-\nabla_g f$, and let β_i consist of the points that flow to Q_i under $+\nabla_g f$. Then $\Sigma(\alpha_1, \ldots, \alpha_g, \beta_1, \ldots, \beta_g)$ is a Heegaard diagram for M.
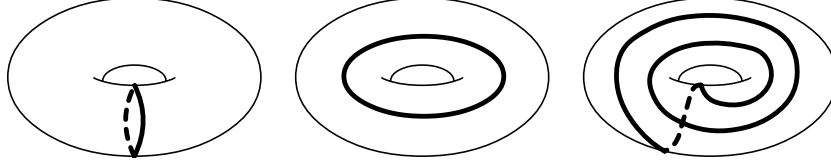
Figure 3.7: Heegaard diagrams of $S^1 \times S^2, S^3, \mathbb{R}P^3$ respectively

Example 3.4. There is only one Heegaard splitting of genus 0. In this case we have a manifold M expressed as $M = D_1 \cup D_2$ where D_1, D_2 are 3-balls. Let $\varphi : D_1 \to S^3$ be an embedding of the first 3-ball onto the upper-hemisphere of S^3. The restriction of φ to $\partial D_1 = \partial D_2$ can be extended to a homeomorphism from D_2 to the lower-hemisphere of S^3 (by Alexander's lemma). Pasting these homeomorphisms together gives a homeomorphism of M with S^3, hence all genus 0 Heegaard splittings are S^3.

Example 3.5. The standard Heegaard splitting of S^3 is to take the standardly embedded solid torus H_1 in S^3, and let H_2 be the complement of H_1 in S^3. The gluing map is just the identity. The standard orientation on S^3 induces orientations on H_1 and H_2. If we let n denote the outward pointing normal vector field on ∂H_1, then the orientation of H_1 induces an orientation on ∂H_1 by choosing the equivalence class of bases at a tangent space such that adding n gives the equivalence class of bases from the induced orientation on H_1. Do something similar to get an orientation on ∂H_2 induced by H_2. Clearly the two orientations induced on $\partial H_1 = \partial H_2$ are different, therefore the gluing map is orientation reversing. This is why we require the gluing map to be orientation reversing in our definition of Heegaard splitting.

Example 3.6. We can also classify all genus 1 splittings. All such splittings are given by an orientation reversing homeomorphism $f : T \to T$, and the isotopy classes of these maps are in bijective correspondence with integer matrices

$$A = \begin{pmatrix} -q & r \\ p & s \end{pmatrix} \qquad \det A = pr + qs = 1$$

so p and q are relatively prime or one is 0 and the other is ± 1. We use $-q$ here because we required our gluing homeomorphism to be orientation reversing, so the determinant of this matrix should be -1, which can be written simply as $pr + qs = 1$ with our sign choice. This homeomorphism maps the meridian μ to the curve $-q \cdot \mu + p \cdot \lambda$, i.e. the meridian wraps $-q$ times around the meridian and p times around the longitude. Let M be the manifold resulting from gluing to copies of the solid torus via f. It is easy to see that M depends only on the image of μ under f by using an argument similar to theorem 3.2. Thus, M is completely determined by the relatively prime pair of integers (p, q), and we call this space the (p, q) **Lens space** and denote it by $L(p, q)$.

The Heegaard diagram of $L(p, q)$ is just a curve on the torus that wraps $-q$ times around the meridian and p times around the longitude. One can prove that $L(p, q) = L(-p, q)$ and $L(p, q) = L(p, q + np)$ for any integer n, so we can assume $p \geq 0$ and $0 \leq q \leq p - 1$. But, if $p = 0$ then we are forced to have $q = \pm 1$, and if $q = 0$ we are also forced to have $p = \pm 1$. In these cases we have $L(\pm 1, n) = S^3$ for all integers n and $L(0, \pm 1) = S^1 \times S^2$, so disregarding these examples we can assume $p \geq 2$ and $1 \leq q \leq p - 1$. We also have $L(2, 1) = \mathbb{R}P^3$.

Example 3.7. Heegaard diagrams can be quite complicated, for example see figure 3.8. We will show later that this 3-manifold is simply connected.
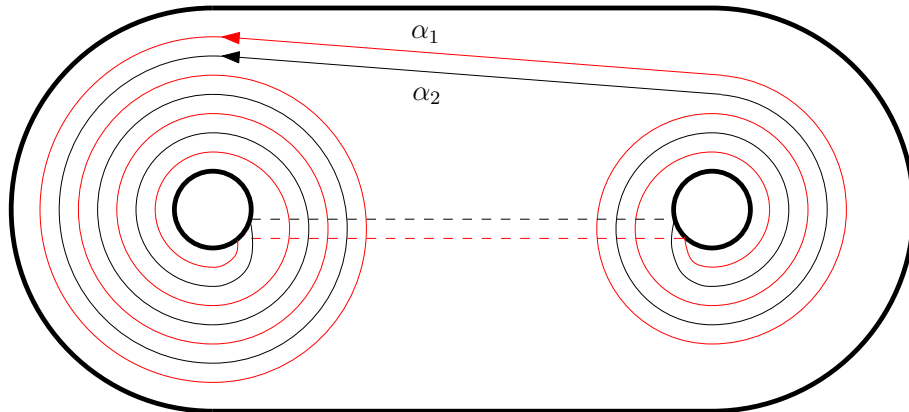
Figure 3.8: Heegaard diagram of a simply connected 3-manifold

Let us reconcile the above theory of Heegaard splittings of 3-manifolds with our general theory of handlebodies. A handlebody of genus g is a 0-handle $\cup$ 1-handles such that the result is orientable. This is just the first step in building a 3-dimensional handlebody, so let us denote it by X_1. Next we could attach some 2-handles (we certainly cannot attach a 3-handle yet, unless $g = 0$). Attaching a 2-handle is completely determined by the image of the attaching sphere $\partial D^2 \times 0$ (which we will call the attaching circle in this case) and the framing of the attaching circle, which are parameterized by $\pi_1(O(1)) = 0$. Therefore, attaching the 2-handles is equivalent to specifying a collection of embedded circles in X_1, which is just a Heegaard diagram, and let X_2 be the result. If ∂X_2 is a collection of S^3's, then there is a unique way to cap these off with 3-handles so that the result is a closed 3-manifold. This is precisely the case we draw g embedded curves on X_1 such that the complement of their union is connected, so such a picture uniquely determines a 3-manifold.

Now we see that a Heegaard splitting is simply a 3-dimensional handlebody, and a Heegaard diagram is just a collection of curves drawn on X_1, we would like to translate our machinery of handle moves, handlebody homology and computation of fundamental groups into the Heegaard setting. We start with handle moves. We fix X_1 to be the standardly embedded handlebody of genus g since all the information in a 3-dimensional handlebody is in the attaching circles of the 2-handles. In the handlebody theory we could isotopy the attaching maps, which just means that we can isotopy the α and β curves on the surface of genus g. To create a canceling pair of 1- and 2-handles simply increase the genus of X_1 by one (i.e. add a 1-handle) and add the attaching circle that goes once around the longitude of the new 1-handle but does not venture into the old part of X_1. Attaching the 2- and 3-handles to this new Heegaard diagram will give the original manifold connect summed with S^3 (recall the Heegaard diagram of S^3, figure 3.7), hence it will be diffeomorphic to the original manifold.

Handle slides are a little more difficult to understand. Sliding the 1-handles is very simple, and does not produce anything interesting since we required X_1 to be orientable, hence allowed only one type of framing on the 1-handles. To get a feeling for sliding 2-handles, consider sliding two handles that are attaching to D^3. Figure 3.9 shows how to do this in the obvious way. Here we sliding the attaching circle of one handle through the disc $D^2 \times \{1\}$ of the other handle. A better way to look at this is just to look at the attaching circles. Then we see that sliding 2-handles is equivalent to taking a parallel of one circle, and forming the banded sum of this circle with the other attaching circle (see figure 3.10).

Now to slide a 2-handle h_1 across another 2-handle h_2 in a Heegaard diagram make a parallel
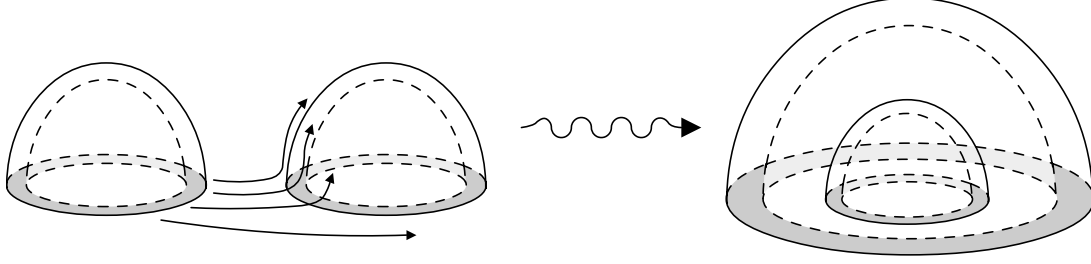
Figure 3.9: Sliding 3-dimensional 2-handles
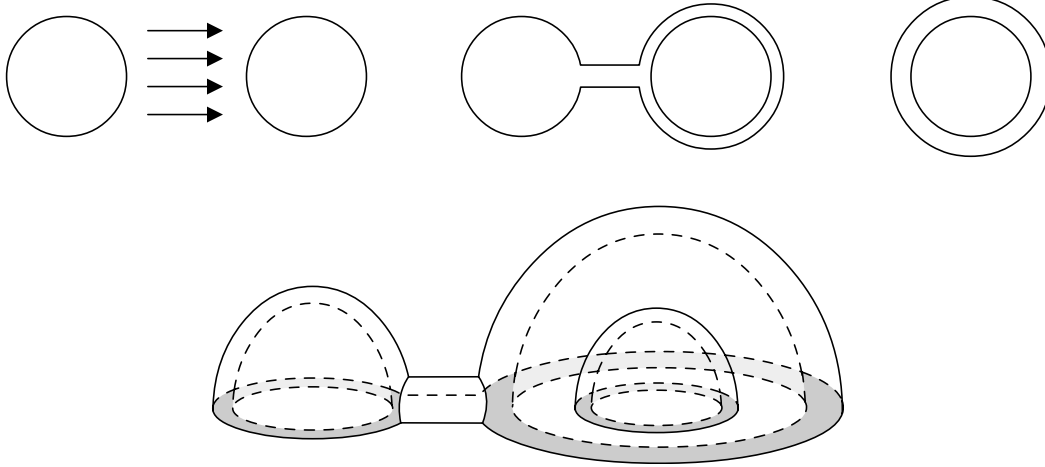
Figure 3.10: Sliding 3-dimensional 2-handles

copy of the attaching circle α_2 of h_2 and form the *band sum* of the attaching circle α_1 of h_1 with the parallel of h_2. If we let γ denote the band sum of α_1 and the parallel of α_2, then note that α_1, α_2, and γ bound a pair of pants in X_1. Conversely, if there are two attaching circles α_1, α_2 on ∂X_1 and another curve γ on ∂X_1 such that γ does not intersect any other attaching curves and α_1, α_2, and γ bound a pair of pants, then we can discard any one of these three curves without changing the resulting 3-manifold. See figure 3.11 for a non-trivial example of this.

Next we see how to compute the fundamental group of a 3-manifold X from one of its Heegaard diagrams. Choose a basis $x_1, \ldots, x_g$ for the free group $\pi_1(X_1)$, and let r_i be the word in $x_1, \ldots, x_g$ representing the i-th attaching circle. Then a presentation of $\pi_1(X)$ is simply $\langle x_1, \ldots, x_g : r_1, \ldots, r_g \rangle$.

Example 3.8. Consider the Heegaard diagram in figure 3.8. Let x_1 denote the standard generator around the first hole (counter-clockwise orientation) and x_2 the generator around the second hole (counter-clockwise orientation). One can easily check that the curve α_1 is represented by the curve $x_1^3 x_2^2$ and α_2 is represented by the curve $x_1^4 x_2^3$. Therefore a presentation for the manifold M associated to this Heegaard splitting is $\pi_1(M) = \langle x_1, x_2 : x_1^3 x_2^2 = x_1^4 x_2^3 = 1 \rangle$. The first relation implies that $x_1^3 x_2^3 = x_2$, and plugging this into the second relation gives $x_1 x_2 = 1$. Therefore x_1 and x_2 are inverses in $\pi_1(M)$, hence $1 = x_1^3 x_2^2 = x_1$ and $x_2 = 1$, and so M is simply connected.

Now we compute the homology of a 3-manifold X from one of its Heegaard diagrams. We clearly have $H_0(X) = H_3(X) = \mathbb{Z}$, $C_1(X) = C_2(X) = \mathbb{Z}^g$ (if X has g 1- and 2-handles). Further, $\partial_1 = 0$ since the attaching spheres of the 1-handles intersect ∂D^3 at two points of opposite sign, and $\partial_3 = 0$ for similar reasons. Hence all the homology information of X is contained in ∂_2. Given
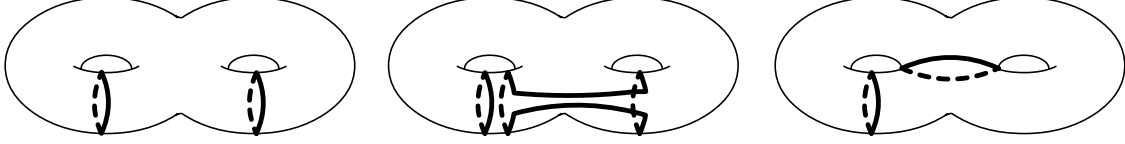
Figure 3.11: Sliding 3-dimensional 2-handles

a 2-handle h, $\partial_2 h$ is a linear combination of the one handles $\{h_i\}_{i=1}^{g}$, where the coefficient of h_i is the intersection number of h's attaching circle with h_i's belt sphere. The belt sphere of h_i is just the meridian that bounds a disc (the non-bolded curves in figure 3.6).

Example 3.9. If we label the left and right 1-handles of figure 3.8 by h_1 and h_2 respectively, and label the 2-handle corresponding to the black attaching circle by k_1 and the other 2-handle by k_2, then a close examination of this diagram shows that

$$\partial_2 h_1 = 3k_1 + 2k_2$$

$$\partial_2 h_2 = 4k_1 + 3k_2$$

We could also see this by just abelianizing the relations of π_1. As a linear transformation $H_2(X) = \mathbb{Z}^2 \to \mathbb{Z}^2 = H_1(X)$ it is invertible since its determinant is $+1$, therefore ∂_2 is an isomorphism. This implies that $H_1(X) = 0 = H_2(X)$, and so X is a homology sphere. In fact, X is homeomorphic to S^3, as can be checked by sliding handles.

3.6 Kirby Diagrams

A **Kirby diagram** is a diagram we can associate to a 4-dimensional handlebody X. We start with a 4-ball D^4, and think of its boundary $\partial D^4 = S^3$ as $\mathbb{R}^3 \cup \{\infty\}$. We attach 1-handles to this via an embedding of $D^3 \sqcup D^3$. There is the issue of framing (there are two framings), but for now let us choose the framing such that $X_1 = \natural m S^1 \times D^3$ (here we are attaching m 1-handles). We can represent this in $\mathbb{R}^3 \cup \{\infty\}$ by drawing $2m$ disjoint embedded 3-balls in $\mathbb{R}^3 \cup \{\infty\}$ and pairing them so that when we identify pairs in an orientation reversing manner we get $X_1 = \natural m S^1 \times D^3$.

To attaching a 2-handle to X_1 we specify a knot in X_1 (the attaching circle) and label it with a framing. So, a diagram of X_2 consists of m pairs of 3-balls and a collection of arcs that either start and end on the boundary of the 3-balls, or form a closed loop. We can specify a framing on a knot component by a nowhere zero normal vector field, or equivalently, a parallel knot. Once we specify which parallel corresponds to $0 \in \pi_1(O(2)) = \mathbb{Z}$ we can get any other framing by adding full left or right twists. We will choose the parallel which has linking number 0 with the knot to be the framing corresponding to 0.

It turns out that there is essentially a unique way to attach 3-handles and 4-handles so that we obtain a closed manifold, so our diagram of X can stop here. Things are more complicated when $\partial X \neq \emptyset$, but we will not deal with that now.

(**todo:** need more on this...)

Example 3.10. Let us find a Kirby diagram of $S^2 \times S^2$. The 2-sphere has the obvious decomposition $S^2 = D_- \cup D_+$, where D_- is a 0-handle, D_+ a 2-handle, which we think of as the southern and northern hemispheres. The attaching map for the 2-handle is just the identity. The handlebody decomposition of $S^2 \times S^2$ is just

$$S^2 \times S^2 = \overbrace{D_- \times D_-}^{0-\text{handle}} \cup \overbrace{D_+ \times D_-}^{2-\text{handle}} \cup \overbrace{D_- \times D_+}^{2-\text{handle}} \cup \overbrace{D_+ \times D_+}^{4-\text{handle}}$$

First we determine the attaching maps for the 2-handles. One is the map $\phi_1 : \partial D_+ \times D_- \to \partial(D_- \times D_-)$ defined by $\phi_1(x, y) = (x, y)$, and the other is the map ϕ_2 :

$$D^2 \times D^0 \times D^0 \times D^2$$

UNFINISHED

Example 3.11. Let us find a handlebody decomposition of $S^2 \tilde{\times} S^2$. Recall that this is the boundary of the non-trivial D^3 bundle over S^2.

4 Morse Theory

4.1 Morse Functions

We start with a smooth n-manifold M and a smooth function $f : M \to \mathbb{R}$. Take a chart (x, U), that is an open set $U \subset M$ and a homeomorphism $x : U \to \mathbb{R}^n$. If we write x in components $x = (x^1, \ldots, x^n)$, then we define the derivative of f at some point $p \in U$ with respect to this chart by

$$\frac{\partial f}{\partial x^i}(p) = D_i(f \circ x^{-1})(x(p))$$

where D_i is the regular derivative of $f \circ x^{-1} : \mathbb{R}^n \to \mathbb{R}$ in the i-th component. Recall that if (y, V) is another chart, then the derivative of f with respect to these two charts on the overlap are related by

$$\frac{\partial f}{\partial y^i} = \sum_{j=1}^{n} \frac{\partial f}{\partial x^j} \frac{\partial x^j}{\partial y^i} \tag{4.1}$$

Another way of saying this is to first define the **Jacobian** matrix of two charts (x, U) and (y, V). This is the matrix whose (i, j)-entry is equal to $\frac{\partial x^i}{\partial y^j}$, and it is denoted by J_y^x. If we let $\frac{\partial f}{\partial x}$ denote the column vector of derivatives of f, then (4.1) simply says

$$\frac{\partial f}{\partial y} = J_y^x \cdot \frac{\partial f}{\partial x}$$

Since $\frac{\partial f}{\partial x^i}$ is just another smooth map from U to $\mathbb{R}$ we can take a second derivative with respect to a chart. The second derivatives calculated with respect to different charts are related by

$$\frac{\partial^2 f}{\partial y^i \partial y^j} = \frac{\partial^2 f}{\partial x^k \partial x^\ell} \frac{\partial x^k}{\partial y^i} \frac{\partial x^\ell}{\partial y^j} \tag{4.2}$$

As with first derivatives there is another way of looking at this. Let us define the **Hessian** of f with respect to the chart x be the matrix whose (i, j)-entry is $\frac{\partial^2 f}{\partial x^i \partial x^j}$, and it is denoted by H_f^x. Since partial derivatives commute we have that H_f^x is symmetric. We now see that (4.2) can be written as

$$H_f^y = (J_y^x)^t \cdot H_f^x \cdot J_y^x \tag{4.3}$$

This tells us that even though the Hessian as a matrix is not well-defined (i.e. it depends on the chart), it is well-defined as a bilinear form. There is also a nice coordinate free way of describing the Hessian. Let $\nabla : \Gamma(TM) \times \Gamma(T^*M) \to \Gamma(T^*M)$ be any connection on T^*M, and for any $p \in M$ define the map $H_f(p) : T_pM \to T_p^*M$ by

$$H_f(p)(X) = \nabla_X(df) \tag{4.4}$$

This does not depend on the choice of connection, for if $\widetilde{\nabla}$ is another connection, then the difference $\nabla - \widetilde{\nabla}$ is tensorial in both arguments and $df_p = 0$, hence $\nabla_X(df) - \widetilde{\nabla}_X(df) = 0$ at p.

Suppose that a point $p \in M$ is a critical point of f, i.e. $df_p = 0$. We say that p is a **non-degenerate** critical point if $\det H_f^x(p)$ is nonzero for some chart x, hence every chart by (4.3). We define the **index** of a critical point p to be the number of negative eigenvalues of the Hessian at p, which is well-defined by (4.3), and is denoted by $\mu(p)$. Equivalently, we call the critical point p non-degenerate if $H_f(p)$ does not have 0 for an eigenvalue, and $\mu(p)$ is the number of

negative eigenvalues of $H_f(p)$. Finally, we define a **Morse function** is a smooth function with all non-degenerate critical points.

It turns out that Morse functions look very nice around critical points. The Morse lemma gives a precise statement of this, and will be very important in our discussion of cell decompositions of spaces via Morse functions. First we recall a basic fact from calculus using the fundamental theorem.

Lemma 4.1. *Let U be a convex subset of $\mathbb{R}^n$ containing 0 and $f : U \to \mathbb{R}$ a smooth function with $f(0) = 0$. Then we can write*

$$f(x_1, \ldots, x_n) = \sum_{i=1}^n x_i g_i(x_1, \ldots, x_n)$$

for some $g_i : U \to \mathbb{R}$ with $g_i(0) = \frac{\partial f}{\partial x_i}(0)$.

Proof. For fixed $x_1, \ldots, x_n$ we apply the fundamental theorem of calculus and chain rule to get

$$f(x_1, \ldots, x_n) = \int_0^1 \frac{df(tx_1, \ldots, tx_n)}{dt} \, dt = \int_0^1 \sum_{i=1}^n x_i \cdot \frac{\partial f}{\partial x_i}(tx_1, \ldots, tx_n) \, dt$$

So let $g_i(x_1, \ldots, x_n) = \int_0^1 \frac{\partial f}{\partial x_i}(tx_1, \ldots, tx_n) \, dt$, then clearly $g_i(0) = \frac{\partial f}{\partial x_i}(0)$ and we are done. $\qquad \square$

Lemma 4.2 (Morse Lemma)**.** *Let $f : M \to \mathbb{R}$ be a smooth function and $p \in M$ a non-degenerate critical point. Then there is a chart $y = (y^1, \ldots, y^n)$ on a neighborhood U of p such that $y^i(p) = 0$ and*

$$f = f(p) - (y^1)^2 - \cdots - (y^\mu)^2 + (y^{\mu+1})^2 + \cdots + (y^n)^2$$

on U, where μ is the index of p.

Proof. First note that finding a chart so that f can be expressed in the above form is equivalent to finding a chart (x, U) such that on U we have

$$(f \circ x^{-1})(t_1, \ldots, t_n) = -t_1^2 - \cdots - t_\mu^2 + t_{\mu+1}^2 + \cdots t_n^2$$

where t_i are real numbers. Therefore we can, without loss of generality, assume that f is defined on some open set V of $\mathbb{R}^n$ (say containing zero), and by performing some translations we can assume $p = 0$ and $f(0) = 0$.

Next it is clear that if f can be represented in the above form for some chart (y, U), then we have

$$\frac{\partial^2 f}{\partial y^i \partial y^j}(p) = \begin{cases} -2 & i = j, j \le \mu \\ 2 & i = j, j > \mu \end{cases}$$

Therefore the Hessian H_f^x is already diagonalized in this coordinate system, and clearly has μ negative entries along the diagonal. Therefore μ really will be the index of p if we can get f into the above form.

Let $x = (x^1, \ldots, x^n)$ be the standard coordinates on $\mathbb{R}^n$. By lemma 4.1 we can find functions $g_i : V \to \mathbb{R}$ such that

$$f(x^1, \ldots, x^n) = \sum_{i=1}^n x^i \cdot g_i(x^1, \ldots, x^n)$$

44

and $g_i(0) = \frac{\partial f}{\partial x^i}(0)$. However, since $p = 0$ is assumed to be a critical point we have $\frac{\partial f}{\partial x^i}(0) = 0$, so we can apply lemma 4.1 to the functions g_i to obtain functions $h_{ij} : V \to \mathbb{R}$ such that

$$g_i(x^1, \ldots, x^n) = \sum_{j=1}^{n} x^j h_{ij}(x^1, \ldots, x^n)$$

and $h_{ij}(0) = \frac{\partial g_i}{\partial x^j}(0) = \frac{\partial^2 f}{\partial x^i \partial x^j}(0)$. Combining the above two expressions yields

$$f(x^1, \ldots, x^n) = \sum_{i,j=1}^{n} x^i x^j h_{ij}(x^1, \ldots, x^n) \tag{4.5}$$

If we define $\widetilde{h}_{ij} = \frac{1}{2}(h_{ij} + h_{ji})$, then we see that $\widetilde{h}_{ij} = \widetilde{h}_{ji}$ and yet still $f = \sum x^i x^j \widetilde{h}_{ij}$. Therefore we can assume that $h_{ij} = h_{ji}$. It is also clear that the matrix $(h_{ij}(0))_{ij}$ is equal to $\frac{1}{2} H_f^x$.

We now simply mimic the proof of the diagonalization of quadratic forms, applied to (h_{ij}), to show that we can find coordinates to bring f into the desired form. By assumption the Hessian is non-singular, so $(h_{ij}(0))$ is also non-singular. Reordering the coordinate system can assume that $h_{11}(0) \neq 0$, and so by continuity we have that $h_{11}(0) \neq 0$ on some small neighborhood V' of 0. Let us also assume that $h_{11} > 0$ on V' (the other case is treated similarly). We introduce a new coordinate $y^1 : V' \to \mathbb{R}$ defined by

$$y^1 = \sqrt{h_{11}} \left(x^1 + \sum_{i=2}^{n} \frac{x^i h_{1i}}{h_{11}} \right) \tag{4.6}$$

The coordinates $(y^1, x^2, \ldots, x^n)$ form a chart (in a possibly smaller neighborhood V'' of 0). Squaring this new coordinate gives

$$
\begin{aligned}
(y^1)^2 \;\; &= h_{11} \left(x^1 + \sum_{i=2}^{n} \frac{x^i h_{1i}}{h_{11}} \right)^2 \\
&= h_{11} \left((x^1)^2 + 2 \sum_{i=2}^{n} \frac{x^1 x^i h_{1i}}{h_{11}} + \left(\sum_{i=2}^{n} \frac{x^i h_{1i}}{h_{11}} \right)^2 \right) \\
&= h_{11}(x^1)^2 + 2 \sum_{i=2}^{n} x^1 x^i h_{1i} + \frac{1}{h_{11}} \left(\sum_{i=2}^{n} x^i h_{1i} \right)^2
\end{aligned}
\tag{4.7}
$$

This expression allows us to rewrite $(x^1)^2$ as

$$h_{11}(x^1)^2 = (y^1)^2 - 2 \sum_{i=2}^{n} x^1 x^i h_{1i} - \frac{1}{h_{11}} \left(\sum_{i=2}^{n} x^i h_{1i} \right)^2 \tag{4.8}$$

Now, using (4.5) and (4.8) we can rewrite f in coordinates $(y^1, x^2, \ldots, x^n)$ as

$$f = h_{11}(x^1)^2 + 2\sum_{i=2}^{n} x^1 x^i h_{1i} + \sum_{i,j=2}^{n} x^i x^j h_{ij}$$

$$= (y^1)^2 - 2\sum_{i=2}^{n} x^1 x^i h_{1i} - \frac{1}{h_{11}}\left(\sum_{i=2}^{n} x^i h_{1i}\right)^2 + 2\sum_{i=2}^{n} x^1 x^i h_{1i} + \sum_{i,j=2}^{n} x^i x^j h_{ij}$$

$$= (y^1)^2 + \sum_{i,j=2}^{n} x^i x^j h_{ij} - \frac{1}{h_{11}}\left(\sum_{i=2}^{n} x^i h_{1i}\right)^2$$

Note that had we assumed $h_{11} < 0$ on V', then we would have $-(y^1)^2$ in the above. The terms in this last expression only involve x^i's for $i \geq 2$, so we can repeat the above arguments to find a new coordinate y^2 such that $(y^1, y^2, x^3, \ldots, x^n)$ is a coordinate system and

$$f = \pm(y^1)^2 \pm (y^2)^2 \pm \text{ terms with } x^3, x^4, \ldots, x^n$$

Continuing this we get a coordinate system $y = (y^1, \ldots, y^n)$ such that

$$f = \pm(y^1)^2 \pm \cdots + (y^n)^2$$

Since the number of minus signs is equal to μ, which does not depend on coordinate charts, we are done. $\qquad\square$

A simple consequence of the Morse lemma is the following.

Corollary 4.1. *The critical points of a Morse function are isolated.*

4.2 Handlebody Decompositions Induced by Morse Functions

For a manifold M with Morse function $f : M \to \mathbb{R}$, let us define the sets $M^a = f^{-1}(-\infty, a]$. If a is not a critical value, then M^a is a submanifold with boundary. We will now analyze what happens to M^a as a crosses a critical value. Our first theorem towards this says that the diffeomorphism type of M^a does not change between critical values. First let us develop the notion of gradient vector fields and flows.

Let (M, g) be a Riemannian manifold with Morse function $f : M \to \mathbb{R}$. Recall that the metric gives us a canonical isomorphism of tangent spaces with cotangent spaces by mapping a vector $X \in T_p M$ to the covector $g(X, -) \in T_p^* M$. The gradient vector field of f is the unique vector field $\nabla_g f$ corresponding to the 1-form df under this isomorphism, hence

$$g(X, \nabla_g f) = df(X)$$

for any vector field X. In coordinates $(x^1, \ldots, x^n)$ we can write

$$\nabla_g f = \sum_{i,j=2}^{n} g^{ij} \frac{\partial f}{\partial x^j} \frac{\partial}{\partial x^i}$$

where g^{ij} are the components of the inverse metric. It is clear that $\nabla_g f$ vanishes at precisely the critical points of f.

Theorem 4.1. *Let $[a, b]$ be an interval that does not contain any critical values and such that $f^{-1}[a, b]$ is compact. Then M^a is diffeomorphic to M^b. In fact, M^a is a deformation retract of M^b.*

Proof. First note that the compactness hypothesis would be immediate if we assumed that M is compact. Let X be a vector field given by

$$X = \frac{\nabla_g f}{||\nabla_g f||^2}$$

on $f^{-1}[a, b]$, and vanish identically outside some compact set containing $f^{-1}[a, b]$. Since this vector field has compact support it generates a unique 1-parameter group $\phi_t : M \to M$ of diffeomorphisms. For fixed $p \in M$ such that $\phi_t(p)$ lies in $f^{-1}[a, b]$, the function $f \circ \phi(p)$ is a map $\mathbb{R} \to \mathbb{R}$, and we can compute (using the chain rule and definition of $\nabla_g f$)

$$\frac{df(\phi_t(p))}{dt} = df(d\phi_t(d/dt)) = g(d\phi_t(d/dt), \nabla_g f) = \frac{g(\nabla_g f, \nabla_g f)}{g(\nabla_g f, \nabla_g f)} = 1$$

So, the function $f \circ \phi(p) : \mathbb{R} \to R$ is linear; in fact, $f(\phi_t(p)) = t + f(\phi_0(p)) = t + f(p)$. So, if we take a point p with $f(p) = a$, then $f(\phi_{b-a}(p)) = (b - a) + a = b$, hence ϕ_{b-a} maps M^a into M^b. Similarly, $\phi_{b-a}^{-1} = \phi_{a-b}$ maps M^b into M^a, hence M^b is diffeomorphic to M^a, and this proves the first part of the theorem.

For the second part of the theorem we can construct the deformation retract explicitly. Consider the homotopy $r_t : M^b \to M^b$ defined by

$$r_t(x) = \begin{cases} x & f(x) \leq a \\ \phi_{t(a-f(x))}(x) & a \leq f(x) \leq b \end{cases}$$

We clearly have $r_0 = \mathrm{id}$ and r_1 is a retract of M^b onto M^a, therefore M^a is a deformation retract of M^b. $\qquad\square$

Theorem 4.2. *Let $f : M^n \to \mathbb{R}$ be a smooth function with a non-degenerate critical point p of index μ. Let $c = f(p)$ and suppose $f^{-1}[c - \varepsilon, c + \varepsilon]$ is compact, and contains no critical points of f other than p. Then for small enough ε, the manifold $M^{c+\varepsilon}$ is diffeomorphic to $M^{c-\varepsilon}$ with an n-dimensional μ-handle attached.*

Proof. See figure 4.1 for the local behavior of f around the critical point p. Take a chart (x, U) around p as in the Morse lemma. Then the lightly shaded region is the set of points $(x^1, \ldots, x^n)$ with $f(x^1, \ldots, x^n) \leq c - \varepsilon$, and the medium shaded region is the set of points with $c - \varepsilon \leq f(x^1, \ldots, x^n) \leq c + \varepsilon$. The intersection of the medium shaded region with the $(x^1, \ldots, x^\mu)$-axis is the set of points $(x^1, \ldots, x^\mu)$ with $(x^1)^2 + \cdots (x^\mu)^2 \leq \varepsilon$, hence a μ-dimensional ball. The intersection of the medium shaded region with the $(x^{\mu+1}, \ldots, x^n)$-axis is the set of points $(x^{\mu+1}, \ldots, x^n)$ with $(x^{\mu+1})^2 + \cdots + (x^n)^2 \leq \varepsilon$, hence a $(n - \mu)$-dimensional ball.

The dark region h in the figure consists of the set of points $(x^1, \ldots, x^n)$ such that

$$(x^1)^2 + \cdots + (x^\mu)^2 - (x^{\mu+1})^2 - \cdots - (x^n)^2 \leq \varepsilon \quad \text{and} \quad (x^{\mu+1})^2 + \cdots + (x^n)^2 \leq \delta$$

for sufficiently small δ. Clearly this region is diffeomorphic to $D^\mu \times D^{n-\mu}$, and intersects $M^{c-\epsilon}$ at the points $\partial D^\mu \times D^{n-\mu}$, i.e.

$$(x^1)^2 + \cdots + (x^\mu)^2 - (x^{\mu+1})^2 - \cdots - (x^n)^2 = \varepsilon \quad \text{and} \quad (x^{\mu+1})^2 + \cdots + (x^n)^2 \leq \delta$$
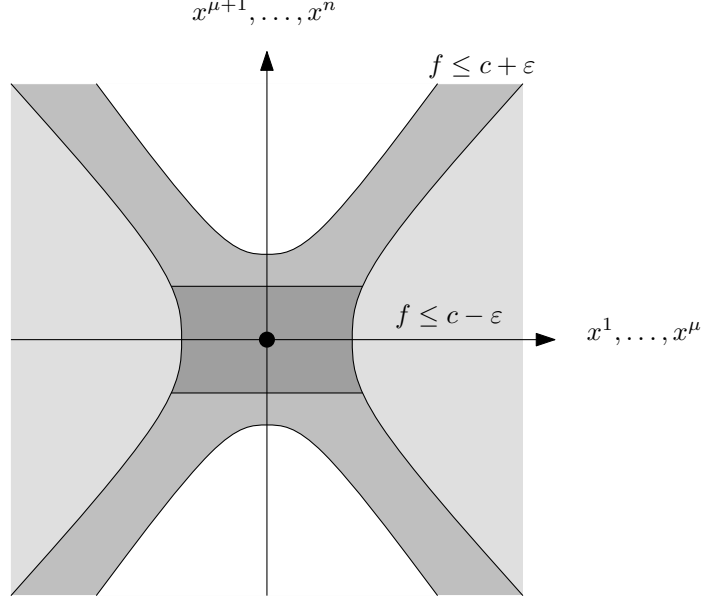
Figure 4.1: Behavior of f around a critical point

One can now use the gradient flow of f to show that $M^{c+\epsilon}$ is diffeomorphic to $M^{c-\epsilon} \cup h$, hence a μ-handle was attached.

$\square$

Theorem 4.2 and figure 4.1 also give an explicit description of the attaching spheres of the handle attached. If the handle h is attached to $M^{c-\varepsilon}$ to obtain $M^{c+\varepsilon}$, then the attaching sphere of h is precisely where the descending manifold of p intersects the boundary of $M^{c-\varepsilon}$.

Theorem 4.3. *If f is a Morse function on a manifold M, and if each M^a is compact, then M has a handlebody decomposition with one μ-handle for each critical point of index μ.*

4.3 Properties of Gradient Flows

Let us further investigate the properties of flows of gradient vector fields. For this section all of our manifolds will be closed.

Proposition 4.1. *Every function decreases along its negative gradient flow lines.*

Proof. Let $f : M \to \mathbb{R}$ be C^1 and $u : \mathbb{R} \to M$ a flow line of $-\nabla_g f$. Since $f \circ u$ is C^1 we have, by

the fundamental theorem of calculus, that for any $b > a$

$$f(u(a)) - f(u(b)) = \int_b^a \frac{d}{dt} f \circ u(t)\, dt$$

$$= -\int_a^b df_{u(t)}(du/dt)\, dt$$

$$= -\int_a^b df_{u(t)}(-\nabla_g f(u(t)))\, dt$$

$$= -\int_a^b g(\nabla_g f(u(t)), -\nabla_g f(u(t)))\, dt$$

$$= \int_a^b ||\nabla_g f(u(t))||\, dt$$

$$\geq 0$$

Therefore $f(u(a)) \geq f(u(b))$, and so f decreased along u. $\qquad\square$

Proposition 4.2. *Letting u be a flow line of f, if $u(\infty) = p$ or $u(-\infty) = p$, then p is a critical point of f.*

Proof. First suppose $u(\infty) = p$ (the other case is similar). We have

$$\lim_{t \to \infty} \int_0^t ||du/dt(s)||^2\, ds = \lim_{t \to \infty} (f(u(0)) - f(u(t))) = f(u(0)) - f(p) < \infty$$

This forces $||du/dt(t)|| \to 0$ as $t \to \infty$. Therefore

$$||\nabla_g f(p)|| = \lim_{t \to \infty} ||du/dt(t)|| = 0$$

and so p is a critical point of f. $\qquad\square$

Proposition 4.3. *If p is a non-degenerate critical point of f, and u is a negative gradient flow line with $u(\infty) = p$, then there are constants $C > 0, \epsilon > 0$ and $T >> 0$ such that*

$$d(u(t), p) \leq Ce^{-\epsilon t}$$

for all $t \geq T$.

4.4 Existence of Morse Functions

We have proved quite a bit concerning Morse functions, but we have not yet shown that they always exist. It turns out that not only do they exist, but in some sense, almost all smooth functions satisfy the Morse property. We will prove this result and show that Morse functions can be chosen with special properties, such as with "separated" critical points and self-indexing critical points.

Since most theorems that involve the phrase "almost every" invoke Sard's theorem in some way, we begin with this result. Recall that a smooth function $f : M \to N$ is said to have a **critical point** at $p \in M$ if $\operatorname{rank} f_{*p} < \dim N$, and a point $q \in N$ is said to be a **critical value** of f if $q = f(p)$ for some critical point $p \in M$.

Theorem 4.4 (Sard's Theorem)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a smooth function. The critical values of f have Lebesgue measure zero in $\mathbb{R}^m$.*

Sard's theorem gives us an easy proof that Morse functions are numerous on open sets of $\mathbb{R}^n$.

Proposition 4.4. *Let $f : U \to \mathbb{R}$ be a smooth function on an open set $U \subseteq \mathbb{R}^n$. Then there exists numbers $a_1, \ldots, a_n$ such that*

$$\widetilde{f}(x^1, \ldots, x^n) = f(x^1, \ldots, x^n) - a_1 x^1 - \cdots - a_n x^n$$

has all non-degenerate critical points. Further, the a_i's can be chosen arbitrary small.

Proof. Define a function $h : U \to \mathbb{R}^n$ by

$$h = \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} = \begin{pmatrix} \frac{\partial f}{\partial x^1} \\ \vdots \\ \frac{\partial f}{\partial x^n} \end{pmatrix}$$

Then the Jacobian of h is

$$J_h = \left(\frac{\partial h_i}{\partial x^j} \right) = \left(\frac{\partial^2 f}{\partial x^i \partial x^j} \right)$$

which is just the Hessian H_f of f. So, the critical points of h are precisely where $\det H_f = 0$. By theorem 4.4 we can find a point $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$ that is not a critical value of h. In fact, there are many such points, and we can choose this point as close to 0 as we want. For this point, define $\widetilde{f} : U \to \mathbb{R}$ by

$$\widetilde{f}(x^1, \ldots, x^n) = f(x^1, \ldots, x^n) - a_1 x^1 - \cdots - a_n x^n$$

Let $p \in U$ be a critical point of $\widetilde{f}$, then

$$0 = \frac{\partial \widetilde{f}}{\partial x^i}(p) = \frac{\partial f}{\partial x^i}(p) - a_i$$

hence $h(p) = a$. Since a is not a critical value of h we have that p is not a critical point of h, and so $\det H_f \neq 0$. Since f and $\widetilde{f}$ differ only by a linear function we have that their Hessians are equal, so $\det H_{\widetilde{f}} = \det H_f \neq 0$, hence p is a non-degenerate critical point of $\widetilde{f}$. $\qquad\square$

We can use this proposition to prove the more general result

Theorem 4.5. *There exists a Morse function on any closed manifold M.*

Proof. Start with any smooth function $f_0 : M \to \mathbb{R}$. We will make a sequence of small changes to this function to remove degenerate critical points over open sets.

Let $\{U_i\}$ be a finite open cover of M by charts, and let $\varphi_i : U_i \to \mathbb{R}$ be a collection of bump functions for $\{U_i\}$. Recall that this means that there is a compact set $K_i \subset U_i$ such that $\varphi_i \equiv 1$ on K_i, $0 \leq \varphi \leq 1$ on U_i and $\varphi_i \equiv 0$ outside U_i. Without loss of generality we can assume that $\{K_i\}$ covers M. Let $\mathcal{K}_m = K_1 \cup \cdots \cup K_m$ and set $K_0 = \emptyset$.

We will use an induction argument to remove degenerate critical points from $\mathcal{K}_m$. Clearly f_0 has no degenerate critical points in $\mathcal{K}_0$, so let us assume that f_{m-1} has been constructed with no degenerate critical points in $\mathcal{K}_{m-1}$. Let $x : U_m \to \mathbb{R}^n$ be a chart for U_m, then $f_{m-1} \circ x^{-1} : x(U_m) \to \mathbb{R}$ is a function defined on an open set of $\mathbb{R}^n$. By proposition 4.4 we can find small vector $a = (a_1, \ldots, a_n)$ such that

$$f_{m-1} \circ x^{-1}(x^1, \ldots, x^n) - a_1 x^1 - \cdots - a_n x^n$$

has only non-degenerate critical points. We can extend this function to all of M by multiplying by the bump function, so we define

$$f_m(p) = \begin{cases} f_{m-1}(p) - a \cdot x(p)\varphi_m(p) & p \in U_m \\ f_{m-1}(p) & p \notin \operatorname{supp}\varphi_m \end{cases}$$

This is a well-defined smooth function on M since $\varphi_m \equiv 0$ on the overlap of U_m and the complement of $\operatorname{supp}\varphi_m$. Now f_m does not have any degenerate critical points on K_m, and since f_{m-1} did not have any degenerate critical points on $\mathcal{K}_{m-1}$ it follows that f_m is a Morse function on $\mathcal{K}_m$. This completes the proof by induction. $\qquad\square$

The proof of theorem 4.5 actually shows a little more. Since we could pick the numbers $a_1, \ldots, a_n$ very small at each stage we could actually show that the constructed Morse function is C^2-close to the initial function f_0, hence there is a Morse function close to any smooth function.

(**note:** Other theorems to prove)

Theorem 4.6. *There is a Morse function* $f : M \to \mathbb{R}$ *on any closed manifold* M *such that if* $p, q \in M$ *are critical points, then* $f(p) \leq f(q)$ *if and only if* $\mu(p) \leq \mu(q)$.

Such Morse functions are called **self-indexing**, and clearly the image of M under f is precisely $[0, n]$, where n is the dimension of the manifold.

5 Kirby Calculus

5.1 Surgery on Links

We have seen that Heegaard diagrams provide a nice combinatorial picture for 3-manifolds, even if they are a little unwieldy to work with. This is mostly due to the fact that self-homeomorphisms of high genus surfaces are difficult to understand. We will now present another combinatorial method of constructing 3-manifolds, called surgery on links. It has the advantage that we now only need to understand self-homeomorphisms of the torus, but the disadvantage that we now need to understand links better.

Given a 3-manifold M, let k be an embedded knot in M and $N(k)$ a tubular neighborhood, which is homeomorphic to $D^2 \times S^1$. Surgery on this knot is the proccess of cutting out the tubular neighborhood and gluing it back in. Since $M\backslash N(k)$ has boundary homeomorphic to $S^1 \times S^1$ we can find a homeomorphism $\varphi : \partial(D^2 \times S^1) \to \partial(M\backslash N(k))$ and form the quotient space $M_{k,\varphi} = (M\backslash N(k)) \cup_\varphi (D^2 \times S^1)$, which we call the **Dehn surgery on k via φ**, or simply **surgery on k** since we are not going to discuss more general surgeries. The construction generalizes to links by performing it on each link component. Every closed, orientable 3-manifold can be obtained by such a construction applied to $M = S^3$, as we will see later.

Proposition 5.1. *A surgery $M_{k,\varphi}$ is completely determined by the image of $\partial D^2 \times \{*\}$ under φ.*

Proof. The solid torus $D^2 \times S^1$ can be attached in two steps. Let J be a small segment of S^1 containing $*$. Then we can start by attaching $D^2 \times J$ to $M\backslash N(k)$. This leaves $D^2 \times J^c$ to be attached along its boundary, which is just a 3-ball being attached along a 2-sphere. There is only one way to attach this piece since every orientation preserving homeomorphism of S^2 is isotopic to the identity. $\square$

So now a surgery can be described by a knot k and a curve on $\partial(M\backslash N(k))$, which is where the meridian will be mapped under the attaching map. We can come up with a more combinatorial description of this curve, or more specifically, closed curves on the torus. Let α and β be the meridian and longitude of the torus $T^2 = S^1 \times S^1$ so that $\pi_1(T^2) = \mathbb{Z} \oplus \mathbb{Z}$ is generated by $[\alpha]$ and $[\beta]$. For p, q integers let $\gamma = p\alpha + q\beta$ denote a curve in the homotopy class $p[\alpha] + q[\beta]$, i.e. γ wraps p times around the meridian and q times around the longitude. We have the following characterization of these curves.

Proposition 5.2.

1. *If a curve γ in the homotopy class $p[\alpha] + q[\beta]$ has no self-intersections, then either p and q are relatively prime, or one of them is 0 and the other ± 1.*

2. *If two closed curves on the torus are homotopic and without self-intersections, then they are isotopic.*

Therefore isotopy classes of curves on T^2 are in one-to-one correspondence with pairs of relatively prime integers.

Let us apply this to a knot k in S^3. Let μ be a canonical meridian of $\partial N(k)$, and let ℓ be the canonical longitude of $\partial N(k)$. Recall that this means that $\mathrm{lk}(\ell, k) = 0$. Choose orientations on μ and ℓ such that the ordered basis (μ, ℓ, n) matches the orientation of S^3, where n is the outward pointing normal to $\partial N(k)$. If we glue a solid torus $D^2 \times S^1$ into $S^3\backslash N(k)$ via $\varphi : \partial D^2 \times S^1 \to \partial(M\backslash N(k))$, then we know φ must map the meridian $\partial D^2 \times \{*\}$ to a curve $p\mu + q\ell$, for some relatively prime

integers p, q (or possibly one equal to 0 and the other ± 1). Therefore a surgery can now be described simply by a knot and an extended rational number $p/q \in \mathbb{Q} \cup \{\infty\}$. This is also called **rational surgery**. If $q = \pm 1$, then the surgery is called an **integral surgery**. We call the rational number p/q the **framing index** of k, or simply the **framing**. This discussion applies equally well to knots embedded in an integral homology sphere.

Here is a simple proposition relating surgeries to Heegaard splittings, as well as a powerful corollary.

Proposition 5.3. *Let $f_1, f_2 : \partial H \to \partial H'$ be homeomorphisms of the surfaces of two handlebodies such that $f_1 = f_2 \circ \tau_c$, where τ_c is a Dehn twist along a simple closed curve c in ∂H. Set $M_i = H \cup_{f_i} H'$ for $i = 1, 2$. Then M_2 is obtained from M_1 by an integral surgery along a knot $k \in M_1$ isotopic to c.*

Proof. Let k be a knot in H obtained from pushing c a little H. Connect a tubular neighborhood $N(k)$ of k to ∂H via an annulus $A \cong S^1 \times I$ such that $S^1 \times \{0\}$ corresponds to a longitude k' on $\partial N(k)$ and $S^1 \times \{1\}$ corresponds to c. There is a self-homeomorphism φ of $H \backslash N(k)$ that is the result of cutting $H \backslash N(k)$ along A, performing a full twist on one of the rims, and then gluing the rims back together. Clearly φ is just an extension of τ_c to $H \backslash N(k)$, but also its restricted to the torus $\partial N(k)$ is a Dehn twist around k'. Let $M_i' = (H \backslash N(k)) \cup_{f_i} H'$, i.e. M_i' is M_i with a solid torus removed. Then M_1' and M_2' are homeomorphic via the map

$$h(x) = \begin{cases} \varphi(x) & x \in H \backslash N(k) \\ x & x \in H' \end{cases}$$

The fact that M_1 and M_2 become homeomorphic after removing a solid torus means that M_1 is obtained from M_2 via a surgery on k. In particular, the meridian μ of $\partial N(k)$ is wrapped around $\mu + k'$, hence it is an integral surgery. $\qquad\square$

Theorem 5.1 (Lickorish-Wallace). *Every closed, orientable 3-manifold M can be obtained from an integral surgery on a link embedded in S^3.*

Proof. By theorem 3.1 we know that $M = H \cup_{f_2} H'$, where H, H' are handlebodies and f_2 is an orientation reversing homeomorphism of their boundaries. Similarly $S^3 = H \cup_{f_1} H'$. Then $f_2^{-1} \circ f_1$ is an orientation preserving homomeomorphism of a surface of genus g, hence by the Dehn-Lickorish theorem ((**todo:** do this theorem)) we have $f_1 = f_2 \circ \tau_{c_1} \circ \cdots \tau_{c_n}$, where τ_{c_i} is a Dehn twist around a simple, closed curve c_i. Therefore by proposition 5.3 we have that M is a sequence of integral surgeries on knots in S^3, or equivalently an integral surgery on a link. $\qquad\square$

A **surgery diagram** for S^3 is a link diagram with each component labeled with an element of $\mathbb{Q} \cup \{\infty\}$. The canonical longitude of a knot k can be drawn in a diagram by just drawing a parallel copy of the knot next to k such that its linking number with k is 0. However, take note that the first Reidemeister move will change this canonical longitude, so one needs to be careful. We will now explain the subtleties of this situation.

Consider the case of integral surgeries on knots in S^3, so let k be a knot with integral framing n. This framing uniquely determines (up to isotopy) a longitude ℓ of k such that $\mathrm{lk}(k, \ell)$, which is the curve we glue the meridian to when we perform the surgery. What is the meaning of the Reidemeister moves in a surgery diagram? It is clear that performing R_2 and R_3 moves does not change the manifold resulting from performing surgery on the knot, so we will continue calling these moves R_1 and R_2 on surgery diagrams. However, performing an R_1 move makes the longitude
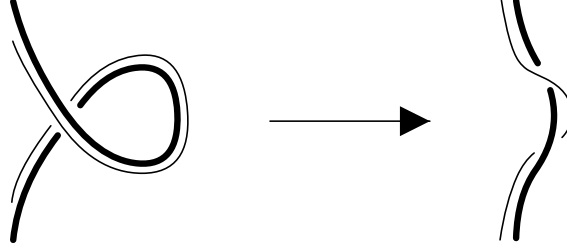
Figure 5.1: Belt Trick
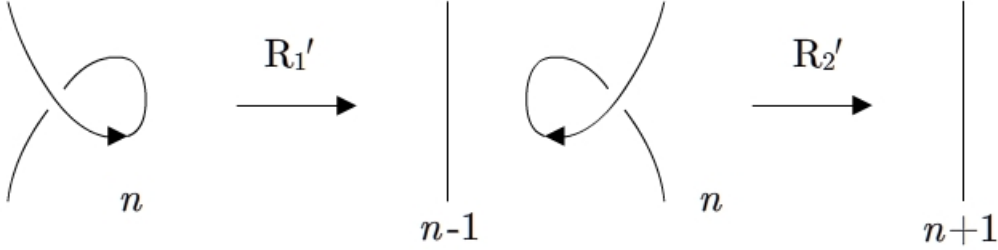
Figure 5.2: Ribbon Reidemeister I

wrap one more time around the knot. This is the so-called "belt trick," and is shown in figure 5.1. This means that the framing of the knot will need to change by ± 1, depending on which way you twist. This move on surgery diagrams will be called R_1' due to its equivalent interpretation as a Reidemeister move for ribbons, and is shown in figure 5.2. The inverse of the first R_1' is called adding a *left twist*, and the inverse of the other R_1' move is called adding a *right twist*. Now two surgery diagrams related by a finite number of R_1', R_2 and R_3 moves give homeomorphic manifolds, but does the converse hold? Unfortunately no, but there are two more moves one can add, called Kirby moves, for which the converse holds. This is discussed in section 5.2.

Example 5.1. The $1/0 = \infty$ surgery on the unknot in S^3 results in S^3 again since we are just cutting out an unknotted, solid torus and gluing it back in the same way. Performing a 0-surgery on the unknot means to cut out an unknotted, solid torus H from S^3 and glue it back in such that the meridian wraps once around the longitude. However, this longitude is precisely the meridian of the the solid torus $S^3 \backslash H$, so this surgery is really just gluing two solid tori together so that their meridians are identified, hence this is $S^1 \times S^2$.

More generally, a surgery on the unknot in S^3 with surgery coefficient $-p/q$ (p and q relatively prime or one equal to 0 and the other 1) results in $L(p, q)$. This is because we cut out an unknotted, solid torus H from S^3 and glue it back it so that the meridian wraps p times around the longitude and q times around the meridian. However, H's meridian is precisely $S^3 \backslash H$'s longitude, and H's longitude is precisely $S^3 \backslash H$'s meridian, hence we are just gluing together two solid tori so that the meridian of one gets wrapped p times around the longitude and q times around the meridian of the other torus. This is precisely the Heegaard splitting description of $L(p, q)$. Compare this with example 3.6. We can also find a surgery diagram for $L(p, q)$ that has only integral surgery coefficients. See figure 5.3.

Let k_i, $1 \leq i \leq n$, denote the components of a surgery diagram of some link, let k_i be framed by p_i/q_i and let M be the 3-manifold resulting from performing this surgery. If we let $K = S^3 \backslash (N(k_1) \cup \cdots \cup N(k_n))$ be the knot complement, then $H_1(K) = \mathbb{Z}^n$ and is generated by the

54
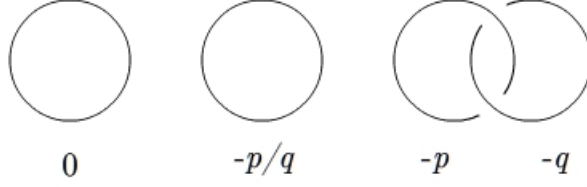
Figure 5.3: Surgery diagrams for $S^1 \times S^2$, $L(p,q)$ and $L(pq - 1, q)$

meridians $\{\mu_i\}$ of the link. When we glue in a solid torus $S^1 \times D^2$ so that $\{*\} \times \partial D^2$ is wrapped p_i times around the meridian of k_i and q_i times around the longitude, when add a relation to $H_1(K)$. Namely, we add the relation $p_i\mu_i + q_i\lambda_i = 0$, where λ_i is the canonical longitude of k_i, since we are essentially filling this curve with a disc. The longitudes can be expressed as a relation in terms of the μ_i's. Since $\mathrm{lk}(k_i, \lambda_i) = 0$ we have that λ_i bounds a surface F_i (i.e. a Seifert surface) in $S^3 \backslash N(k_i)$ (but not necessarily in K). Cut a small disc out of F_i at each intersection point of a different component k_j with F_i, and orient it positively if the intersection is positive and negatively if the intersection is negative. Since k_j and F_i intersect $\mathrm{lk}(k_i, k_j)$ times (counted with sign), we have $\lambda_i = \sum_{j \neq i} \mathrm{lk}(k_i, k_j)\mu_j$. We have proved the following.

Proposition 5.4. *For a surgery diagram with components $k_1, \ldots, k_n$ and framings $p_1/q_1, \ldots, p_n/q_n$, the first homology of this surgery is the free abelian group on n generators modulo the relations $p_i\mu_i + q_i \sum_{i \neq j} \mathrm{lk}(k_i, k_j)\mu_j = 0$.*

Corollary 5.1. *For a surgery diagram with components $k_1, \ldots, k_n$ and integral framings $p_1, \ldots, p_n$, let A be the linking matrix of this link with respect to any orientation. Then the first homology of this surgery is isomorphic to $\mathbb{Z}^n / \mathrm{im}\, A$.*

Corollary 5.2. *If the linking matrix A has $\det A = \pm 1$, then the surgery is a homology sphere. In particular, a surgery on any knot with framing ± 1 is a homology sphere.*

5.2   Kirby Moves

We now describe two basic moves that can be performed on surgery diagrams that do not change the resulting surgered manifold. If a link L can be written as $L = L_1 \cup L_2$, where L_1 and L_2 are links that can be separated by a ball, then clearly a surgery on L is diffeomorphic to the connected sum of the surgeries on each L_1 and L_2. So, if we add an unknotted, unlinked component to any link L with framing ± 1, surgery on this new link is the connected sum of the surgery on L and the surgery on the unknot with framing ± 1, which is just S^3, hence the manifold is unchanged. The addition or removal of an unknotted, unlinked component with framing ± 1 from a surgery diagram is called the **first Kirby move** (K_1).

Next, suppose our link L is expressed as $L = k_1 \cup k_2$, where k_1 and k_2 are knots with framing n_1 and n_2, and are not necessarily unlinked. Let ℓ be a longitude of k_2 defining its framing (i.e. $\mathrm{lk}(k_2, \ell) = n_2$), and let k_1' be the connected sum of k_1 with ℓ along a band disjoint from L and ℓ, and let $L' = k_1' \cup k_2$. We then give k_1' the framing $n_1 + n_2 + 2\,\mathrm{lk}(k_1, k_2)$, where the linking number is computed with respect to the orientations on k_1 and k_2 that induce an orientation on k_1' (there are only two, and both have the same linking number). Then surgery on L gives the same manifold as surgery on L', but we will wait to show this until later. This process of "adding" the component k_1 to k_2, sometimes also called "sliding" k_1 across k_2, is the **second Kirby move** (K_2). Note that
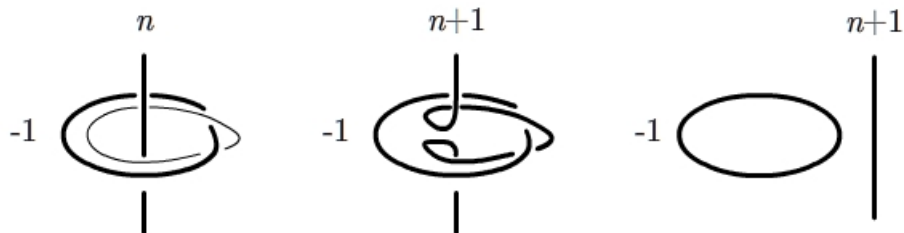
Figure 5.4: Separating an unknotted component from a strand

the link L' will of course depend on the band we choose in forming k_1'. Amazingly these two moves are enough to get between any two surgery diagrams of the same manifold.

Theorem 5.2 (Kirby). *Two manifolds presented as a surgery on some links are orientation preservingly homeomorphic if and only if one link can be obtained from the other via a finite sequence of K_1 and K_2 moves.*

We shall now develop many "secondary" moves from K_1 and K_2 that are easier to apply directly.

Example 5.2. Suppose there is an unknotted component of L with framing ± 1 that is linked once with a single strand from another component of L with framing n. We claim that we can separate the unknotted component from the rest of the link, in which case the framing of the linked stranded changes by ∓ 1. See figure 5.4 for the case of the unknot framed by -1, where the bolded curves are the components of the link and the non-bolded curve is the longitude of the unknot determining its framing. Note that the linking number of the vertical strand with the unknotted component is $+1$, so the framing of the vertical strand changes to $n - 1 + 2 \cdot 1 = n + 1$. The framing of this strand does not change from the 2nd frame to the 3rd because the two R_1' moves performed cancel each other out. The case where the unknot is framed by $+1$ is similar.

Example 5.3. Suppose now there is an unknotted component of L with framing ± 1 that is linked once with *two* strands from other components of L. All the steps are shown in figure 5.5. We can see that separating this unknotted component resulted in a full right twist in the two strands. If the two strands belong to two different link components, then each of their framings will increase by 1, and if they belong to the same component, the framing will increase by 2.

Example 5.4. The general picture of the previous two examples is that if we can find an unknotted component of L with framing ± 1 that is linked once with k strands, then we can separate the unknot from these strands by adding a full left/right twist in the strands, and changing their framings accordingly. This type of move (see figure 5.6) is called the **Fenn-Rourke move**. There is another name for this operation due to its relation to the 4-manifold theory of Kirby calculus. If we use the first Kirby move to introduce an unknotted, unlinked component with framing ± 1, we say that we have performed a **blow-up** operation. Conversely, if we find an unknotted component with framing ± 1 that links some strands once, then the operation of sliding the unknotted component off the strands and using K_1 to discard it is called the **blow-down** operation.

Example 5.5. Suppose our link L can be written as $L = \cup_{i=0}^n L_i$ such that L_0 is the unknot with framing 0 and links with only L_1 once (the other L_i's are free to link with L_1 in any way). We claim that we can use K_2 moves to change any crossing of some component L_i $(i \geq 1)$ passing *under* L_1 to an over crossing. By changing all such crossings to over crossings we would have completely separated $L_0 \cup L_1$ from all of $\cup_{i=2}^n L_i$, and we would have completely unknotted L_1.
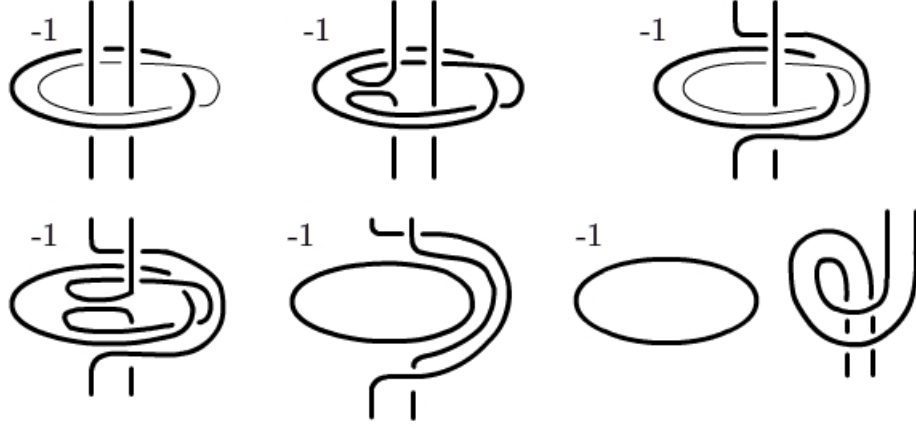
Figure 5.5: Separating an unknotted component from two strands
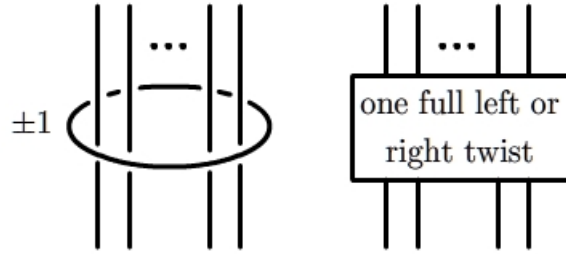
Figure 5.6: Separating an unknotted component from many strands

Suppose a strand of L_i crosses *underneath* L_1. We can isotopy L_0 along L_1 so that it is close to this crossing (in a diagram we only need R_1 and R_2 moves). Sliding L_i over L_0 and then applying R_1 and R_2 allows us to change this crossing to an over crossing (see figure 5.7). If $i > 1$, then the framing of L_i does change from this Kirby move since L_0 has framing 0 and L_i is not not linked with L_0. However, if $i = 1$, then the framing of L_1 is changed to $n_1 + 2 \operatorname{lk}(L_0, L_1)$, where n_1 is the framing of L_1.

Once all the crossings have been changed so that $L_0 \cup L_1$ is separated from the rest of the link and L_1 is unknotted to become L_1', we are left to deal with a Hopf link $L_0 \cup L_1'$ with one component framed by 0 and the other by some integer, say p. By sliding L_1' over L_0 like we did above we see that the framing of L_1' changes by ± 2, so doing this repeatedly we can reduce this to the case when the Hopf link is framed by 0 and 1, or 0 and 0. Performing surgery on either of these links just gives S^3, so we can remove them entirely. In the end, we have separated $L_0 \cup L_1$ from the rest of L and cancelled it.

Example 5.6. There is a simple way of expressing a surgery on the connected sum of two knots $k_1 \# k_2$ and the surgery on each knot separately. Suppose k_1 and k_2 have framings p and q respectively, and are linked once by a 0-framed unknot k_0. If we slide k_1 over k_2 to get k_1', then we see that k_0 now only links k_2. So, applying the move in example 5.6 allows us to separate $k_0 \cup k_2$ from k_1' and cancel it. The knot that is left, k_1', is precisely $k_1 \# k_2$ with framing $p + q$. See **??**. (**todo: make this pic**)

A graph embedded in the plane with vertices v_i decorated by elements $p_i \in \mathbb{Q} \cup \{\infty\}$ induces a surgery diagram in the following way. Place an unknotted, unlinked circle k_i at each vertex v_i and
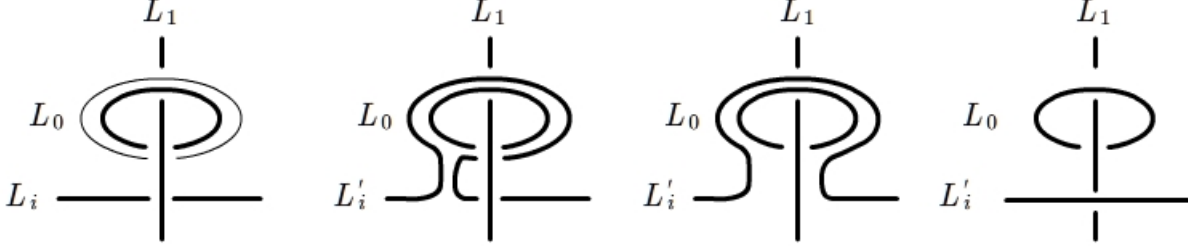
Figure 5.7: Changing under crossings to over via a 0-framed unknot

frame it by p_i. If two vertices are connected by an edge, then link the corresponding circles once. If the graph is a tree, then there is a unique way to do this. These surgery tries provide a simply way to express surgery diagrams that are composed of unknotted components linked together in the simplest way.

We have seen that a surgery on any knot with framing ± 1 results in an integral homology sphere. There is a particular homology sphere that is ubiquitous in low dimensional topology called the Poincaré sphere. We will define it to be the space obtained by surgery on the left-hand trefoil with framing -1.

Example 5.7. We will show that the Poincaré sphere is homeomorphic to performing surgery on the E_8 Dynkin diagram where each vertex is framed by -2. We blow down this diagram three times by adding 3 unknotted components with framings $+1$ and sliding them over the 3 ends of the diagram. Then we blow up 8 times to remove all components with framing -1. At this point we have 3 components with framing 3, 2 and 5 linked to a component with framing 1. Blowing this up we get two components with framing 2 and 4 linked to a component with framing 1. Blowing up again gives a component of framing 3 linked with a component of framing 1, and a final blow of gives a single knot with framing 2. Performing a few R_1' moves gives the trefoil with framing -1. (**todo:** make pic)

5.3 The Rational Calculus

Now we will extend our moves on integral surgeries to rational surgeries. The first move we will consider is called a **Rolfsen twist**. Recall that the lens spaces $L(p, q)$ and $L(p, q + np)$ are homeomorphic for any integer n; that is, the surgeries on the unknot with rational coefficients $\frac{p}{q}$ and $\frac{p}{np+q}$ are homeomorphic. We will describe how to explicitly visualize this homeomorphic, which will then lead us to a new move on rational surgeries.

Let H be the standardly embedded solid torus in S^3 and T its boundary torus. We can construct $L(p, q)$ by a p/q surgery on the unknot in S^3, which is equivalent to gluing two solid tori together via a map $f : T \to T$ that maps the meridian μ to the curve $q\mu + p\ell$, where ℓ is the longitude. If $\tau : T \to T$ is a right Dehn twist around μ, then each time the curve $q\mu + p\ell$ intersects μ (which is p times) the curve $\tau(q\mu + p\ell)$ picks up another intersection with ℓ, hence $\tau(q\mu + p\ell) = (q + p)\mu + p\ell$. Since τ extends to a homeomorphism of the solid torus H that T bounds we have that it also extends to a homeomorphism of $H \cup_f H$ and $H \cup_{\tau \circ f} H$ by our gluing lemma (lemma 3.1). This shows explicitly that $L(p, q)$ is homeomorphic to $L(p, p + q)$. We can generalize this construction by letting τ be an n-fold Dehn twist around μ (where it is a right twist if $n > 0$ and a left twist if $n < 0$), in which case we get an explicit homeomorphism of $L(p, np + q)$ with $L(p, q)$. If $r = p/q$, then we can also rephrase this as the r surgery on the unknot is homeomorphic to the $\frac{1}{n + \frac{1}{r}}$ surgery
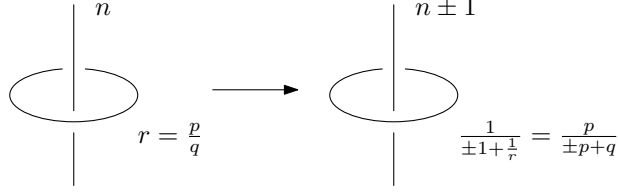
Figure 5.8: The Rolfsen Twist

on the unknot for any n.

We can now describe the Rolfsen twist. Suppose we have a rationally framed link in S^3 such that one component k is framed with an integer n and is linked once with an unknotted component U which has any rational framing $r = p/q$. We claim that surgery on this link is equivalent to surgery on the link formed by changing k's framing to $n \pm 1$ and U's framing to $\frac{p}{\pm p + q} = \frac{1}{\pm 1 + \frac{1}{r}}$. We will show this by mimicking the above arguments to describe an explicit homeomorphism between the spaces.

First we remove a tubular neighborhood νU of U from S^3 to obtain the space X, and let $f : \partial \nu U \to \partial X$ be the gluing homeomorphism for the surgery on U; that is, f maps the meridian of $\partial \nu U$ to q times around the meridian and p times around the longitude of ∂X. Let τ be the homeomorphism of X that cuts the space open along a disc $D \subset X$ such that ∂D is the longitude on $\partial \nu U$, performs a full right twist, then glues everything back, and is the identity outside a small ball containing D. Now glue U back into X via the map $\tau \circ f$. Then clearly this surgery is equivalent to the original surgery, except now U is framed by $\frac{p}{p+q}$ and k is framed by $n + 1$.

It turns out that the Rolfsen twists are nearly enough to get between any rational surgery diagrams for the same 3-manifold.

Theorem 5.3. *If two 3-manifolds presented as a rational surgery diagram are orientation preservingly diffeomorphic, then one diagram can be brought into the other via a finite sequence of Rolfsen twists and adding/removing a knot component with framing ∞.*

Let us look at many examples using these rational calculus moves.

Example 5.8.

Proposition 5.5. *We can convert any rational surgery diagram L for a 3-manifold into an integral surgery diagram L' for the same 3-manifold by adding unknotted components.*

(**todo:** surgery diagram for $L(p, q)$ in terms of continued fraction expansion of p/q)
(**todo:** any rational surgery can be turned into an integral surgery)

6 4-Manifold Theory

6.1 Intersection Forms on 4-Manifolds

Let M be a closed, oriented, connected and simply connected 4-manifold, and let $[M]$ denote its fundamental class. Since M is simply connected we have $H_1(M) = 0$, hence $H^3(M) = 0$ by Poincaré duality. Further, the split exact sequence of the universal coefficient theorem says that

$$H^2(M) = \text{Ext}(H_1(M), \mathbb{Z}) \oplus \text{Hom}(H_2(M), \mathbb{Z}) = \text{Hom}(H_2(M), \mathbb{Z})$$

Since $\text{Hom}(A, B)$ is torsion free if B is free, we have that $H^2(M)$ is torsion free, and hence $H_2(M)$ is torsion free by Poincaré duality. Define a bilinear form $Q_M : H^2(M) \otimes H^2(M) \to \mathbb{Z}$ by

$$Q_M(\alpha, \beta) = (\alpha \smile \beta, [M])$$

This symmetric bilinear form is called the **intersection form** of M. This form is non-degenerate, which means that its matrix in *any* basis is non-singular, i.e. determinant $= \pm 1$.

There is a more geometric way of defining this bilinear form in terms of counting intersections of embedded surfaces. If M is a closed, smooth, oriented n-manifold, we say that a homology class $a \in H_k(M)$ is *represented* by an oriented, smoothly embedded k-submanifold $i : N \to M$ if $i_*[N] = a$. Then we have the following

Lemma 6.1. *Let M be a closed, oriented, smooth 4-manifold. Any homology class $a \in H_2(M)$ can be represented by a smoothly embedded surface in M.*

Proof. We will prove this for the case when M is simply connected, which is the case we need the most for our later discussions. Since M is simply connected, the Hurewicz homomorphism $h : \pi_2(M) \to H_2(M)$ is an isomorphism. Recall that this map is defined by $h([f]) = f_*[S^2]$, so every homology class $\alpha \in H_2(M)$ has a corresponding classifying map $f : S^2 \to M$ such that $f_*[S^2] = \alpha$. Since $\dim M = 4$ we have that f can be slightly homotoped so that f is an immersion with finitely many isolated double points. Locally, a double point looks like two planes in $\mathbb{R}^4$ intersecting at a point P. We can cut a small disc out of each plane around P and connect the planes with a cylinder that misses P. Performing this operation on each double point of $f(S^2)$ we get a smoothly embedded surface in M representing α. $\square$

So, for a 4-manifold M and two homology classes $a, b \in H_2(M)$, let $F_a, F_b \subset M$ be smoothly embedded, oriented surfaces representing a and b respectively. We can slightly deform the embeddings, if necessary, so that F_a and F_b intersect transversally in a finite collection of points. Assign a sign of ± 1 to each intersection point p depending on whether or not the induced orientation on $T_p F_a \oplus T_p F_b$ from F_a and F_b agrees with $T_p M$. We call the sum of these numbers the *intersection product* of a and b, and denote it by $a \cdot b$. Then

Proposition 6.1. $Q_M(\alpha, \beta) = PD(\alpha) \cdot PD(\beta)$

If we fix a basis $e_1, \ldots, e_k$ of $H_2(M)$, then we call the matrix $[Q_M] = \{e_i \cdot e_j\}_{ij}$ the matrix form of Q_M. It is clearly symmetric.

Example 6.1. Let $M = S^2 \times S^2$, then $H_2(M)$ is generated by $a = S^2 \times \{y_0\}$ and $b = \{x_0\} \times S^2$ for some base point $(x_0, y_0) \in M$. We can slightly deform a to be completely disjoint from the original a, hence $a \cdot a = 0$, and similarly $b \cdot b = 0$. We cannot deform a so that it is disjoint from b, but

a and b clearly intersect transversally at one point, and by choosing orientations appropriately we have $a \cdot b = 1$. Therefore the matrix form of Q_M with respect to the basis a, b is

$$[Q_M] = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Example 6.2. We will see soon that the intersection form on $\mathbb{C}P^2$ is (1) and on $\overline{\mathbb{C}P}^2$ is (-1). More generally, the intersection form on the D^2-bundle over S^2 with Euler number n is just (n).

Example 6.3. An easy application of the Mayer-Vietoris sequence shows that $H_2(M \# N) \cong H_2(M) \oplus H_2(N)$, and so $Q_{M \# N} = Q_M \oplus Q_N$.

If M is a compact, oriented, connected and simply connected 4-manifold with boundary, then $H^2(M)$ and $H^2(M, \partial M)$ are still torsion free, and so Poincaré-Lefschetz duality gives us a non-degenerate pairing $Q_M : H^2(M) \otimes H^2(M, \partial M) \to \mathbb{Z}$ defined by $Q_M(\alpha, \beta) = (\alpha \smile \beta, [M, \partial M])$. If ∂M is an integral homology sphere, then $H^2(M, \partial M) = H^2(M)$, and so Q_M is a symmetric, non-degenerate pairing of $H^2(M)$ with itself. In terms of intersections of surfaces, we can still define the intersection product $\cdot : H_2(M) \otimes H_2(M) \to \mathbb{Z}$ in the usual way, and if M is an integral homology sphere we still have $Q_M(\alpha, \beta) = PD(\alpha) \cdot PD(\beta)$. There is a converse to the non-degeneracy of Q_M.

Proposition 6.2. *The intersection form Q_M on a compact, oriented, connected and simply connected 4-manifold M with boundary is non-degenerate if and only if ∂M is an integral homology sphere.*

Proof. If ∂M is a disjoint union of integral homology spheres, then we have already seen that Q_M is non-degenerate by Poincaré-Lefschetz duality.

Suppose Q_M is non-degenerate. Since M is simply connected we have $H^1(M) = 0$ by the universal coefficient theorem, hence $H_3(M, \partial M) = 0$ by Poincaré-Lefschetz duality. A part of the long exact sequence of the pair $(M, \partial M)$ looks like

$$0 \longrightarrow H_2(\partial M) \overset{i_*}{\longrightarrow} H_2(M) \overset{j_*}{\longrightarrow} H_2(M, \partial M) \longrightarrow H_1(\partial M) \longrightarrow 0$$

So we have i_* is injective, hence $\operatorname{im} i_* = 0$ if and only if $H_2(\partial M) = 0$. If $\operatorname{im} i_*$ were not trivial, then for any $a \in \operatorname{im} i_*$ and $b \in H_2(M)$ we have $a \cdot b = 0$ since we can move any surface representing b so that it does not intersect the boundary at all. But, Q_M is assume to be non-degenerate, so we do not allow this to happen, hence $\operatorname{im} i_* = 0$, and so $H_2(\partial M) = 0$. Finally, this implies $H_1(\partial M) = 0$, therefore ∂M is an integral homology sphere. $\square$

Intersection forms are very important in 4-dimensional topology. In fact, they nearly classify all simply connected 4-manifolds by a theorem of Freedman. So, it will be useful to recall some basic facts of non-degenerate, symmetric, bilinear forms $Q : \mathbb{Z}^n \times \mathbb{Z}^n \to \mathbb{Z}$. We say that two such forms Q_1, Q_2 are isomorphic if there is an isomorphism $\varphi : \mathbb{Z}^n \to \mathbb{Z}^n$ such that $Q_1(x, y) = Q_2(\varphi(x), \varphi(y))$. The rank of Q is defined to be n. If we think of Q as a bilinear form over $\mathbb{R}$, then we can diagonalize over the reals and there will be b_+ positive values and b_- negative values on the diagonal. These numbers are invariants of the isomorphism type of Q, as is the signature $\operatorname{sign} Q = b_+ - b_-$. We call Q a **definite** form if $b_- = 0$ or $b_+ = 0$, and call Q an **indefinite** form otherwise. We say that Q is **even** if $Q(x, x) = 0 \mod 2$ for all x, and otherwise we say Q is **odd**.

There is a nice relationship between surgeries on links and intersection forms on cobordisms of surgeries. First we will consider a fourth way of defining the linking number of two disjoint, oriented knots (see the other three definitions in section 2.1). For disjoint, oriented knots k_i and k_j

in S^3, let F_i and F_j be Seifert surfaces with their induced orientations. We can push the interiors of these surfaces into the 4-ball D^4 that bounds S^3 so that $F_i \cap S^3 = k_i$ and $F_j \cap S^3 = k_j$. Then, perturbing F_i and F_j slightly so that they intersect transversely if necessary, one can compute that $\mathrm{lk}(k_i, k_j) = F_i \cdot F_j$.

Let X be a 4-dimensional 2-handlebody; that is, a 4-manifold that can be decomposed as 0-handle $\cup$ 2-handles $\cup$ 4-handle. This is equivalent to specifying a link $L = k_1 \cup \cdots \cup k_n$ in ∂D^4 with k_i framed by an integer p_i. For each component k_i, let F_i denote a surface of k_i embedded in $\partial D^4 = S^3$, then we can push this surface a little bit into D^4 to obtain a surface $\hat{F}_i$ such that $\hat{F}_i \cap S^3 = \partial \hat{F}_i = k_i$. Let G_i be the core disc of h_i, then by gluing $\hat{F}_i$ and G_i together along k_i we obtain a closed, oriented, smoothly embedded surface α_i in X. Put any orientation on k_i, then we get induced orientations on $\hat{F}_i$ and G_i, hence α_i also gets an induced orientation.

Proposition 6.3. *The surfaces $\alpha_1, \ldots, \alpha_n$ form a basis for $H_2(X)$.*

Proof. First consider the space obtained from attaching one 2-handle $Y = D^4 \cup_\varphi (D^2 \times D^2)$. Let N be a small open neighborhood of $\varphi(\partial D^2 \times D^2)$, and let $U = D^4 \cup N$ and $V = (D^2 \times D^2) \cup N$. Then U is homotopy equivalent to D^4, V is homotopy equivalent to $D^2 \times D^2$, and $U \cap V$ is homotopy equivalent to $S^1 \times D^2$. The Mayer-Vietoris sequence gives the following exact sequence.

$$H_2(U) \oplus H_2(V) \longrightarrow H_2(Y) \overset{\Delta}{\longrightarrow} H_1(U \cap V) \longrightarrow H_1(U \oplus V)$$

which implies that the connecting homomorphism $\Delta : H_2(Y) \to H_1(U \cap V)$ is an isomorphism, where clearly $H_1(U \cap V) \cong \mathbb{Z}$. To understand the generator of $H_2(Y)$ let us recall how the connecting homomorphism is defined. A cycle $c \in Z_2(Y)$ can be expressed as $c = u + v$ with $u \in C_2(U)$ and $v \in C_2(V)$. This implies $\partial u = -\partial v$, hence $\partial u \in Z_1(U \cap V)$, and so we define $\Delta[c] = [\partial u]$.

The attaching sphere knot k of the 2-handle is the generator of $H_1(U \cap V)$, so we need to determine what element α of $H_2(Y)$ gets mapped to k. By our description of Δ we see that α can be represented by a 2-chain in U and a 2-chain in V whose boundaries are k. Therefore we can take these chains to be the Seifert surface of k pushed into D^4 and the core disc of the 2-handle $D^2 \times D^2$, and so α is can be represented by gluing these surfaces together.

Repeating this argument for each 2-handle in X we get a set of n generators of $H_2(X)$, which we know to be of rank n from its handlebody homology, therefore these generators form a basis. $\square$

We call the basis $\alpha, \ldots, \alpha_n$ the canonical basis of $H_2(X)$. Let us compute the matrix of the intersection form with respect to this canonical basis. First we compute the self intersection number $\alpha_i \cdot \alpha_i$. Translate the core disc G_i to a parallel disc G_i' in the boundary $D^2 \times \partial D^2$. This new disc will intersect the boundary of D^4 at a longitude of k_i, call it k_i', with linking number $\mathrm{lk}(k_i, k_i') = p_i$. Isotope the Seifert surface F_i to a new surface F_i' such that $\partial F_i' = k_i'$, and then push this surface into D^4 to obtain $\hat{F}_i'$ such that $\hat{F}_i' \cap S^3 = \partial \hat{F}_i' = k_i'$ and so that $\hat{F}_i$ and $\hat{F}_i'$ intersect transversally. By gluing $\hat{F}_i'$ to G_i' we obtain a closed surface α_i' that represents the same homology class as α, and clearly

$$\alpha_i \cdot \alpha_i' = (\hat{F}_i \cup G_i) \cdot (\hat{F}_i' \cup G_i') = F_i \cdot F_i' = \mathrm{lk}(k_i, k_i') = p_i$$

Repeating this argument for two different components k_i and k_j, shows that $\alpha_i \cdot \alpha_j = \mathrm{lk}(k_i, k_j)$. So, we have proved the following.

Proposition 6.4. *The intersection form of a 2-handlebody in the canonical basis is equal to the linking matrix of the associated link.*
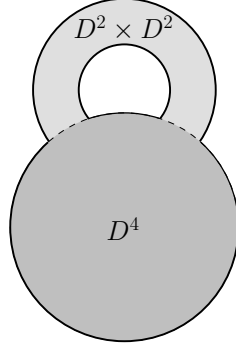
Figure 6.1: Surgery Cobordism

Let $M = \partial X$, i.e. M is the result of performing the surgery on the link L in S^3. We saw in corollary 5.1 that $H_1(M)$ could be computed as $\mathbb{Z}^n/\text{im}[L]$, where $[L]$ is the linking matrix of L. As a corollary we now have

Corollary 6.1. *Let X be a 2-handlebody with intersection form Q_X. Then $H_1(\partial X)$ is finite if and only $\det Q_X \neq 0$, in which case $|H_1(\partial X)| = |\det Q_X|$.*

6.2 Kirby Calculus Reinterpreted

Kirby calculus is essentially a 4-dimensional technique, and so we take another look at the Kirby moves. To bring Kirby calculus into the 4-dimensional world we show that surgeries on manifolds induce cobordisms. Recall that two compact, oriented, smooth n-manifolds M_1 and M_2 are said to be **cobordant** if there is a compact, oriented, smooth $(n+1)$-manifold W such that $\partial W = -M_1 \sqcup M_2$, where $-M_1$ is M_1 with the opposite orientation. A manifold M is said to be **cobordant to zero** if it is cobordant to the empty manifold, i.e. $\partial W = M$.

Proposition 6.5. *The manifold resulting from any surgery on any knot in S^3 is cobordant to zero.*

Proof. Let k be a knot in S^3 with tubular neighborhood $N(k)$, and let $\varphi : \partial(S^1 \times D^2) \to \partial(M \setminus N(k))$ be the attaching map that determines the surgery on k. Let c be the image of any meridian $\{*\} \times \partial D^2$ under φ so that c is a curve in $\partial N(k)$. There is a diffeomorphism $\phi : \partial D^2 \times D^2 \to N(k)$ such that $\phi(\partial D^2 \times \{0\}) = k$ and $\phi(\partial D^2 \times \{*\}) = c$, where $*$ is some point on the boundary of the 2-disc. Let W be the manifold resulting from attaching the 2-handle $D^2 \times D^2$ to the the 4-ball D^4 via $\phi : \partial D^2 \times D^2 \to N(k) \subset \partial D^4$.

The boundary of the 4-ball has changed by attaching this 2-handle. The part of the boundary of D^4 outside of $N(k)$ (which is the knot complement) have remained boundary points, and all the points in $N(k)$ have now become interior points of M. The "free" part $D^2 \times \partial D^2$ of the boundary of the 2-handle have now become boundary points of M (see figure 6.1). This means we have removed a solid torus from M and glued in another solid torus. In fact, the solid torus was glued in by a diffeomorphism that maps the meridian $\partial D^2 \times \{*\}$ to the curve c, hence this is just performing the surgery on k determined by φ. Therefore we have constructed a cobordism from the surgered manifold to the empty manifold, and so it is cobordant to zero. $\qquad \square$

Corollary 6.2. *Every closed, oriented 3-manifold is cobordant to zero.*
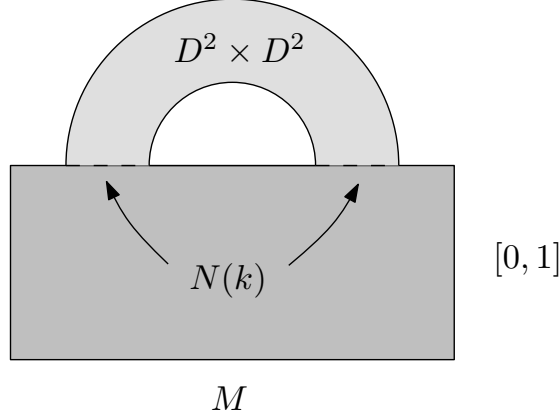
Figure 6.2: Surgery Cobordism

Proof. By proposition 6.5 we have that every surgery on S^3 is cobordant to zero. But, by theorem 5.1 every closed, oriented 3-manifold is a surgery on S^3, hence every such manifold is cobordant to zero. $\qquad\square$

Example 6.4. We saw earlier that D^2-bundles over S^2 are canonically parameterized by $\mathbb{Z}$ by attaching a 4-dimensional 2-handle along an unknotted circle in ∂D^4. Let E_n denote this 4-manifold with boundary, and since ∂E_n is the result of performing a surgery on the unknot with framing n, we see that $\partial E_n = L(n, 1)$. In particular E_0 is the trivial bundle $S^2 \times D^2$. We also saw that $\mathbb{C}P^2$ is a union of a 0, 2 and 4-handle, and the 2-handle is attached to the 0-handle along an unknotted circle. So, which of the 4-manifolds E_n corresponds to $\mathbb{C}P^2 \backslash D^4$, i.e. $\mathbb{C}P^2$ without its 4-handle? In $\mathbb{C}P^2$ the 2-handle $D^2 \times D^2$ is attached to $D^4 = D^2 \times D^2$ via the map $\varphi : \partial D^2 \times D^2 \to \partial(D^2 \times D^2)$ defined by $\varphi(z, w) = (\bar{z}, \bar{z}w)$. By fixing a z and restricting $|w| = 1$ we see that a longitude of $S^1 \times D^2$ is wrapped once around the meridian and once around the longitude of $S^1 \times D^2$. So, when we attach this handle the boundary changes by performing a surgery on the unknot with framing $+1$, hence $\mathbb{C}P^2 \backslash D^4$ is diffeomorphic to E_1. We also have E_{-1} is diffeomorphic to $\overline{\mathbb{C}P}^2 \backslash D^4$.

A slight generalization of the construction in proposition 6.5 gives an explicit cobordism between any manifold and any surgery on that manifold.

Corollary 6.3. *Any integral surgery on any 3-manifold M is cobordant to M.*

Proof. Mimic the construction in proposition 6.5 by attaching a 2-handle to the component $M \times \{1\}$ of the 4-manifold $M \times I$. After the handle is attached one component will be M and the other component will be the result of the sugery (see figure 6.2). $\qquad\square$

For a framed link L, let W_L denote the 4-dimensional cobordism of the surgery on L with zero, i.e. $W_L = $ 0-handle $\cup$ a 2-handle for each component of L. This association of a 4-manifold with boundary to any framed link is what allows us to recast Kirby calculus as a 4-dimensional theory. In fact, this is the way Kirby calculus was originally described. The first Kirby move added an unknotted, unlinked component to L with framing ± 1 to get the link L'. On the level of 4-manifolds this amounts to $W_{L'}$ being diffeomorphic to $W_L \# \mathbb{C}P^2$ or $W_L \# \overline{\mathbb{C}P}^2$, depending on the framing of the added unknot. This follows from example 6.4 where we saw that $E_1 = \mathbb{C}P^2 \backslash D^4$ and $E_{-1} = \overline{\mathbb{C}P}^2 \backslash D^4$.

Next, we claim that performing the second Kirby move on two components of L is equivalent to sliding one handle correspond to one component across the handle corresponding to the other component. Since this whole operation is technically taking place in a 4-dimensional manifold it is difficult to picture it. Figures 3.9 and 3.10 showed how to slide 3-dimensional 2-handles over each other, and the idea in 4-dimensions is similar. Let us attach two 2-handles h_1, h_2 to the 4-ball D^4 with attaching circles k_1, k_2 and framings n_1, n_2 and let $X = D^4 \cup h_1 \cup h_2$. To slide h_1 over h_2 we need to find a disc parallel to the core of h_2 and then isotope the attaching circle of h_1 through this disc. We can take any disc parallel to the core, say $D^2 \times \{p\}$ where $p \neq 0$. The boundary of this disc is precisely where it intersects the 3-sphere, which is at a longitude k_2' of k_2 such that $\mathrm{lk}(k_2, k_2') = n_2$, i.e. its a longitude determined by the framing. As we saw in figure 3.10, isotoping the attaching circle of h_1 through this disc is the same as form the banded sum of the attaching circle with the boundary of the disc, so the attaching circle of h_1 after sliding across h_2 is the connect sum of k_1 with k_2 along some band disjoint from k_1 and k_2. We could compute the framing of this handle after the slide directly, or appeal to proposition 6.4. Namely, let α_1, α_2 be the canonical basis of $H_2(X)$, then after a handle slide this canonical basis has changed to $\alpha_1 + \alpha_2, \alpha_2$. The framing on h_1 after the slide is

$$(\alpha_1 + \alpha_2) \cdot (\alpha_1 + \alpha_2) = \alpha_1 \cdot \alpha_1 + \alpha_2 \cdot \alpha_2 + 2\alpha_1 \cdot \alpha_2 = n_1 + n_2 + 2\,\mathrm{lk}(k_1, k_2)$$

In example 5.5 we saw how an unknotted component with framing 0 that links one other strand exact once allows us to separate and cancel those two components from the rest of the link. What really happened is that we were able to separate the 0-framed unknot and the linked component from the rest of the link, and then we could completely unknot that linked component so that we were left with a Hopf link with one component framed by 0 and the other framed by some integer n. By repeatedly sliding handles we arrived at a Hopf link framed by 0 and 0, or 0 and 1 depending on the parity of n. The 4-manifolds corresponding to these framed Hopf links are $S^2 \times S^2$ and $S^2 \tilde{\times} S^2$ respectively. So, example 5.5 has the following two immediate corollaries.

Corollary 6.4. *Let M be a 4-dimensional handlebody and let k_0, k_1 be attaching circles of two 2-handles such that k_1 lies entirely in ∂D^4 and k_0 is a 0-framed meridian of k_1. Finally, let N be the handlebody with all the handles of M except for k_1 and k_2. Then M is diffeomorphic to $N \# S^2 \times S^2$ if the framing coefficient of k_1 is even, and otherwise M is diffeomorphic to $N \# S^2 \tilde{\times} S^2$.*

Corollary 6.5. *If M is a 2-handlebody with odd intersection form, then $M \# S^2 \times S^2$ is diffeomorphic to $M \# S^2 \tilde{\times} S^2$.*

Proof. Since Q_M is odd we must have that any matrix representation of Q_M has an odd number of the diagonal. In particular, if we take the matrix representation with respect to the canonical basis of $H_2(M)$ we see that some 2-handle has odd framing, so let k denote the attaching circle for this handle. Form $M \# S^2 \times S^2$ by adding a Hopf link $\ell_1 \cup \ell_2$, both framed as 0, and slide ℓ_1 over k to obtain ℓ_1'. Then ℓ_1' has odd framing and is linked once by ℓ_2, so we can separate them from the rest of the link, and reduce it until it is a Hopf link framed by 0 and 1, hence we have $M \# S^2 \tilde{\times} S^2$. $\qquad \square$

7 Morse Homology

7.1 The Morse-Smale Condition and Moduli Spaces of Flow Lines

Let $f : X \to \mathbb{R}$ be a Morse function on a closed, Riemannian n-dimensional manifold $(X, \widetilde{g})$ with the Levi-Civita connection $\widetilde{\nabla}$. The metric $\widetilde{g}$ and connection $\widetilde{\nabla}$ will be used for background computations, and is not really significant. The set of critical points of f of index i will be denoted by $\mathrm{crit}_i(f)$, and we set $\mathrm{crit}(f) = \cup_i \mathrm{crit}_i(f)$. Recall that at a critical point $p \in \mathrm{crit}(f)$, the tangent space $T_p X$ splits as a direct sum $E^- T_p X \oplus E^+ T_p X$, where $E^- T_p X$ and $E^+ T_p X$ denote the negative and positive eigenspaces of the Hessian H_f at p and $\dim E^- T_p X = \mu(p)$. A negative gradient flow line is a curve $u : \mathbb{R} \to X$ such that $du/dt = -\nabla_g f \circ u$. Since X is compact, $-\nabla_g f$ generates a 1-parameter group of diffeomorphisms $\varphi_t : X \to X$, i.e φ_t is a diffeomorphism for all $t \in \mathbb{R}$ such that $\varphi_{s+t} = \varphi_s \circ \varphi_t$, $\varphi_0 = \mathrm{id}$ and for fixed point $p \in X$ we have

$$\frac{d\varphi_t(p)}{dt} = -\nabla_g f(\varphi_t(p))$$

For a critical point $p \in \mathrm{crit}(p)$ we define the **ascending** and **descending manifolds** to be

$$\mathscr{A}(p) = \left\{ x \in X : \lim_{t \to \infty} \varphi_t(x) = p \right\}$$

$$\mathscr{D}(p) = \left\{ x \in X : \lim_{t \to -\infty} \varphi_t(x) = p \right\}$$

These sets consist of points that flow to critical points in forward or backward time respectively. The ascending manifold is sometimes called the **stable** manifold and the descending manifold is sometimes called the **unstable** manifold. A study of the dynamics of the negative gradient flow allows one to prove the following important statement.

Proposition 7.1. *If $p \in \mathrm{crit}_i(f)$, then $\mathscr{D}(p)$ and $\mathscr{A}(p)$ are smoothly embedded discs of dimension i and $n-i$ respectively. Further, $T_p\mathscr{D}(p) = E^- T_p X$ and $T_p\mathscr{A}(p) = E^+ T_p X$.*

We will not prove this because it is not useful in our route to defining Morse homology. A pair (f, g) is said to be **Morse-Smale** if f is a Morse function such that for every pair of critical points $p, q \in \mathrm{crit}(f)$ the ascending and descending manifolds intersect transversely $\mathscr{D}(p) \pitchfork \mathscr{A}(q)$. Since codimension is additive under transverse intersections we have $\dim \mathscr{D}(p) \cap \mathscr{A}(q) = \mu(p) - \mu(q)$. We do not yet know such pairs even exist (they do, in fact, in abundance), but we immediately get the following lemma.

Lemma 7.1. *If (f, g) is a Morse-Smale pair on X, then a non-constant flow line of $-\nabla_g f$ from p to q has $\mu(p) > \mu(q)$.*

Proof. Since (f, g) is Morse-Smale we have $\mathscr{D}(p) \cap \mathscr{A}(q)$ is a manifold of dimension $\mu(p) - \mu(q)$. If $\mu(p) < \mu(q)$, then this manifold is empty, hence there are no flow lines from p to q. If $p = q$, then this is a 0-dimensional manifold whose points correspond to constant flow lines. Therefore if there is a non-constant flow line from p to q we must have $\mu(p) > \mu(q)$. $\square$

Example 7.1. If we stand the torus T^2 vertically with the induced metric from its embedding in $\mathbb{R}^2$, and let $f : T^2 \to \mathbb{R}$ be the height function, then f does *not* satisfy the Morse-Smale condition. If we let p, q, r, s denote the critical points of f such that $f(p) > f(q) > f(r) > f(s)$, then we can see that $\mathscr{D}(q)$ does not intersect $\mathscr{A}(r)$ transversely. Unfortunately this type of failure of the

Morse-Smale condition is common for "obvious" choices of Morse functions, which tend to have many symmetries. We can perturb f a little so that it satisfies the Morse-Smale condition (for example, slightly tilt T^2 off its vertical axis), but sometimes it is better to consider a more relaxed requirement than Morse-Smale, such as the Morse-Bott condition.

Example 7.2. Consider the flat torus $T^2 = \mathbb{R}^2/\mathbb{Z}^2$ with the smooth function $f : T^2 \to \mathbb{R}$ defined by $f(x, y) = \cos(2\pi x) + \cos(2\pi y)$. This function is clearly well defined, and the critical points are at the equivalence classes of $(0, 0), (1/2, 0), (0, 1/2), (1/2, 1/2) \in \mathbb{R}^2$. It is easy to compute the Hessian at these critical points and see that they are non-degenerate, so f is a Morse function. It is also easy to see that all the ascending and descending manifolds intersect transversally, so f satisfies the Morse-Smale condition.

There is another way of discussing the Morse-Smale condition that is better suited for infinite dimensional generalizations of Morse homology. For two critical points $p, q \in \text{crit}(f)$, let $\mathcal{M}^f(p, q)$ be the moduli space of flow lines from p to q, i.e.

$$\mathcal{M}^f(p, q) = \left\{ u : \mathbb{R} \to X : \frac{du}{dt} = -\nabla_g f(u(t)), u(-\infty) = p, u(\infty) = q \right\}$$

We will show how to put a topology on this set momentarily. This set carries a free $\mathbb{R}$ action given by pre-composition with translation: $(s \cdot u)(t) = u(t + s)$. Let $\widehat{\mathcal{M}}^f(p, q)$ denote the quotient of $\mathcal{M}^f(p, q)$ by this action, which we call the moduli space of unparameterized gradient flow lines. There is an obvious isomorphism between $\mathscr{D}(p) \cap \mathscr{A}(q)$ and $\mathcal{M}^f(p, q)$ as sets since any flow line is uniquely determined by its initial condition, so we would like to know that $\mathcal{M}^f(p, q)$ is actually a manifold.

To pursue this we need to develop significant functional analysis machinery. Let $\mathcal{C}_{p,q}$ be the subset of $W^{1,2}_{loc}(\mathbb{R}, X)$ such that if $u \in \mathcal{C}_{p,q}$, then

1. $u(-\infty) = p$ and $u(\infty) = q$

2. There is an $R < 0$ sufficiently negative such that $u(-\infty, R]$ is contained in a coordinate chart around p and $u|_{(-\infty, R]} \in W^{1,2}((-\infty, R], X)$.

3. There is an $R > 0$ sufficiently positive such that $u[R, \infty)$ is contained in a coordinate chart around p and $u|_{[R, \infty)} \in W^{1,2}([R, \infty), X)$.

Then $\mathcal{C}_{p,q}$ is a Banach manifold modeled on $W^{1,2}(\mathbb{R}, \mathbb{R}^n)$. Let $\pi_{\mathcal{E}} : \mathcal{E} \to \mathcal{C}_{p,q}$ be the Banach bundle over $\mathcal{C}_{p,q}$ defined by

$$\mathcal{E} = \bigcup_{u \in \mathcal{C}_{p,q}} \{u\} \times L^2(u^* TX)$$

The pullback bundle $u^* TX$ makes sense because the Sobolev embedding $W^{1,2} \hookrightarrow C^0$ ensures that u is continuous, so $\mathcal{E}$ is a Banach bundle with characteristic fiber $L^2(\mathbb{R}, \mathbb{R}^n)$. For any metric g on X we can define a section $F : \mathcal{C}_{p,q} \to \mathcal{E}$ by

$$F(u) = (u, u' + \nabla_g f \circ u)$$

where u' is the weak derivative of u. Technically F depends on the Morse function f, metric g and the points p, q, but we omit this from the notation. If we let $0_{\mathcal{E}}$ denote the zero section of $\mathcal{E} \to \mathcal{C}_{p,q}$, then we have the following:

Proposition 7.2. $F^{-1}(0_{\mathcal{E}}) = \mathcal{M}^f(p, q)$

Proof. If $u \in F^{-1}(0_{\mathcal{E}})$, then $u' = -\nabla_g f \circ u$. Since $\nabla_g f$ and u are continuous, we have u' is continuous, hence u' is equal to the regular derivative du/dt. Therefore u is C^1 and so is an integral curve from p to q, i.e. $u \in \mathcal{M}^f(p,q)$. On the other hand, if $u \in \mathcal{M}^f(p,q)$, then u converges exponentially to the critical points p and q, so $u \in W^{1,2}(\mathbb{R}, X)$ and hence $u \in F^{-1}(0_{\mathcal{E}})$.

By applying the standard bootstrapping argument we see that u is actually as smooth as g (assuming f is smooth), hence u is actually C^k. $\qquad\square$

In this context we say that (f, g) is a Morse-Smale pair if $F \pitchfork 0_{\mathcal{E}}$, which of course implies that $\mathcal{M}^f(p,q)$ is a submanifold of the Banach manifold $\mathcal{C}_{p,q}$. If we could show that dF is Fredholm at the points in $F^{-1}(0_{\mathcal{E}})$, then we would have that $F^{-1}(0_{\mathcal{E}}) = \mathcal{M}^f(p,q)$ is a finite dimensional submanifold of $\mathcal{C}_{p,q}$. If we work a little harder we can actually generalize the above constructions a little to show something stronger. Let $\mathcal{A}^k$ denote the Banach manifold consisting of C^k-metrics on X. Note that this is just a convex subset of the Banach space (i.e. linear space) of C^k sections of $S^2 T^* X \to X$. Let $\pi_{\widetilde{\mathcal{E}}} : \widetilde{\mathcal{E}} \to \mathcal{A}^k \times \mathcal{C}_{p,q}$ be the Banach bundle that is the pullback of $\mathcal{E} \to \mathcal{C}_{p,q}$ along the projection $\mathcal{A}^k \times \mathcal{C}_{p,q} \to \mathcal{C}_{p,q}$. In particular, $\widetilde{\mathcal{E}}$ can be written as

$$\widetilde{\mathcal{E}} = \bigcup_{(g,u) \in \mathcal{A}^k \times \mathcal{C}_{p,q}} \{(g,u)\} \times L^2(u^* TX)$$

So, the characteristic fiber of $\widetilde{\mathcal{E}}$ is $L^2(\mathbb{R}, \mathbb{R}^n)$. We define a section $\widetilde{F} : \mathcal{A}^k \times \mathcal{C}_{p,q} \to \widetilde{\mathcal{E}}$ of this bundle by $\widetilde{F}(g, u) = (g, u, u' + \nabla_g f \circ u)$. This section depends only on f, but we do not include this in the notation. Then the most important proposition we want to prove is the following.

Proposition 7.3. *The section $\widetilde{F}$ is transverse to the zero section: $\widetilde{F} \pitchfork 0_{\widetilde{\mathcal{E}}}$. Further, if $\pi : \mathcal{A}^k \times \mathcal{C}_{p,q} \to \mathcal{A}^k$ is the projection onto the second factor, then the restriction of π to $\pi : \widetilde{F}^{-1}(0_{\widetilde{\mathcal{E}}}) \to \mathcal{A}^k$ is Fredholm of index $\mu(p) - \mu(q)$, i.e. $d\pi_{(g,u)}$ is Fredholm of index $\mu(p) - \mu(q)$ at each $(g,u) \in \widetilde{F}^{-1}(0_{\widetilde{\mathcal{E}}})$.*

The immediate corollary of the this proposition is that a generic $g \in \mathcal{A}^k$ is a regular value of the restricted map π, hence $\pi^{-1}(g)$ is a finite dimensional submanifold of dimension $\mu(p) - \mu(q)$. Clearly $\pi^{-1}(g) = \mathcal{M}^f(p,q)$, where this moduli space is taken with respect to g, therefore we have the following.

Corollary 7.1. *The moduli space of flow lines $\mathcal{M}^f(p,q)$ from p to q for a Morse-Smale pair is a $(\mu(p) - \mu(q))$-dimensional submanifold of $\mathcal{C}_{p,q}$. The moduli space of unparameterized flow lines $\widehat{\mathcal{M}}^f(p,q)$ is a $(\mu(p) - \mu(q) - 1)$-dimensional manifold.*

Proposition 7.3 will actually follow reasonably easily from the fact about the vertical differential DF_u of the section constructed for the first Banach bundle $\mathcal{E} \to \mathcal{C}_{p,q}$ is Fredholm.

Proposition 7.4. *For each $u \in F^{-1}(0_{\mathcal{E}})$, the vertical differential $DF_u : T_u \mathcal{C}_{p,q} \to \mathcal{E}_u$ is Fredholm of index $\mu(p) - \mu(q)$.*

So, now we will work to prove proposition 7.4. The first thing we can do is introduce coordinates on our manifolds and bundles so that we can write the vertical differential in a nicer way. To put coordinates on $\mathcal{C}_{p,q}$ fix $u \in \mathcal{C}_{p,q}$ and a small open set U around u. Let $N(TX)$ be a small enough convex, tubular neighborhood of the zero section of TX so that the exponential map $\exp : N(TX) \to X$ is defined, where this is the exponential map with respect to the background metric $\widetilde{g}$ and Levi-Civita connection $\widetilde{\nabla}$ floating around. This exponential map induces coordinates on $\mathcal{C}_{p,q}$ in the following way. Let $u \in \mathcal{C}_{p,q}$ and take another curve $v \in \mathcal{C}_{p,q}$ close to u. Then there is a unique
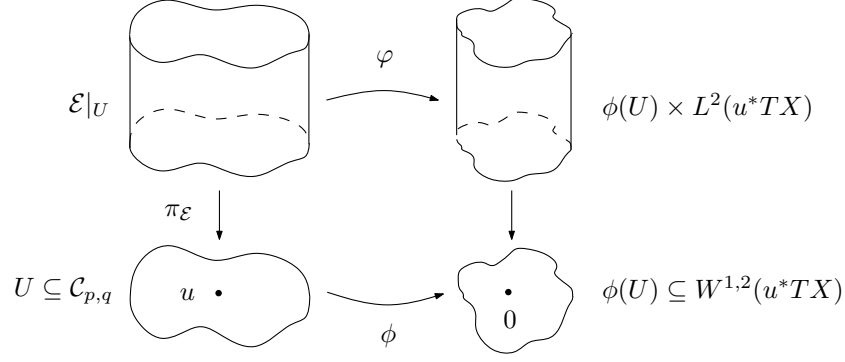
Figure 7.1: Trivializing $\mathcal{E} \to \mathcal{C}_{p,q}$

section $\phi(v) \in W^{1,2}(u^*TX)$ (i.e. a $W^{1,2}$ vector field along u) such that $v(t) = \exp(u(t), \phi(v)(t))$. So, for a small neighborhood U around u we have a chart $\phi : U \to W^{1,2}(u^*TX) = T_u\mathcal{C}_{p,q}$ with $\phi(u) = 0$.

Next we trivialize $\pi_\mathcal{E} : \mathcal{E} \to \mathcal{C}_{p,q}$ over U. For $v \in U$ we can define the bundle map $P_u^v : u^*TX \to v^*TX$ that parallel transports a vector V_t in the fiber of u^*TX over t along the geodesic $s \mapsto (u(t), s\phi(v)(t))$ to a vector in the fiber of v^*TX over t. This map allows us to identify a section in $L^2(u^*TX)$ with a section in $L^2(v^*TX)$, which are precisely the fibers of $\mathcal{E}$ over u and v. Now that we have coordinates on (ϕ, U) around u in $\mathcal{C}_{p,q}$ and a trivialization of $\mathcal{E}|_U$ (and hence coordinates on $\mathcal{E}|_U$), we can write the coordinate representation $F : \phi(U) \to \phi(U) \times L^2(u^*TX)$ of the section F as

$$F(\xi) = \left(\xi, (P_u^v)^{-1}\left(\exp(u, \xi)' + \nabla_g f \circ \exp(u, \xi)\right)\right)$$

Computing the vertical differential $DF_u : \phi(U) \to L^2(u^*TX)$ in these coordinates gives

$$\begin{aligned}
DF_u(\xi)(t) &= \left.\frac{d}{ds}\right|_{s=0} F(s\xi(t)) \\
&= \left.\frac{d}{ds}\right|_{s=0} \left((P_u^v)^{-1}\left(\exp(u(t), s\xi(t))' + \nabla_g f(\exp(u(t), s\xi(t)))\right)\right) \\
&= \widetilde{\nabla}_{\xi(t)}\left(u'(t) + \nabla_g f(u(t))\right) \\
&= \widetilde{\nabla}_{\xi(t)}u'(t) + \widetilde{\nabla}_{\xi(t)}\left(\nabla_g f(u(t))\right)
\end{aligned}$$

where the second from last line follows from the fact that any connection can be recovered from parallel translations along its geodesics. Also, since the connection is torsion free we have $\widetilde{\nabla}_{\xi(t)}u'(t) = \widetilde{\nabla}_{u'(t)}\xi(t)$, so

$$DF_u(\xi) = \widetilde{\nabla}_{u'}\xi + \widetilde{\nabla}_\xi\left(\nabla_g f(u)\right)$$

If we trivialize u^*TX using parallel translation with respect to $\widetilde{\nabla}$, then vertical differential $DF_u : W^{1,2}(\mathbb{R}, \mathbb{R}^n) \to L^2(\mathbb{R}, \mathbb{R}^n)$ can be written as

$$(DF_u\xi)(t) = \left.\frac{d}{ds}\right|_{s=t} \xi(s) - A_t\xi(t)$$

where $A \in C^k(\mathbb{R}, \text{End}(\mathbb{R}^n))$ is defined by $A_t(V) = -\widetilde{\nabla}_V(\nabla_g f(u(t)))$. By (4.4) we have

$$A_{-\infty} = -H_f(p) \qquad A_\infty = -H_f(q)$$

69

With this set up we will see that proposition 7.4 follows from a more general proposition concerning spectral flows (see [4]). Before stating this proposition we recall the definition of spectral flow, at least for the special case we need. Let $A \in C_b(\mathbb{R}, \mathrm{End}(\mathbb{R}^n))$ be a bounded path of operators on $\mathbb{R}^n$ such that $A_{\pm\infty}$ are symmetric and do not have zero as an eigenvalue. Then we define the spectral flow of the path A to be the number of eigenvalues of A_s that change from negative to positive as s ranges from $-\infty$ to ∞. We denote this integer by $\mathrm{SF}(A)$, and clearly in this special case we have

$$\mathrm{SF}(A) = \dim E^+(A_\infty) - \dim E^+(A_{-\infty})$$

Proposition 7.5. *Let $A \in C_b(\mathbb{R}, \mathrm{End}(\mathbb{R}^n))$ such that $A_{-\infty}$ and A_∞ are symmetric and do not have zero as an eigenvalue. Then the operator $D_A : W^{1,2}(\mathbb{R}, \mathbb{R}^n) \to L^2(\mathbb{R}, \mathbb{R}^n)$ given by*

$$D_A \xi = \frac{d}{dt}\xi - A\xi$$

is Fredholm with

$$\mathrm{ind}\, D_A = -\,\mathrm{SF}(A)$$

Proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Applying proposition 7.5 to the vertical differential $DF_u = d/dt - A_s$ we see that DF_u is Fredholm of index

$$\begin{aligned}
\mathrm{ind}\, DF_u &= -\,\mathrm{SF}(A) \\
&= -(\dim E^+(-H_f(q)) - \dim E^+(-H_f(p))) \\
&= \dim E^-(H_f(p)) - \dim E^-(H_f(q)) \\
&= \mu(p) - \mu(q)
\end{aligned}$$

and this finishes the proof proposition 7.4. Let us see how to use proposition 7.4 to prove proposition 7.3.

Proof of proposition 7.3. Our derivation of the local expression of the vertical differential DF_u at $u \in F^{-1}(0_{\mathcal{E}})$ can be slightly adapted to give the following local expression for the vertical differential $D\widetilde{F}_{(g,u)} : T_{(g,u)}(\mathcal{A}^k \times \mathcal{C}_{p,q}) \to T_{(g,u,0)}\widetilde{\mathcal{E}}$ at a point $(g,u) \in \widetilde{F}^{-1}(0_{\widetilde{\mathcal{E}}})$:

$$D\widetilde{F}_u(\eta, \xi) = DF_u(\xi) + \partial_\eta(\nabla_g f(u))$$

where ∂_η denotes the derivative with respect to the η variable (remember that $\mathcal{A}^k$ is just a convex, open subset of a linear space).

Let $(g,u) \in \widetilde{F}^{-1}(0_{\widetilde{\mathcal{E}}})$ and consider the differential $d\pi_{(g,u)} : T_{(g,u)}\widetilde{F}^{-1}(0_{\widetilde{\mathcal{E}}}) \to T_g\mathcal{A}^k$ of the restriction of π to $\widetilde{F}^{-1}(0_{\widetilde{\mathcal{E}}})$. We claim that the kernel of $d\pi_{(g,u)}$ is isomorphic to the kernel of DF_u. We have that

$$T_{(g,u)}\widetilde{F}^{-1}(0_{\widetilde{\mathcal{E}}}) = (D\widetilde{F}_{(g,u)})^{-1}(T_{(g,u)}0_{\widetilde{\mathcal{E}}})$$

which can be written as

$$T_{(g,u)}\widetilde{F}^{-1}(0_{\widetilde{\mathcal{E}}}) = \left\{ (\eta, \xi) \in T_g\mathcal{A}^k \oplus T_u\mathcal{C}_{p,q} : DF_u(\xi) + \partial_\eta(\nabla_g f(u)) = 0 \right\} \qquad (7.1)$$

using our coordinates. We clearly have $d\pi_{(g,u)}(\eta, \xi) = \eta$, hence

$$\ker d\pi_{(g,u)} = \left\{ (0, \xi) \in T_g\mathcal{A}^k \oplus T_u\mathcal{C}_{p,q} : DF_u(\xi) = 0 \right\} = \ker DF_u$$

70

Next we claim that the cokernel of $d\pi_{(g,u)}$ is isomorphic to the cokernel of DF_u. (**todo:** do this).

Finally, we show that the image of $d\pi_{(g,u)}$ is closed. Using our coordinates and (7.1), the image of $d\pi_{(g,u)}$ can be written as

$$d\pi_{(g,u)}\left(T_{(g,u)}\widetilde{F}^{-1}(0_{\widetilde{\mathcal{E}}})\right) = \left\{\eta \in T_g\mathcal{A}^k : DF_u(\xi) + \partial_\eta(\nabla_g f(u)) = 0 \text{ for some } \xi \in T_u\mathcal{C}_{p,q}\right\}$$
$$= \left\{\eta \in T_g\mathcal{A}^k : \partial_\eta(\nabla_g f(u)) \in \operatorname{im} DF_u\right\}$$

So, we see that $\operatorname{im} d\pi_{(g,u)}$ is the pre-image of $\operatorname{im} DF_u$ under the map $\eta \mapsto \partial_\eta(\nabla_g f(u))$. This map is continuous, and DF_u is closed since it is Fredholm, therefore $\operatorname{im} d\pi_{(g,u)}$ is closed. $\qquad\square$

Proposition 7.6. *A generic C^k function $f : X \to \mathbb{R}$ ($k \geq 2$) is Morse.*

Proof. We clearly want to use the Sard-Smale theorem somehow, and in fact we will use theorem E.3. In order to use this we need to formulate a Banach bundle such that the zero section corresponds to what we want. In particular, let $\pi : \mathcal{F} \to C^k(X) \times X$ be the Banach bundles that is the pullback of the cotangent bundle $T^*X \to X$ along the projection $C^k(X) \times X \to X$ so that the fiber of $\mathcal{F}$ over (f,p) is just T_x^*X. Consider the section $s : C^k(X) \times X \to \mathcal{F}$ given by $s(f,p) = df_p \in T_p^*X$. Let $(f,x) \in \pi^{-1}(0_\mathcal{F})$ and let us compute the vertical differential $Ds_{(f,x)} : T_{(f,x)}(C^k(X) \times X) \to T_x^*X$ at (f,x). Note that since $C^k(X)$ is a linear space we have $T_f C^k(X)$ is naturally isomorphic to $C^k(X)$, so we describe how $Ds_{(f,x)}$ acts on a pair $(h,V) \in C^k(X) \times T_xX$.

$$Ds_{(f,p)}(h,V) = dh(X) + \nabla_V(df)$$

    UNFINISHED $\qquad\square$

7.2 The Compactness and Gluing Theorems

The space of unparameterized flow lines $\widehat{\mathcal{M}}^f(p,q)$ is not always compact; for example, in figure 7.2 we see a sequence of flow lines (the thin curves) on the deformed sphere that converge to a "broken" curve (the bolded curve). A **broken flow line** from p to q is a sequence of critical points $r_1 = p, r_2, \ldots, r_{k-1}, r_k = q \in \operatorname{crit}(f)$ and curves $\gamma_1, \ldots, \gamma_{k-1}$ in X such that γ_i is a flow line from r_i to r_{i+1}. However, there is a canonical compactification of $\widehat{\mathcal{M}}^f(p,q)$ to a manifold with corners by adding "broken" flow lines. There are two parts to this compactification process. The first part (called the compactness result) consists of identifying how a sequence can fail to converge in $\widehat{\mathcal{M}}^f(p,q)$, trying to get a reasonable object out of a diverging sequence, and then adding these points to $\widehat{\mathcal{M}}^f(p,q)$ to get a compact space. The second part (called the gluing result) consists of showing that a broken flow line can be slightly perturbed to an honest flow line so that our compactification is a manifold with corners.

Recall that a second countable, Hausdorff space X is said to be a **smooth n-manifold with corners** if every point $x \in X$ has an open neighborhood U_α and homeomorphism $\varphi_\alpha : U_\alpha \to \mathbb{R}^{n-k} \times [0,\infty)^k$ (for some $0 \leq k \leq n$) such that the transition functions $\varphi_\beta \circ \varphi_\alpha^{-1}$ are smooth. For $0 \leq k \leq n$, we define the **codimension k stratum** of X to be the set X_k of points $x \in X$ with a chart $\varphi : U \to \mathbb{R}^{n-k} \times [0,\infty)^k$ such that at least one of the last k coordinates of $\varphi(x)$ is zero. Note that X_0 is just the interior of X, and if $X_2 = X_3 = \cdots = X_n = \emptyset$, then X is just a manifold with boundary and $X_1 = \partial X$. So we will use the notation ∂X to denote X_1 even when the higher strata are non-empty.

Now we look at what it means for a sequence in $\widehat{\mathcal{M}}^f(p,q)$ to converge, and how this convergence can fail. In $\mathcal{M}^f(p,q)$, a sequence of elements converges if it converges in the subspace topology

71
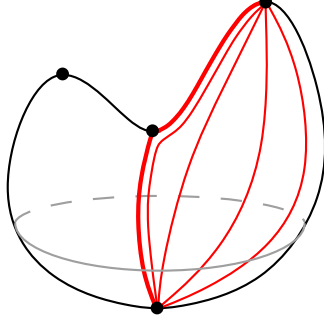
Figure 7.2: Flow lines converging to a broken flow line

induced by the Banach manifold $\mathcal{C}_{p,q}$. Since modding out by the $\mathbb{R}$ action identifies flow lines that differ by only a translation we have that a sequence $(\widehat{u}_n)$ of unparameterized flow lines in $\widehat{\mathcal{M}}^f(p,q)$ converge to a flow line $\widehat{u}$ if for any sequence of lifts (u_n) of $(\widehat{u}_n)$ and lift u of $\widehat{u}$ there is a sequence (s_n) of real numbers such that $s_n \cdot u_n$ converges to u in $\mathcal{M}^f(p,q)$.

(**todo:** finish compactifications and gluing)
‒ ‒ ‒ ‒ ‒ ‒ ‒ ‒ ‒

The BIG theorem:

Theorem 7.1. *If X is a closed n-manifold and (f,g) is a Morse-Smale pair, then for any two critical points p, q, the manifold $\widehat{\mathcal{M}}^f(p,q)$ has a natural compactification to a smooth manifold with corners $\overline{\widehat{\mathcal{M}}^f(p,q)}$ whose codimension k stratum is*

$$\overline{\widehat{\mathcal{M}}^f(p,q)}_k = \bigcup_{\substack{r_1, \ldots, r_k \,\in\, \mathrm{crit}(f) \\ p, r_1, \ldots, r_k, q \ \text{distinct}}} \widehat{\mathcal{M}}^f(p,r_1) \times \widehat{\mathcal{M}}^f(r_1, r_2) \times \cdots \widehat{\mathcal{M}}^f(r_{k-1}, r_k) \times \widehat{\mathcal{M}}^f(r_k, q)$$

In particular, if $k = 1$, then as oriented manifolds we have

$$\partial\overline{\widehat{\mathcal{M}}^f(p,q)} = \bigcup_{\substack{r \,\in\, \mathrm{crit}(f) \\ p, r, q \ \text{distinct}}} (-1)^{\mu(p)+\mu(r)+1} \widehat{\mathcal{M}}^f(p,r) \times \widehat{\mathcal{M}}^f(r,q)$$

The two most important special cases of this theorem are when $\mu(q) = \mu(p) - 1$, in which case $\mathcal{M}^f(p,q)$ is a compact 0-manifold (hence a finite collection of points), and when $\mu(q) = \mu(p) - 2$, in which case $\widehat{\mathcal{M}}^f(p,q)$ has a compactification to a 1-manifold with boundary $\overline{\widehat{\mathcal{M}}^f(p,q)}$ such that $\partial\overline{\widehat{\mathcal{M}}^f(p,q)}$ consists of just once broken flow lines:

$$\partial\overline{\widehat{\mathcal{M}}^f(p,q)} = \bigcup_{r \in \mathrm{crit}_{\mu(p)-1}(f)} \widehat{\mathcal{M}}^f(p,r) \times \widehat{\mathcal{M}}^f(r,q)$$

72

7.3 The Morse Complex

We can now define the **Morse complex**[1] associated to a Morse-Smale pair (f, g). Let $CM_i(f, g)$ be the $\mathbb{Z}/2$ vector space generated by the critical points of index i

$$CM_i(f, g) := \bigoplus_{p \in \mathrm{crit}_i(f)} \mathbb{Z}/2 \langle p \rangle$$

We can define this complex with $\mathbb{Z}$ coefficients if we work harder and define orientations on the moduli spaces $\mathcal{M}^f(p, q)$. It is simpler to work with $\mathbb{Z}/2$ coefficients for now, and we will discuss orientation issues later. The differential $\partial : CM_i(X) \to CM_{i-1}(X)$ counts the number of flow lines between critical points. Specifically, if $p \in \mathrm{crit}_i(f)$, then

$$\partial(p) = \sum_{q \in \mathrm{crit}_{i-1}(f)} \#\widehat{\mathcal{M}}^f(p, q) \cdot q$$

and extend linearly to $CM_i(f)$. The count $\#\widehat{\mathcal{M}}^f(p, q)$ is of course done modulo 2 (at least until we can define orientations).

Proposition 7.7. $\partial \circ \partial = 0$

Proof. First recall that the number of points in the boundary of a 1-manifold is zero when counted modulo 2. For $p \in \mathrm{crit}_i(f)$ we can easily compute

$$\begin{aligned}
\partial \circ \partial(p) &= \partial \left(\sum_{r \in \mathrm{crit}_{i-1}(f)} \#\mathcal{M}^f(p, r) \cdot r \right) \\
&= \sum_{r \in \mathrm{crit}_{i-1}(f)} \#\mathcal{M}^f(p, r) \cdot \partial(r) \\
&= \sum_{q \in \mathrm{crit}_{i-2}(f)} \left(\sum_{r \in \mathrm{crit}_{i-1}(f)} \#\mathcal{M}^f(p, r) \cdot \#\mathcal{M}^f(r, q) \right) \cdot q
\end{aligned}$$

So, the coefficient of $q \in \mathrm{crit}_{i-2}(f)$ in $\partial \circ \partial(p)$ is precisely $\#\partial\overline{\mathcal{M}^f(p, q)}$, which is zero since it is a compact 1-dimensional manifold. $\square$

We define the **Morse homology** $HM_*(f, g)$ of the pair (f, g) to be the homology of the chain complex $(CM_*(f, g), \partial)$.

7.4 A Natural Isomorphism between Morse and Singular Homology

We would like to show that the homology groups do not depend on our choice of Morse function or metric. There are many ways of doing this, but the first way we will consider is to show that $HM_*(f, g)$ is naturally isomorphic to some other known homology theory, such as singular homology, which of course implies that $HM_*(f, g)$ does not depend on f or g.

The homology theory we will show is naturally isomorphic to HM_* is defined as follows. Recall that an *i-dimensional current* on X is a functional on the space of differential i-forms $\Omega^i(X)$.

---

[1]Some might call this construction the Morse-Smale-Witten complex due to its complicated history. See [1] for the interesting history of Morse theory and Morse homology

The topology on $\Omega^i(X)$ is defined by saying that a sequence of differential forms $\{\omega_n\}$ converges to ω if and only if for every compact set $K \subseteq X$ all partial derivatives of ω_n (in any coordinate system) converges uniformly to the partial derivatives of ω. A singular simplex $\sigma : \Delta^i \to X$ determines an i-dimensional current $[\sigma]$ by

$$[\sigma](\omega) = \int_{\Delta^i} \sigma^* \omega$$

We define the chain groups $C_i^c(X)$ of our homology theory to be the free abelian group generated by $[\sigma]$, where $\sigma : \Delta^i \to X$ is a singular i-simplex that is transverse to $\mathscr{A}(p)$ for all $p \in \mathrm{crit}(f)$. The differential $\partial : C_i^c(X) \to C_{x-1}^c(X)$ is defined by

$$\partial[\sigma](\omega) = \int_{\Delta^i} \sigma^* d\omega$$

where $\sigma \in C_i(X)$ and $\omega \in \Omega^{i-1}(X)$. We clearly have $\partial^2 = 0$ since $d^2 = 0$. The homology of this chain complex will be denoted by $H_i^c(X)$, and it is naturally isomorphic to singular homology.

There is a natural compactification of $\mathscr{D}(p)$, for $p \in \mathrm{crit}(f)$, that is similar to theorem 7.1 by adding broken flow lines.

Proposition 7.8. *There is a natural compactification of $\mathscr{D}(p)$ to a smooth manifold with corners $\overline{\mathscr{D}(p)}$, whose codimension k stratum is*

$$\overline{\mathscr{D}(p)}_k = \bigcup_{\substack{q_1, \ldots, q_k \in \mathrm{crit}(f) \\ p, q_1, \ldots, q_k \ \mathit{distinct}}} \widehat{\mathcal{M}}^f(p, q_1) \times \widehat{\mathcal{M}}^f(q_1, q_2) \times \widehat{\mathcal{M}}^f(q_{k-1}, q_k) \times \mathscr{D}(q_k)$$

In particular, for $k = 1$ we have

$$\partial \overline{\mathscr{D}(p)} = \bigcup_{p \neq q \in \mathrm{crit}(f)} (-1)^{\mu(p) + \mu(q) + 1} \widehat{\mathcal{M}}^f(p, q) \times \mathscr{D}(q)$$

as oriented manifolds. Further, the projections of $\overline{\mathscr{D}(p)}$ onto the last factor $\mathscr{D}(q_k)$ patch together to form a smooth map $e : \overline{\mathscr{D}(p)} \to X$ whose restriction to $\mathscr{D}(p)$ is the inclusion.

The compactified spaces $\overline{\mathscr{D}(p)}$ are homeomorphic to a closed ball of dimension $\mu(p)$, and so the maps $e : \overline{\mathscr{D}(p)} \to X$ give X a CW-structure with one cell of dimension i for each critical point of index i. A consequence of this is that $\chi(X)$ is equal to $\sum_i (-1)^i c_i$ where $c_i = \# \mathrm{crit}_i(f)$.

We now define maps $D : CM_i(X) \to C_i^c(X)$ and $A : C_i^c(X) \to CM_i(X)$ that are chain homotopy inverses of each other. For $p \in \mathrm{crit}_i(f)$ we have $\overline{\mathscr{D}(p)}$ is a closed disc of dimension i.

UNFINISHED

7.5 Direct Invariance of Morse Homology

Now we will show that $HM_*(f, g)$ does not depend on f and g directly. This line of reasoning is more applicable to Floer homologies than section 7.4 since there might not be another well-known homology theory to compare to. Fix two Morse-Smale pairs (f_0, g_0) and (f_1, g_1) with associated Morse complexes (C_*^0, ∂^0) and (C_*^1, ∂^1). Take any path of pairs $\Gamma = \{(f_t, g_t) : t \in I\}$ from (f_0, g_0) to (f_1, g_1). Two things to note: the pairs (f_t, g_t) do not have to be Morse-Smale for all t (in fact, it may be impossible), and such a path is always possible as all functions $M \to \mathbb{R}$ are homotopic and any two metrics can be connected by a straight line of metrics. We will show how such a path (with some simple requirements) induces a chain map $\Phi_\Gamma : C_*^0 \to C_*^1$.

Let V_t denote the negative gradient vector field of f_t with respect to g_t, and define the vector field V_Γ on $I \times X$ by

$$V_\Gamma = (t+1)t(t-1)\frac{\partial}{\partial t} + V_t \tag{7.2}$$

Since the two components of this vector field are gradient like vector fields, we can define critical points, flow lines, ascending and descending manifolds in the regular way. In particular, the first coordinate of V_Γ is just the gradient of the function $\mathbb{R} \to \mathbb{R}$ defined by $t \mapsto \frac{1}{4}(t+1)^2(t-1)^2$. This function only has critical points at $-1, 0, 1$ of index $0, 1, 0$ respectively. Thus, the critical points of index i of V_Γ are precisely

$$\mathrm{crit}_i(V_\Gamma) = \{0\} \times \mathrm{crit}_{i-1}(f_0) \cup \{1\} \times \mathrm{crit}_i(f_1) \tag{7.3}$$

We will say that the path Γ is **admissible** if the ascending and descending manifolds of the critical points of V_Γ intersect transversely. Just as almost all pairs (f, g) are Morse-Smale we also have that almost all paths of pairs are admissible.

For critical points P, Q of V_Γ we define $\mathcal{M}^\Gamma(P, Q)$ to be the moduli space of flow lines of V_Γ from P to Q, and $\widehat{\mathcal{M}}^\Gamma(P, Q)$ the moduli space of unparameterized flow lines. If Γ is admissible, then $\mathcal{M}^\Gamma(P, Q)$ is a $(\mu(P) - \mu(Q))$-dimensional manifold, and so $\widehat{\mathcal{M}}^\Gamma(P, Q)$ is a $(\mu(P) - \mu(Q) - 1)$-dimensional manifold. The spaces $\widehat{\mathcal{M}}^\Gamma(P, Q)$ are without boundary, but are not generally compact. We can compactify them via broken lines as before. We state this result only for the two special cases we need. If $\mu(P) = \mu(Q) + 1$, then $\widehat{\mathcal{M}}^\Gamma(P, Q)$ is already compact, and since it is 0-dimensional we have that it is just a finite collection of points. If $\mu(p) = i = \mu(q) + 1$, then $\mu((0, p)) = i + 1$ and $\mu((1, q)) = i - 1$, then $\widehat{\mathcal{M}}^\Gamma((0, p), (1, q))$ is a 1-dimensional manifold with a compactification $\overline{\widehat{\mathcal{M}}^\Gamma((0, p), (1, q))}$ by adding broken flow lines that pass through one intermediate critical points of index i. These points correspond to eq. (7.3), and so we have

$$\partial\overline{\widehat{\mathcal{M}}^\Gamma((0, p), (1, q))} = \bigcup_{r \in \mathrm{crit}_{i-1}(f_0)} \widehat{\mathcal{M}}^\Gamma((0, p), (0, r)) \times \widehat{\mathcal{M}}^\Gamma((0, r), (1, q))$$

$$= \bigcup_{r' \in \mathrm{crit}_i(f_1)} \widehat{\mathcal{M}}^\Gamma((0, p), (1, r')) \times \widehat{\mathcal{M}}^\Gamma((1, r'), (1, q))$$

Note that a flow line between two points in the $\{0\} \times X$ slice must stay in that slice for all time, and similarly for the $\{1\} \times X$ slice. The flow lines in the these slices correspond to the flows of f_0 and f_1 in X, so we have

$$\widehat{\mathcal{M}}^\Gamma((0, p), (0, r)) = \widehat{\mathcal{M}}^{f_0}(p, r)$$

$$\widehat{\mathcal{M}}^\Gamma((1, r'), (1, q)) = \widehat{\mathcal{M}}^{f_1}(r', q)$$

For $p \in \mathrm{crit}_i(f_0)$ and $q \in \mathrm{crit}_i(f_1)$ we have $\mu((0, p)) = i + 1$ and $\mu((0, q)) = i$, hence $\widehat{\mathcal{M}}^\Gamma((0, p), (1, q))$ is a compact 0-dimensional manifold. We now define $\Phi_\Gamma : C_i^0 \to C_i^1$ on $p \in \mathrm{crit}_i(f_0)$ by

$$\Phi_\Gamma(p) = \sum_{q \in \mathrm{crit}_i(f_1)} \#\widehat{\mathcal{M}}^\Gamma((0, p), (1, q)) \cdot q$$

and extend this linearly to all of C_*^0 (where $\widehat{\mathcal{M}}^\Gamma((0, p), (1, q))$ is counted modulo 2).
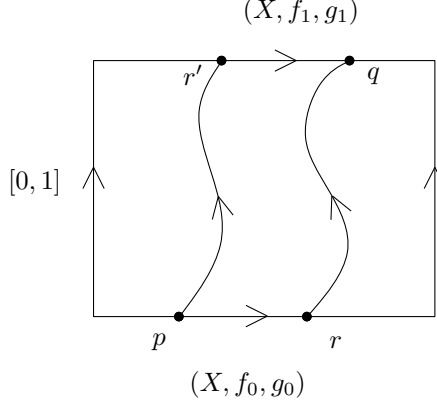
Figure 7.3: Counting flow lines from $(0, p)$ to $(1, q)$

Proposition 7.9. Φ_Γ *is a map of chain complexes, i.e. the following diagram commutes*

$$
\begin{array}{ccc}
C_i^0 & \xrightarrow{\partial^0} & C_{i-1}^0 \\
{\scriptstyle\Phi_\Gamma}\downarrow & & \downarrow{\scriptstyle\Phi_\Gamma} \\
C_i^1 & \xrightarrow{\partial^1} & C_{i-1}^1
\end{array}
$$

Proof. We will use the idea from the proof of proposition 7.7 by showing the coefficient of $q \in \text{crit}_i(f_1)$ in $(\Phi_\Gamma \circ \partial^0 - \partial^1 \circ \Phi_\Gamma)(p)$ is equal to a count of the number of boundary components of a compact 1-manifold, and hence is zero (modulo 2).

For $p \in \text{crit}_i(f_0)$ we have

$$
\begin{aligned}
\Phi_\Gamma \circ \partial^0(p) &= \sum_{r \in \text{crit}_{i-1}(f_0)} \sum_{q \in \text{crit}_{i-1}(f_1)} \#\widehat{\mathcal{M}}^{f_0}(p, r) \cdot \#\widehat{\mathcal{M}}^\Gamma((0, r), (1, q)) \cdot q \\
\partial^1 \circ \Phi_\Gamma(p) &= \sum_{r' \in \text{crit}_i(f_1)} \sum_{q \in \text{crit}_{i-1}(f_1)} \#\widehat{\mathcal{M}}^\Gamma((0, p), (1, r')) \cdot \#\widehat{\mathcal{M}}^{f_1}(r', q) \cdot q
\end{aligned}
\tag{7.4}
$$

Now it is clear that the coefficient of $q \in \text{crit}_{i-1}(f_1)$ in $\underline{(\Phi_\Gamma \circ \partial^0 - \partial^1 \circ \Phi_\Gamma)(p)}$ is precisely the number of boundary components in the compact 1-manifold $\widehat{\mathcal{M}}^\Gamma((0, p), (1, q))$, hence zero. Therefore Φ_Γ is a chain map.

$\square$

We now have that admissible paths Γ induce maps on Morse homology $\Phi_{\Gamma *} : HM_*(f_0, g_0) \to HM_*(f_1, g_1)$. If we take Γ to be the constant path at a Morse-Smale pair (f, g), then we expect the induced map Φ_Γ to be the identity. This is indeed the case, as is shown in the following proposition.

Proposition 7.10. *Let Γ be the constant path at the Morse-Smale pair (f, g). Then Γ is admissible and $\Phi_\Gamma = \text{id}$*

Proof. Let V be the vector field on $[0, 1] \times X$ induced by the constant path Γ, i.e. $V = (t + 1)t(t - 1)\frac{\partial}{\partial t} - \nabla_g f$. Since the vector field $(t + 1)t(t - 1)\frac{\partial}{\partial t}$ on $[0, 1]$ is directed from 0 to 1, we have that the descending manifold of a critical point $(0, p)$ of V is $\mathscr{D}(0, p) = [0, 1) \times \mathscr{D}(p)$ and the ascending manifold of a critical point $(1, q)$ is $\mathscr{A}(1, q) = (0, 1] \times \mathscr{A}(q)$. These submanifolds clearly

76

intersect transversally. On the other hand, the ascending manifold of $(0, p)$ is just $\{0\} \times \mathscr{A}(p)$ and the descending manifold of $(1, q)$ is just $\{1\} \times \mathscr{D}(q)$, which do not intersect at all. Therefore V is admissible.

For $p, q \in \mathrm{crit}_i(f)$, a flow from $(0, p)$ to $(1, q)$ in $[0, 1] \times X$ projects to a flow from p to q in the $\{0\} \times X$ slice. This implies $p = q$ since there are no flows between critical points of equal index, unless it is the constant flow. Therefore $\Phi_\Gamma(p) = p$. $\qquad \square$

Next, we would like to know that Φ_Γ depends only on the homotopy type of the path Γ. We will use a different model for homotopies than the typical 1-parameter family of maps. Let D be a 2-dimensional manifold with corners such that D has two edges, call them e and e', and two vertices, call them v and w. Such a space is sometimes called a **digon**. If Γ_1, Γ_2 are two paths between pairs (f_0, g_0) and (f_1, g_1), then a homotopy is just a family of pairs $\Psi = \{(f_d, g_d) : d \in D\}$ such that when restricted to e it coincides with Γ_1 and when restricted to e' it coincides with Γ_2. This means that $(f_v, g_v) = (f_0, g_0)$ and $(f_w, g_w) = (f_1, g_1)$.

Put a metric $\widetilde{g}$ on D such that the length of the edges is 1, and let $\widetilde{f} : D \to \mathbb{R}$ be a smooth function with only two critical points, one at v of index 2 and one at w of index 0. Further, assume that the negative gradient vector field $\widetilde{V} = -\nabla_{\widetilde{g}} \widetilde{f}$ with respect to $\widetilde{g}$ is tangent to the edges and is equal to $(t + 1)t(t - 1)$ on the edges. Finally, define the vector field V_Ψ on $D \times X$ by $V_\Psi = \widetilde{V} + V_d$, where V_d is the negative gradient vector field of f_d with respect to g_d. Note that these constructions are a higher dimensional version of what we constructed to define V_Γ. We see that V_Ψ restricted to the slices $e \times X$ and $e' \times X$ coincide with the vector fields V_{Γ_1} and V_{Γ_2}, respectively. We can again define critical points, flow lines, ascending and descending manifolds of V_Ψ. Also, just like last time, we will say that Ψ is admissible if the ascending and descending manifolds of critical points intersect transversely, and almost all families are admissible.

The critical points of index i of V_Ψ are precisely

$$\mathrm{crit}_i(V_\Psi) = \{v\} \times \mathrm{crit}_{i-2}(f_0) + \{w\} \times \mathrm{crit}_i(f_1)$$

For critical points P, Q of V_Ψ let $\mathcal{M}^\Psi(P, Q)$ be the moduli space of flow lines of V_Ψ from P to Q, and $\widehat{\mathcal{M}}^\Psi(P, Q)$ the moduli space of unparameterized flow lines. For any admissible family Ψ we have that $\mathcal{M}^\Psi(P, Q)$ is a $(\mu(P) - \mu(Q))$-dimensional manifold, and $\widehat{\mathcal{M}}^\Psi(P, Q)$ is a $(\mu(P) - \mu(Q) - 1)$-dimensional manifold. Compactifying these spaces in full generality can be quite complicated, so we state the results for the two special cases that we need.

If $\mu(P) = \mu(Q) + 1$, then $\widehat{\mathcal{M}}^\Psi(P, Q)$ is already compact, and since it is 0-dimensional we have that it is just a finite collection of points. If $P = (v, p)$ and $Q = (w, q)$ with $\mu(p) = i = \mu(q)$, then $\mu((v, p)) = i + 2$ and $\mu((w, q)) = i$, hence $\widehat{\mathcal{M}}^\Psi((v, p), (w, q))$ is a 1-dimensional manifold and has a natural compactification to a 1-dimensional manifold $\overline{\widehat{\mathcal{M}}}^\Psi((v, p), (w, q))$ with boundary by adding broken flow lines. The boundary of this space has more points than one might think at first. The broken flow lines that pass through one intermediate critical point of index $i + 1$ are of course included in the boundary, and these intermediate critical points are precisely

$$\mathrm{crit}_{i+1}(V_\Psi) = \{v\} \times \mathrm{crit}_{i-1}(f_0) \cup \{w\} \times \mathrm{crit}_{i+1}(f_1)$$

But, we also have boundary points coming from those flow lines which stay in the $e \times X$ and $e' \times X$ for all time. These flow lines are in correspondence with $\widehat{\mathcal{M}}^{\Gamma_1}((v, p), (w, q))$ and $\widehat{\mathcal{M}}^{\Gamma_2}((v, p), (w, q))$,

hence

$$\partial \overline{\widehat{\mathcal{M}}^\Psi((v,p),(w,q))} = \widehat{\mathcal{M}}^{\Gamma_1}((v,p),(w,q)) \cup \widehat{\mathcal{M}}^{\Gamma_2}((v,p),(w,q))$$
$$\cup \bigcup_{r\in\mathrm{crit}_{i-1}(f_0)} \widehat{\mathcal{M}}^\Psi((v,p),(v,r)) \times \widehat{\mathcal{M}}^\Psi((v,r),(w,q))$$
$$\cup \bigcup_{r'\in\mathrm{crit}_{i+1}(f_1)} \widehat{\mathcal{M}}^\Psi((v,p),(w,r')) \times \widehat{\mathcal{M}}^\Psi((w,r'),(w,q))$$

Note that the flow lines in $D \times X$ that stay in the $v \times X$ slice correspond to the flows in X with respect to (f_0,g_0), and similarly for flow lines in the $w \times X$ slice, so in the above boundary we have

$$\begin{aligned}\widehat{\mathcal{M}}^\Psi((v,p),(v,r)) &= \widehat{\mathcal{M}}^{f_0}(p,r)\\ \widehat{\mathcal{M}}^\Psi((w,r'),(w,q)) &= \widehat{\mathcal{M}}^{f_1}(r',q)\end{aligned}$$

Let us define the map $H_\Psi : CM_i(f_0,g_0) \to CM_{i+1}(f_1,g_1)$ on $p \in \mathrm{crit}_i(f_0)$ by

$$H_\Psi(p) = \sum_{q\in\mathrm{crit}_{i+1}(f_1)} \#\widehat{\mathcal{M}}^\Psi((v,p),(w,q))\cdot q$$

and extend linearly to $CM_i(f_0,g_0)$.

Proposition 7.11. *For any two admissible paths Γ_1,Γ_2 from (f_0,g_0) to (f_1,g_1), the induced maps Φ_{Γ_1} and Φ_{Γ_2} are chain homotopic.*

Proof. Let $\Psi = \{(f_d,g_d) : d \in D\}$ be an admissible homotopy between Γ_1 and Γ_2. We claim that H_Ψ defined above is a chain homotopy; that is, if we set $E = \partial^1 \circ H_\Psi + H_\Psi \circ \partial^0 - \Phi_{\Gamma_1} + \Phi_{\Gamma_2}$ then $E(p) = 0$ for all $p \in \mathrm{crit}(f)$. We show this by realizing the coefficient of $q \in \mathrm{crit}_i(f_1)$ in $E(p)$, where $p \in \mathrm{crit}_i(f_0)$, as the number of boundary components of $\overline{\widehat{\mathcal{M}}^\Psi((v,p),(w,q))}$, which we know to be zero since the space is a compact, 1-dimensional manifold.

We can easily compute the following

$$H_\Psi \circ \partial^0(p) = \sum_{r\in\mathrm{crit}_{i-1}(f_0)}\sum_{q\in\mathrm{crit}_i(f_1)} \#\widehat{\mathcal{M}}^{f_0}(p,r)\cdot\#\widehat{\mathcal{M}}^\Psi((v,r),(w,q))\cdot q$$

$$\partial^1 \circ H_\Psi(p) = \sum_{r'\in\mathrm{crit}_{i+1}(f_1)}\sum_{q\in\mathrm{crit}_i(f_1)} \#\widehat{\mathcal{M}}^\Psi((v,p),(w,r'))\cdot\#\widehat{\mathcal{M}}^{f_1}(r',q)\cdot q$$

Let us fix $p \in \mathrm{crit}_i(f_0)$ and $q \in \mathrm{crit}_i(f_1)$, then $\mu((v,p)) = i+2$ and $\mu((w,q)) = i$, so $\widehat{\mathcal{M}}^\Psi((v,p),(w,q))$ is a 1-dimensional manifold and has its natural compactification described above. We see that by definition the coefficient of q in $\Phi_{\Gamma_1}(p)$ is precisely $\#\widehat{\mathcal{M}}^{\Gamma_1}((v,p),(w,q))$, and similarly for the coefficient of q in $\Phi_{\Gamma_2}(p)$. Now it is clear that the coefficient of q in $E(p)$ is precisely $\#\partial\overline{\widehat{\mathcal{M}}^\Psi((v,p),(w,q))}$, hence zero.

$\square$

Now we have that $\Phi_{\Gamma*}$ depends only on the homotopy class of the path Γ. However, all paths between any two pairs are homotopic (since $C^\infty(M)$ and the space of metrics is contractible), so $\Phi_{\Gamma*}$ does not even depend on Γ. This means that just by virtue of the construction of Morse homology we always get a natural map $\Phi_* : HM_*(f_0,g_0) \to HM_*(f_1,g_1)$ for any Morse-Smale pairs (f_0,g_0) and (f_1,g_1). We want to show that this naturally defined map is an isomorphism, which means that Morse homology depends only on X.
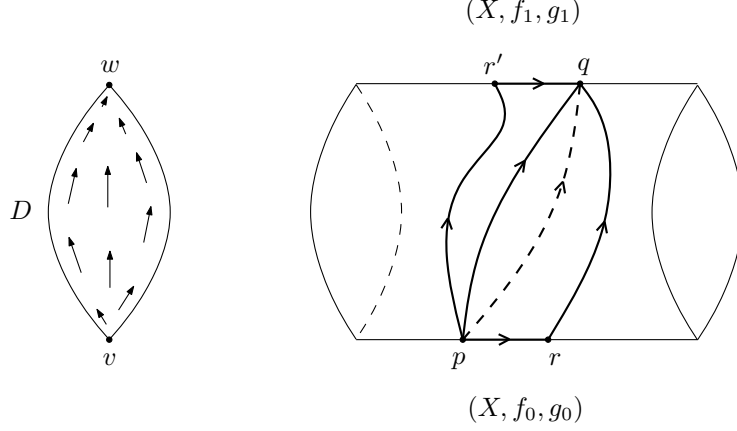
Figure 7.4: Counting flow lines from (v, p) to (w, q) in $D \times X$

In order to show this we need composition of paths $\Gamma_2 * \Gamma_1$ (where Γ_1's ending point is Γ_2's starting point) to correspond to composition of induced maps $\Phi_{\Gamma_2*} \circ \Phi_{\Gamma_1*}$; that is, we need to show that $\Phi_{\Gamma_2*\Gamma_1}$ is chain homotopic to $\Phi_{\Gamma_2*} \circ \Phi_{\Gamma_1*}$. To show this we need to introduce a new apparatus similar to the digon from above. First, slightly homotope $\Gamma_2 * \Gamma_1$ so that it is admissible. Then, let T be a trigon; that is, a 2-manifold with 3 corners, u, v, w, and 3 edges, e_{uv}, e_{vw}, e_{uw}. Let $\Omega = \{(f_d, g_d) : d \in T\}$ be a family of pairs such that $(f_u, g_u) = \Gamma_1(0)$, $(f_v, g_v) = \Gamma_1(1) = \Gamma_2(0)$, $(f_w, g_w) = \Gamma_2(1)$, and such that Ω restricted to e_{uv} is Γ_1, restricted to e_{vw} is Γ_2, and restricted to e_{uw} is $\Gamma_2 * \Gamma_1$. Put a metric $\widetilde{g}$ on T such that the edges are of length one, and take a smooth function $\widetilde{f} : T \to \mathbb{R}$ whose only critical points are u, v, w with $\mu(u) = 2, \mu(v) = 1$ and $\mu(w) = 0$, and the negative gradient $\widetilde{V} = -\nabla_{\widetilde{g}} \widetilde{f}$ is tangent to the edges of T and like $(1 + t)t(1 - t)$. Define the vector field V_Ω on $T \times X$ by $V_\Omega = \widetilde{V} + V_d$, where $V_d = -\nabla_{g_d} f_d$. We can define the critical points, flow lines, ascending and descending manifolds of V_Ω in the regular way. We say that Ω is an admissible family if the ascending and descending manifolds of V_Ω intersect transversely.

The critical points of index i of V_Ω are precisely

$$\mathrm{crit}_i(V_\Omega) = \{u\} \times \mathrm{crit}_{i-2}(f_u) \cup \{v\} \times \mathrm{crit}_{i-1}(f_v) \cup \{w\} \times \mathrm{crit}_i(f_w)$$

For critical points P, Q of V_Ω let $\mathcal{M}^\Omega(P, Q)$ be the moduli space of flow lines from P to Q, and $\widehat{\mathcal{M}}^\Omega(P, Q)$ the moduli space of unparameterized flow lines. For any admissible family Ω we have that $\mathcal{M}^\Omega(P, Q)$ is a $(\mu(P) - \mu(Q))$-dimensional manifold and $\widehat{\mathcal{M}}^\Omega(P, Q)$ is a $(\mu(P) - \mu(Q) - 1)$-dimensional manifold. Like the spaces $\widehat{\mathcal{M}}^\Psi$ considered above, the space $\widehat{\mathcal{M}}^\Omega(P, Q)$ may have boundary, and their compactifications can be quite complicated. The description of this space can vary greatly depending on where P and Q are, so we state the results we need in the only two special cases used.

If $\mu(P) = \mu(Q) + 1$, then $\widehat{\mathcal{M}}^\Omega(P, Q)$ is already compact, and since it is a 0-dimensional manifold we have that it is just a finite set of points. If $P = (u, p)$ and $Q = (w, q)$ with $\mu(p) = i = \mu(q)$, then $\mu((u, p)) = i + 2$ and $\mu((w, q)) = i$, hence $\widehat{\mathcal{M}}^\Omega((u, p), (w, q))$ is a 1-dimensional manifold and has a natural compactification to a 1-dimensional manifold $\overline{\widehat{\mathcal{M}}^\Omega((u, p), (w, q))}$ with boundary by adding broken flow lines. The broken flow lines that pass through one intermediate critical point of index $i + 1$ are contained in the boundary, and these points are precisely

$$\mathrm{crit}_{i+1}(V_\Omega) = \{u\} \times \mathrm{crit}_{i-1}(f_u) \cup \{v\} \times \mathrm{crit}_i(f_v) \cup \{w\} \times \mathrm{crit}_{i+1}(f_w)$$

79

But, we also have boundary points coming from those flow lines that stay in the $e_{uw} \times X$ slice for all time. These flow lines correspond to $\widehat{\mathcal{M}}^{\Gamma_2 * \Gamma_1}((u,p),(w,q))$, hence

$$\partial\overline{\widehat{\mathcal{M}}^{\Omega}((u,p),(w,q))} = \widehat{\mathcal{M}}^{\Gamma_2 * \Gamma_1}((u,p),(w,q))$$
$$\cup \bigcup_{r\in\mathrm{crit}_{i-1}(f_u)} \widehat{\mathcal{M}}^{\Omega}((u,p),(u,r)) \times \widehat{\mathcal{M}}^{\Omega}((u,r),(w,q))$$
$$\cup \bigcup_{r'\in\mathrm{crit}_{i+1}(f_w)} \widehat{\mathcal{M}}^{\Omega}((u,p),(w,r')) \times \widehat{\mathcal{M}}^{\Omega}((w,r'),(w,q))$$
$$\cup \bigcup_{s\in\mathrm{crit}_i(f_v)} \widehat{\mathcal{M}}^{\Omega}((u,p),(v,s)) \times \widehat{\mathcal{M}}^{\Omega}((v,s),(w,q))$$

Note that flow lines that stay in the $u \times X$ slice correspond to $\widehat{\mathcal{M}}^{f_u}(p,r)$, and similarly for flows that stay in the $w \times X$ slice. Therefore we have

$$\widehat{\mathcal{M}}^{\Omega}((u,p),(u,r)) = \widehat{\mathcal{M}}^{f_u}(p,r)$$
$$\widehat{\mathcal{M}}^{\Omega}((w,r'),(w,q)) = \widehat{\mathcal{M}}^{f_w}(r',q)$$

Further, the flows from (u,p) to (v,s) must stay in the $e_{uv} \times X$ slice and similarly for flows from (v,s) to (w,q), so we have

$$\widehat{\mathcal{M}}^{\Omega}((u,p),(v,s)) = \widehat{\mathcal{M}}^{\Gamma_1}((u,p),(v,s))$$
$$\widehat{\mathcal{M}}^{\Omega}((v,s),(w,q)) = \widehat{\mathcal{M}}^{\Gamma_2}((v,s),(w,q))$$

See figure 7.5.

We define $G_{\Omega} : CM_i(f_u, g_u) \to CM_{i+1}(f_w, g_w)$ on $p \in \mathrm{crit}_i(f_u)$ by

$$G_{\Omega}(p) = \sum_{q\in\mathrm{crit}_{i+1}(f_w)} \#\widehat{\mathcal{M}}^{\Omega}((u,p),(w,q)) \cdot q$$

and extend linearly to all of $CM_i(f_u, g_u)$.

Proposition 7.12. $\Phi_{\Gamma_2 * \Gamma_1}$ *is chain homotopic to* $\Phi_{\Gamma_2} \circ \Phi_{\Gamma_1}$.

Proof. Let $\Omega = \{(f_d, g_d) : d \in T\}$ be an admissible family between Γ_1, Γ_2 and $\Gamma_2 * \Gamma_1$. We claim that the map G_{Ω} defined above is a chain homotopy between $\Phi_{\Gamma_2 * \Gamma_1}$ and $\Phi_{\Gamma_2} \circ \Phi_{\Gamma_1}$, i.e. if we set $E = \partial^w \circ G_{\Omega} + G_{\Omega} \circ \partial^u - \Phi_{\Gamma_2 * \Gamma_1} + \Phi_{\Gamma_2} \circ \Phi_{\Gamma_1}$, then the coefficient of $q \in \mathrm{crit}_i(f_w)$ in $E(p)$ is zero for all $p \in \mathrm{crit}_i(f_u)$.

We can easily compute

$$G_{\Omega} \circ \partial^u(p) = \sum_{q\in\mathrm{crit}_i(f_w)} \sum_{r\in\mathrm{crit}_{i-1}(f_u)} \#\widehat{\mathcal{M}}^{f_u}(p,r) \cdot \#\widehat{\mathcal{M}}^{\Omega}((u,r),(w,q)) \cdot q$$
$$\partial^w \circ G_{\Omega}(p) = \sum_{q\in\mathrm{crit}_i(f_w)} \sum_{r'\in\mathrm{crit}_{i+1}(f_w)} \#\widehat{\mathcal{M}}^{\Omega}((u,p),(w,r')) \cdot \#\widehat{\mathcal{M}}^{f_w}(r',q) \cdot q$$
$$\Phi_{\Gamma_2} \circ \Phi_{\Gamma_1}(p) = \sum_{q\in\mathrm{crit}_i(f_w)} \sum_{s\in\mathrm{crit}_i(f_v)} \#\widehat{\mathcal{M}}^{\Gamma_1}((u,p),(v,r)) \cdot \#\widehat{\mathcal{M}}^{\Gamma_2}((v,r),(w,q)) \cdot q$$

Now it is clear that the coefficient of $q \in \mathrm{crit}_i(f_w)$ in $E(p)$, where $p \in \mathrm{crit}_i(f_u)$, is precisely $\#\overline{\partial\widehat{\mathcal{M}}^{\Omega}((u,p),(w,q))}$, hence zero.

$\square$
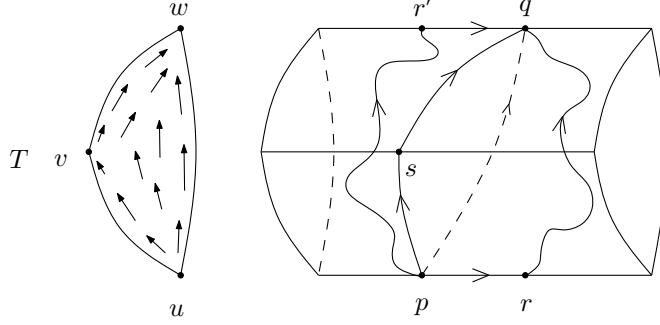
Figure 7.5: Counting flow lines from (u, p) to (w, q) in $T \times X$

Combining propositions 7.10 to 7.12 we get invariance of Morse homology.

Corollary 7.2. *For any two Morse-Smale pairs (f_0, g_0) and (f_1, g_1), the Morse homologies $HM_*(f_0, g_0)$ and $HM_*(f_1, g_1)$ are naturally isomorphic.*

7.6 Orientations on Moduli Spaces

There are (at least) two ways to introduce orientations on our moduli spaces $\mathcal{M}^f(p, q)$ and $\widehat{\mathcal{M}}^f(p, q)$ so that we can define the Morse complex with $\mathbb{Z}$ coefficients. One way is geometric and technically the simplest way to define orientations, but it does not generalize to the infinite dimensional versions of Morse homology. The other way is very technical and uses high powered machinery from functional analysis, but also has obvious interpretations in the infinite dimensional setting. We will give a superficial description of both of these orientations, and explain why they should give the same end result.

For the geometric orientations we must assume that our Riemannian manifold (X^n, g) is oriented. Let us also fix a Morse function f on X so that (f, g) is a Morse-Smale pair. We stated in proposition 7.1 that the ascending manifold $\mathscr{A}(p)$ and descending manifold $\mathscr{D}(p)$ of a critical point $p \in \mathrm{crit}_i(f)$ are smoothly embedded discs of dimension i and $n - i$, respectively. Let us define $\mathscr{M}^f(p, q)$ to be the intersection $\mathscr{D}(p) \cap \mathscr{A}(q)$, which is a submanifold of dimension $\mu(p) - \mu(q)$. We get an $\mathbb{R}$ action on $\mathscr{M}^f(p, q)$ via the 1-paremeter group of diffeomorphisms $\varphi_t : X \to X$ generated by $-\nabla_g f$: for $t \in \mathbb{R}$ and $x \in \mathscr{M}^f(p, q)$ we define $t \cdot x = \varphi_t(x)$. This action is free and smooth, so $\widehat{\mathscr{M}}^f(p, q) = \mathscr{M}^f(p, q)/\mathbb{R}$ is a $(\mu(p) - \mu(q) - 1)$-dimensional manifold.

Proposition 7.13. *The map $\phi : \mathcal{M}^f(p, q) \to \mathscr{M}^f(p, q)$ defined by $\phi(u) = u(0)$ is a diffeomorphism, and for each $t \in \mathbb{R}$ the following diagram commutes*

$$
\begin{array}{ccc}
\mathcal{M}^f(p, q) & \xrightarrow{t \cdot} & \mathcal{M}^f(p, q) \\
\phi \downarrow & & \downarrow \phi \\
\mathscr{M}^f(p, q) & \xrightarrow{t \cdot} & \mathscr{M}^f(p, q)
\end{array}
$$

where $t\cdot$ on the horizontal arrows denotes the $\mathbb{R}$ action on the respective spaces. In particular, $\widehat{\mathscr{M}}^f(p, q)$ is diffeomorphic to $\widehat{\mathcal{M}}^f(p, q)$.

Let us put an arbitrary orientation on $T_p\mathscr{D}(p)$ for all $p \in \mathrm{crit}(f)$. Since $T_p\mathscr{D}(p) \oplus T_p\mathscr{A}(p) = T_pX$, and T_pX is oriented by assumption, we get an induced orientation on $T_p\mathscr{A}(p)$. These orientations

at p uniquely extend to coherent orientations on all of $\mathscr{A}(p)$ and $\mathscr{D}(p)$, and so now all ascending and descending submanifolds are oriented. Now the spaces $\mathscr{M}^f(p,q)$ get an induced orientation, as described in section 1, so we can orient $\mathcal{M}^f(p,q)$ by requiring the isomorphism ϕ in proposition 7.13 to be orientation preserving.

UNFINISHED

7.7 Applications

The Morse inequalities are set of nice results that put topological restrictions on a space according to what type of Morse functions the space admits, and they nearly come for free with the definition of Morse homology. It was these inequalities that inspired Floer to define his infinite dimensional generalization of Morse homology in order to prove the Arnold conjecture. Since $HM_*(f,g)$ is isomorphic to $H_*(X)$ we already have that the Euler characteristic of X can be computed by $\chi(X) = \sum (-1)^i \# \operatorname{crit}_i(f)$ for any Morse function f on X. The Morse inequalities strengthen this quite a bit.

First we need to prove a general result concerning finitely generated chain complexes.

Theorem 7.2 (Euler-Poincaré Theorem)**.** *Take any finitely generated chain complex*

$$0 \longrightarrow C_n \xrightarrow{\partial_n} C_{n-1} \xrightarrow{\partial_{n-1}} \cdots \xrightarrow{\partial_2} C_1 \xrightarrow{\partial_1} C_0 \longrightarrow 0$$

and let $H_i = H_i(C_*, \partial_*)$. Then

$$\sum_{i=0}^{n} (-1)^i \operatorname{rank} C_i = \sum_{i=0}^{n} \operatorname{rank} H_i$$

Proof. At each $i = 0, \ldots, n$ we have a short exact sequence

$$0 \longrightarrow \ker \partial_i \longrightarrow C_i \xrightarrow{\partial_i} \operatorname{im} \partial_i \longrightarrow 0$$

hence $\operatorname{rank} C_i = \operatorname{rank} \ker \partial_i + \operatorname{rank} \operatorname{im} \partial_i$. We also have a short exact sequence

$$0 \longrightarrow \operatorname{im} \partial_{i+1} \longrightarrow \ker \partial_i \longrightarrow H_i \longrightarrow 0$$

hence $\operatorname{rank} \ker \partial_i = \operatorname{rank} \operatorname{im} \partial_{i+1} + \operatorname{rank} H_i$. There is a common $\operatorname{rank} \ker \partial_i$ in each of the equations we derived from these short exact sequences, and solving for it gives

$$\operatorname{rank} \ker \partial_i = \operatorname{rank} C_i - \operatorname{rank} \operatorname{im} \partial_i = \operatorname{rank} \operatorname{im} \partial_{i+1} + \operatorname{rank} H_i$$

Therefore

$$\sum_{i=0}^{n} (-1)^i \left(\operatorname{rank} C_i - \operatorname{rank} \operatorname{im} \partial_i \right) = \sum_{i=0}^{n} (-1)^i \left(\operatorname{rank} \operatorname{im} \partial_{i+1} + \operatorname{rank} H_i \right)$$

On the left side we have that $\operatorname{rank} \operatorname{im} \partial_i$ appears with sign $(-1)^{i+1}$, and on the right $\operatorname{rank} \operatorname{im} \partial_i$ appears with sign $(-1)^{i-1}$, hence they cancel and we are left with

$$\sum_{i=0}^{n} (-1)^i \operatorname{rank} C_i = \sum_{i=0}^{n} (-1)^i \operatorname{rank} H_i$$

$\square$

Theorem 7.3 (Morse Inequalities). *Let $f : X^n \to \mathbb{R}$ be a Morse function. Let $c_i = \# \operatorname{crit}_i(f)$ be the number of critical points of index i, and let $b_i = \operatorname{rank} H_i(X)$ be the i-th Betti number of X. Then*

$$c_k - c_{k-1} + \cdots + (-1)^k c_0 \geq b_k - b_{k-1} + \cdots + (-1)^k b_0 \tag{7.5}$$

for all $0 \leq k \leq n$.

Proof. We have $\operatorname{rank} HM_i(f,g) = \operatorname{rank} H_i(X) = b_i$ and $\operatorname{rank} CM_i(f,g) = \# \operatorname{crit}_i(f)$. Since $HM_i(f,g)$ is a subquotient of $CM_i(f,g)$, we clearly have $c_i \geq b_i$. Let $0 \leq k \leq n$, and let (C_*, ∂_*) denote the chain complex (CM_*, ∂_*) truncated at k:

$$0 \longrightarrow C_k \xrightarrow{\partial} C_{k-1} \xrightarrow{\partial} \cdots \xrightarrow{\partial} C_1 \xrightarrow{\partial} C_0 \longrightarrow 0$$

The chain complex (C_*, ∂_*) is finitely generated, so by the Euler-Poincaré theorem we have

$$\sum_{i=0}^{k} (-1)^{i+k} \operatorname{rank} C_i = \sum_{i=0}^{k} (-1)^{i+k} \operatorname{rank} H_i(C_*, \partial_*) \tag{7.6}$$

The inserted signs $(-1)^k$ are necessary so that the last term $(i = k)$ in the sums are non-negative. The homology of the chain complex (C_*, ∂_*) is equal to the homology of (CF_*, ∂_*) in all gradings except possibly the last, so we have $\operatorname{rank} H_i(C_*, \partial_*) = b_i$ for all $0 \leq i \leq k - 1$. For the last term we have $H_k(C_*, \partial_*) = \ker \partial_k$, hence $H_k(CM_*, \partial_*)$ is a quotient of $H_k(C_*, \partial_*)$, and so we have $b_k \leq \operatorname{rank} H_k(C_*, \partial_*)$. By replacing the last term $\operatorname{rank} H_k(C_*, \partial_*)$ in (7.6) with the smaller term b_k we obtain the Morse inequality

$$c_k - c_{k-1} + \cdots + (-1)^k c_0 \geq b_k - b_{k-1} + \cdots + (-1)^k b_0$$

$\square$

7.8 Witten's Complex

In a paper on supersymmetry in physics, Witten was able to rediscover the Morse inequalities (see theorem 7.3) and Morse (co)homology by considering a simple supersymmetric theory on the exterior bundle $\Lambda^* T^* X$ of a compact, orientable manifold X^n. First we discuss the mathematical content of his results without mentioning any of the physics.

Let us recall some of the basic results of Hodge theory. If we put a metric g on X, then we get an induced volume form vol_g, and we get a star operator $\star : \Omega^k(X) \to \Omega^{n-k}(X)$ which satisfies $\star^2 = (-1)^{k(n-k)} \operatorname{id}_{\Omega^k(X)}$. The de Rham operator $d : \Omega^k(X) \to \Omega^{k+1}(X)$ has a formal adjoint $\delta : \Omega^k(X) \to \Omega^{k-1}(X)$, which can be defined by $\delta = (-1)^{kn+n+1} \star d\star$. We define the Laplace operator $\Delta : \Omega^k(X) \to \Omega^k(X)$ by $\Delta = d\delta + \delta d$. This operator has discrete spectrum, and decomposes $\Omega^k(X)$ into eigenspaces

$$\Omega^k(X) = \bigoplus_{\lambda \in \mathbb{R}} E_\lambda^k \quad \text{where} \quad E_\lambda^k = \left\{ \omega \in \Omega^k(X) : \Delta \omega = \lambda \omega \right\}$$

The fundamental theorem of Hodge theory is the following.

Theorem 7.4 (Hodge's Theorem). $E_0^k \cong H_{dR}^k(X)$

So, we can think of (E_0^*, d) has a finite dimensional chain complex (the differential happens to be zero) that "computes" the homology of the infinite dimensional chain complex $(\Omega^*(X), d)$. We clearly have $d(E_\lambda^k) \subseteq E_\lambda^{k+1}$, so for $\lambda > 0$ we also get a chain complex

$$0 \longrightarrow E_\lambda^0 \xrightarrow{d} E_\lambda^1 \xrightarrow{d} \cdots \xrightarrow{d} E_\lambda^n \longrightarrow 0$$

However, this sequence is exact. To see this, let $\omega \in E_\lambda^k$ such that $d\omega = 0$, then

$$d\left(\lambda^{-1}\delta\omega\right) = \lambda^{-1}d\delta\omega = \lambda^{-1}\Delta\omega = \lambda^{-1}\lambda\omega = \omega$$

therefore ω is exact. This means that the homology of the complex (E_λ^*, d) is zero. Now, for fixed $\lambda_0 > 0$, let

$$\Omega_{\lambda_0}^k = \bigoplus_{\lambda \le \lambda_0} E_\lambda^k$$

That is, $\Omega_{\lambda_0}^*$ consists of all the eigenforms of Δ with eigenvalue less than or equal to λ_0. Now, the restriction of d to $\Omega_{\lambda_0}^*$ gives a *finite* dimensional complex $(\Omega_{\lambda_0}^*, d)$, and since this complex is a direct sum of exact sequences (when $\lambda \in (0, \lambda_0]$) and a sequence with homology equal to de Rham homology (when $\lambda = \lambda_0$), we have

$$H^*(\Omega_{\lambda_0}^*, d) = H_{dR}^*(X)$$

If we have chosen λ_0 large enough so that there is an eigenvalue of Δ in $(0, \lambda_0]$, then we have replaced the *infinite* dimension complex (Ω^*, d) with a non-trivial *finite* dimensional complex $(\Omega_{\lambda_0}^*, d)$ without changing the homology.

We now repeat these same ideas with a slight variation of Hodge theory. For a fixed Morse function $f : X \to \mathbb{R}$ and real parameter t we define the operator $d_t : \Omega^k(X) \to \Omega^{k+1}(X)$ by

$$d_t = e^{-ft}de^{ft}$$

that is, we conjugate the de Rham operator by the linear isomorphism $\omega \mapsto e^{-ft}\omega$. The function f will remained fixed in our discussion, but we will be taking t to be large soon, so we only put t in our notation. We clearly have $d_t^2 = 0$, and this operator has the formal adjoint $\delta_t : \Omega^k(X) \to \Omega^{k-1}(X)$ defined by

$$\delta_t = e^{ft}\delta e^{-ft}$$

and so we define the Laplacian $\Delta_t = d_t\delta_t + \delta_t d_t$. We can also define the homology groups $H_t^*(X) = \ker d_t / \operatorname{im} d_t$, but the isomorphism $\Omega^k \to \Omega^k$ given by $\omega \mapsto e^{-ft}\omega$ descends to an isomorphism $H_{dR}^*(X) \cong H_t^*(X)$, so this homology theory does not give us anything new. Like before we get an eigenspace decomposition

$$\Omega^k(X) = \bigoplus_{\lambda \in \mathbb{R}} E_\lambda^k(t) \qquad \text{where} \qquad E_\lambda^k(t) = \left\{\omega \in \Omega^k(X) : \Delta_t\omega = \lambda\omega\right\}$$

We also have a Hodge-like theorem that says $E_0^k(t) \cong H_{dR}^k(X)$. Also like before, for $\lambda > 0$ we have an exact sequence

$$0 \longrightarrow E_\lambda^0(t) \xrightarrow{d_t} E_\lambda^1(t) \xrightarrow{d_t} \cdots \xrightarrow{d_t} E_\lambda^n(t) \longrightarrow 0$$

So, if for fixed $\lambda_0 > 0$ we define $\Omega_{\lambda_0}^k(t) = \bigoplus_{\lambda \le \lambda_0} E_\lambda^k(t)$, then the *finite* dimensional complex $(\Omega_{\lambda_0}^*(t), d_t)$ has the same homology as (Ω^*, d). At this point Witten claims that for sufficiently

84

large $t >> 0$, the dimensions $\dim \Omega^k_{\lambda_0}(t)$ become independent of t, and so let us write $(\Omega^*_{\lambda_0}(\infty), d_\infty)$ to denote this "stabilized" chain complex. In fact, Witten shows that

$$\dim \Omega^k_{\lambda_0}(\infty) = \# \operatorname{crit}_k(f) = c_k$$

As we saw in theorem 7.3, whenever we have a finite dimensional complex whose chain groups correspond to critical points of a Morse function, and whose homology computes the homology of X, then we get the Morse inequalities for free. So, we have proved the Morse inequalities by considering a variation of Hodge theory.

Now let us see the physical ideas behind the above results. Using the definition of d_t and δ_t, one can easily compute that these operators act on a k-form $\omega \in \Omega^k(X)$ by

$$d_t \omega = d\omega + t df \wedge \omega \qquad \delta_t \omega = \delta \omega - t i_{df} \omega$$

If we choose coordinates (x, U) on X, and define the operators $A_i^* \omega = i_{\frac{\partial}{\partial x^i}} \omega$ and $A_i \omega = dx^i \wedge \omega$, then one can show that Δ_t can be written as

$$\Delta_t = \Delta + t^2 \, ||df||^2 + t \sum_{i=1}^n \frac{\partial^2 f}{\partial x^i \partial x^j} [A_i^*, A_j]$$

As t gets large we expect the eigenforms of this operator to be increasingly concentrated at the critical points of f. This leads us to look around the critical points. If $p \in \operatorname{crit}(f)$, then we can choose normal coordinates (x, U) around p such that f is of the form

$$f = f(p) + \sum_{i=1}^n \lambda_i (x^i)^2 + O(x^3)$$

where the λ_i's are the eigenvalues of the Hessian of f at p. In these coordinates, if we ignore terms of order 3 and above, then Laplacian Δ_t can be approximated by

$$\widetilde{\Delta}_t = \sum_{i=1}^n \left(-\frac{\partial^2}{(\partial x^i)^2} + t^2 \lambda_i^2 (x^i)^2 + t \lambda_i [A_i^*, A_i] \right)$$

This looks like a simple harmonic oscillator, plus some other operator. Recall that for $a > 0$, the simple harmonic oscillator

$$H = -\frac{d^2}{dx^2} + a^2 x^2$$

has spectrum $\{(2k+1)a : k = 0, 1, \ldots\}$. So, let us write our approximation of the Laplacian as

$$\widetilde{\Delta}_t = \sum_{i=1}^n (H_i + t \lambda_i K_i)$$

where $H_i = -\partial^2/(\partial x^i)^2 + t^2 \lambda_i (x^i)^2$ and $K_i = [A_i^*, A_i]$. The H_i's commute with themselves, the K_i's commute with themselves, and the H_i's commute with the K_j's. Therefore the spectrum of $\widetilde{\Delta}_t$ is the sum of the spectrum of the H_i's and K_i's. The spectral of H_i is given by $\{(2N_i + 1)t|\lambda_i| : N_i = 0, 1, \ldots\}$. One can easily compute that

$$K_i dx^{i_1} \wedge \cdots \wedge dx^{i_k} = \begin{cases} +dx^{i_1} \wedge \cdots \wedge dx^{i_k} & i \in \{i_1, \ldots, i_k\} \\ -dx^{i_1} \wedge \cdots \wedge dx^{i_k} & i \notin \{i_1, \ldots, i_k\} \end{cases}$$

so the eigenvalues of this operator are just $+1$'s and -1's. In fact, the number of $+1$ eigenvalues is precisely k. Now, the spectrum of $\widetilde{\Delta}_t$ can be written as the set

$$\left\{ t \sum_{i=1}^{n} \left((2N_i + 1)|\lambda_i| + n_i\lambda_i \right) : \begin{array}{c} N_i = 0, 1, \ldots \\ n_i = \pm 1 \\ \#\{n_i = +1\} = k \end{array} \right\}$$

What choice of N_i and n_i give zero? If we rewrite the summand as $2N_i|\lambda_i| + |\lambda_i n_i| + n_i\lambda_i$, then we see that we are forced to have $N_i = 0$ and $n_i = -\operatorname{sign}(\lambda_i)$. Therefore, there is exactly one zero eigenvalue of $\widetilde{\Delta}_t$, and all other eigenvalues blow up to ∞ since they are proportional to t. Further, associated eigenform is a k-form, where

$$k = \#\{n_i = +1\} = \#\{\lambda_i < 0\} = \mu(p)$$

So, at each critical point $p \in \operatorname{crit}(f)$, we are able to associate a $\mu(p)$-form ω such that $\widetilde{\Delta}_t\omega = 0$. Since $\widetilde{\Delta}_t$ is only an approximation of Δ_t, we do not necessarily have that $\Delta_t\omega = 0$. Therefore, the number of k-forms ω which satisfy $\widetilde{\Delta}_t\omega = 0$ at some critical point of index k (i.e. the number of critical points of index k) is greater than or equal to the number of k-forms with $\Delta\omega = 0$ (i.e. the k-th Betti number of X). This proves the weak Morse inequalities.

8 Heegaard Floer Homology

8.1 Construction for 3-Manifolds

The Morse homology discussed in the previous section has inspired variations that usually go under the heading of Floer-type theories. These theories start with some differential equation (in Morse homology this was the negative gradient flow equation), then one forms a graded group freely generated by some points in a space or some other objects, and then one defines a differential to count the number of solutions to the differential equation that start at one point and end at another. Hopefully the differential squares to zero so that one can define the homology groups, and hopefully the homology groups do not depend on the various choices used in the construction of this homology theory. We now follow this plan to build a homology theory for 3-manifolds from Heegaard diagrams.

General Idea

First, we will give a general idea of the object we are trying to construct. The details in this construction are very difficult, like it was for Morse homology, requiring a lot of machinery from analysis and symplectic geometry. Let $(\Sigma, \alpha_1, \ldots, \alpha_g, \beta_1, \ldots, \beta_g)$ be a genus g Heegaard diagram for a 3-manifold Y, and assume the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ curves are in general position so that they intersect transversally. Let us also fix a base point z in the complement of the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ curves. We can form the **symmetric product** $\mathrm{Sym}^g(\Sigma)$ of Σ, defined to be the quotient of $\Sigma^{\times g}$ under the action of S_g permuting the components of $\Sigma^{\times g}$, i.e. $\mathrm{Sym}^g(\Sigma)$ consists of g-tuples of unordered points of Σ. Even though S_g does not act freely on Σ, we do have that $\mathrm{Sym}^g(\Sigma)$ is a smooth manifold, and in fact a complex structure j on Σ induces a complex structure $\mathrm{Sym}^g(\mathrm{j})$ on $\mathrm{Sym}^g(\Sigma)$.

Since the $\boldsymbol{\alpha}$ curves never intersect each other (and similarly for the $\boldsymbol{\beta}$ curves), we have that the products

$$\mathbb{T}_\alpha = \alpha_1 \times \cdots \times \alpha_g \qquad \mathbb{T}_\beta = \beta_1 \times \cdots \times \beta_g$$

are smoothly embedded submanifolds of $\mathrm{Sym}^g(\Sigma)$, and $\mathbb{T}_\alpha \pitchfork \mathbb{T}_\beta$, hence the intersect consists of finitely many points. The base point z determines a codimension 2 submanifold V_z of $\mathrm{Sym}^g(\Sigma)$ defined by

$$V_z = \{z\} \times \mathrm{Sym}^{g-1}(\Sigma) \subset \mathrm{Sym}^g(\Sigma)$$

Clearly V_z is disjoint from the tori $\mathbb{T}_\alpha$ and $\mathbb{T}_\beta$.

Let $\mathbb{D} \subset \mathbb{C}$ denote the unit disc, and let $e_1 = \partial\mathbb{D} \cap \{\mathrm{Re}(z) \geq 0\}$ and $e_2 = \partial\mathbb{D} \cap \{\mathrm{Re}(z) \leq 0\}$. For two intersection points $\mathbf{x}, \mathbf{y} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$ (this means that $\mathbf{x}$ and $\mathbf{y}$ are unordered g-tuples of intersection points of the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ curves) we define a **Whitney disc** connecting $\mathbf{x}$ to $\mathbf{y}$ to be a continuous map $u : \mathbb{D} \to \mathrm{Sym}^g(\Sigma)$ such that $u(-i) = \mathbf{x}$, $u(i) = \mathbf{y}$, $u(e_1) \subset \mathbb{T}_\alpha$ and $u(e_2) \subset \mathbb{T}_\beta$. Let $\pi_2(\mathbf{x}, \mathbf{y})$ denote the homotopy classes of Whitney discs from $\mathbf{x}$ to $\mathbf{y}$.

We define a function $n_z : \pi_2(\mathbf{x}, \mathbf{y}) \to \mathbb{Z}$ to be the algebraic intersection number of a disc $\phi : \mathbb{D} \to \mathrm{Sym}^g(\Sigma)$ with the submanifold V_z. For a Whitney disc ϕ, let $\mathcal{M}(\phi)$ denote the moduli space of pseudoholomorphic representatives of ϕ, i.e. the set of representatives $u : \mathbb{D} \to \mathrm{Sym}^g(\Sigma)$ of ϕ that satisfies the Cauchy-Riemann equation

$$\frac{\partial u}{\partial s} + i\frac{\partial u}{\partial t} = 0$$

If we introduce the proper perturbations of our set up, then $\mathcal{M}(\phi)$ will be a smooth manifold, and we denote its dimension by $\mu(\phi)$. There is a natural free $\mathbb{R}$ acts on this space, and so we denote the quotient by $\widehat{\mathcal{M}}(\phi) = \mathcal{M}(\phi)/\mathbb{R}$.

With these definitions we can give an intuitive (but slightly incorrect) definition of Heegaard Floer homology. Let $\widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ be the Abelian group freely generated by the points in $\mathbb{T}_\alpha \cap \mathbb{T}_\beta$. We define a differential $\widehat{\partial} : \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \to \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ on $\mathbf{x} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$ by

$$\widehat{\partial}\mathbf{x} = \sum_{\left\{ \begin{array}{c} \mathbf{y} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta \,\big|\, n_z(\phi)=0 \\ \phi \in \pi_2(\mathbf{x},\mathbf{y}) \,\big|\, \mu(\phi)=1 \end{array} \right\}} \#\widehat{\mathcal{M}}(\phi) \cdot \mathbf{y}$$

That is, we are counting the number of pseudoholomorphic discs between two points in $\mathrm{Sym}^g(\Sigma)$. One can show that $\widehat{\partial}^2 = 0$, so we can define the homology groups $\widehat{HF}(\boldsymbol{\alpha}, \boldsymbol{\beta})$. One can also show that $\widehat{HF}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ does not depend on the Heegaard diagram or complex structures chosen in order to define it, hence $\widehat{HF}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is a topological invariant. Now we work to properly define these homology groups.

Morse Functions and Pointed Heegaard Diagrams

Let us fix a Heegaard diagram $(\Sigma, \alpha_1, \ldots, \alpha_g, \beta_1, \ldots, \beta_g)$ for our 3-manifold Y, where Σ is a surface of genus g. Recall that this means we attach a 2-handle for each $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ curve, the result being a 3-manifold with boundary two disjoint 2-spheres, and so we attach two 3-handles in the only way possible. So, the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ curves determine handlebodies U_0 and U_1 of genus g such that $U_0 \cap U_1 = \Sigma$. We also have the following Morse theoretic picture of this set up. Let $f : Y \to [0, 3]$ be a self-indexing Morse function with g critical points of index 1, denote them by $P_1, \ldots, P_g$, and hence g critical points of index 2, denote them by $Q_1, \ldots, Q_g$. Then $U_0 = f^{-1}[0, 3/2], U_1 = f^{-1}[3/2, 3]$ and $\Sigma = f^{-1}(3/2)$, and the attaching circles can be defined by $\alpha_i = \mathscr{A}(P_i) \cap \Sigma$ and $\beta_i = \mathscr{D}(Q_i) \cap \Sigma$. We say that a Morse function is **compatible** with a Heegaard diagram if the Heegaard diagram is equal to the induced Heegaard diagram from the Morse function.

We can make this Heegaard diagram into a **pointed** Heegaard diagram by choosing a base point $z \in \Sigma$ in the complement of the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ curves. We can consider pointed Heegaard moves on pointed Heegaard diagrams by requiring that isotopies of attaching circles never pass through z, and the pair of pants defining a handle slide is not allowed to contain z.

The Algebraic Topology of Symmetric Products

Let D denote the *fat diagonal* in $\mathrm{Sym}^g(\Sigma)$; that is, D consists of those ordered g-tuples of points in Σ such that at least points are equal. The embedded tori $\mathbb{T}_\alpha$ and $\mathbb{T}_\beta$ miss the diagonal, and since the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ curves intersect transversally, we have $\mathbb{T}_\alpha \pitchfork \mathbb{T}_\beta$ in $\mathrm{Sym}^g(\Sigma)$. For a complex structure j on Σ, we get an induced complex structure $\mathrm{Sym}^g(\mathrm{j})$ on $\mathrm{Sym}^g(\Sigma)$ by requiring the quotient map $\pi : \Sigma^{\times g} \to \mathrm{Sym}^g(\Sigma)$ to be holomorphic, where $\Sigma^{\times g}$ is given the product complex structure $\mathrm{j}^{\times g}$. The homotopy and homology groups of $\mathrm{Sym}^g(\Sigma)$ in low dimensions is stated in the following lemmas.

Lemma 8.1. *If Σ is a genus g surface, then $\pi_1(\mathrm{Sym}^g(\Sigma)) \cong H_1(\mathrm{Sym}^g(\Sigma)) \cong H_1(\Sigma)$.*

Proof. One can show that $\pi_1(\mathrm{Sym}^g(\Sigma))$ is abelian, so that gives the first isomorphism. For a fixed point $x \in \Sigma$, the inclusion $\Sigma \times \{x\} \times \cdots \times \{x\} \hookrightarrow \mathrm{Sym}^g(\Sigma)$ induces an isomorphism $H_1(\Sigma) \to H_1(\mathrm{Sym}^g(\Sigma))$ on homology. To construct the inverse of this map we take a generic loop $\gamma : S^1 \to \mathrm{Sym}^g(\Sigma)$. If we pullback the $g!$-fold branched cover $\pi : \Sigma^{\times g} \to \mathrm{Sym}^g(\Sigma)$ along γ we get a $g!$-fold

branched cover $p : \widetilde{S^1} \to S^1$ such that the following commutes

$$
\begin{array}{ccc}
\widetilde{S^1} & \xrightarrow{\widetilde{\gamma}} & \Sigma^{\times g} \\
{\scriptstyle p}\downarrow & & \downarrow{\scriptstyle \pi} \\
S^1 & \xrightarrow[\gamma]{} & \mathrm{Sym}^g(\Sigma)
\end{array}
$$

The space $\widetilde{S^1}$ is a disjoint union of circles and has a natural S_g action on it such that $\widetilde{\gamma}$ is equivariant. Let $\widehat{S^1} = \widetilde{S^1}/S_g$, and let $\widehat{\gamma} : \widehat{S^1} \to \Sigma$ be the map induced by $\pi_1 \circ \widetilde{\gamma}$, where $\pi_1 : \Sigma^{\times g} \to \Sigma$ is the projection onto the first factor. Then the map $H_1(\mathrm{Sym}^g(\Sigma)) \to H_1(\Sigma)$ given by $[\gamma] \mapsto \widehat{\gamma}[\widehat{S^1}]$ is the inverse map. $\qquad \square$

Recall that the fundamental group π_1 of a space naturally acts on the higher homotopy groups π_n. Let π_n' denote the quotient of π_n by the action of π_1.

Lemma 8.2. *If Σ is a surface of genus $g > 1$, then $\pi_2'(\mathrm{Sym}^g(\Sigma)) \cong \mathbb{Z}$. Further, if $g > 2$, then $\pi_1(\mathrm{Sym}^g(\Sigma))$ acts trivially on $\pi_2(\mathrm{Sym}^g(\Sigma))$, so $\pi_2(\mathrm{Sym}^g(\Sigma)) \cong \mathbb{Z}$ for $g > 2$.*

Proof. The isomorphism $\pi_2'(\mathrm{Sym}^2(\Sigma)) \cong \mathbb{Z}$ is given by counting the algebraic intersection number of a sphere with the subvariety $x \times \mathrm{Sym}^{g-1}(\Sigma)$ for generic x. We can find an explicit generator. The hyperelliptic involution on Σ gives rise to a sphere $S_0 \subset \mathrm{Sym}^2(\Sigma)$, and so let $S = S_0 \times x_3 \times \cdots \times x_g \subset \mathrm{Sym}^g(\Sigma)$. Then S maps to 1 under the above isomorphism. $\qquad \square$

Lemma 8.3. *The first Chern class c_1 of $\mathrm{Sym}^g(\Sigma)$ satisfies $\langle c_1, [S] \rangle = 1$.*

Lemma 8.4.
$$
\frac{H_1(\mathrm{Sym}^g(\Sigma))}{H_1(\mathbb{T}_\alpha) \oplus H_1(\mathbb{T}_\beta)} \cong \frac{H_1(\Sigma)}{\langle [\alpha_1], \ldots, [\alpha_g], [\beta_1], \ldots, [\beta_g] \rangle} \cong H_1(Y)
$$

Proof. Since the inclusion of Σ into $\mathrm{Sym}^g(\Sigma)$ induces an isomorphism on first homology, we see that the subgroup of $H_1(\Sigma)$ generated by the $\boldsymbol{\alpha}$ curves gets maps to $H_1(\mathbb{T}_\alpha)$ in $H_1(\mathrm{Sym}^g(\Sigma))$, and similarly for the $\boldsymbol{\beta}$ curves, and so the first isomorphism follows. The second isomorphism is obvious since we building Y from Σ by filling the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ curves with discs. $\qquad \square$

There is a Morse theoretic interpretation of intersection points $\mathbf{x} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$. Let f be a Morse function compatible with our Heegaard diagram with critical points $P_1, \ldots, P_g$ of index 1 and critical points $Q_1, \ldots, Q_g$ of index 2. If $\mathbf{x} = \{x_1, \ldots, x_g\}$ such that $x_i \in \alpha_i \cap \beta_{\sigma(i)}$ (for some permutation $\sigma \in S_g$), then we have $x_i \in \mathcal{A}(P_i) \cap \mathcal{D}(Q_{\sigma(i)})$, hence x_i can be identified with an unparameterized, gradient flow line from P_i to $Q_{\sigma(i)}$. The intersection point $\mathbf{x}$ can therefore be identified with a g-tuple of flow lines from index 1 critical points to index 2 critical points, which we will denote by $\gamma_{\mathbf{x}}$.

The tori $\mathbb{T}_\alpha$ and $\mathbb{T}_\beta$ in $\mathrm{Sym}^g(\Sigma)$ have a nice compatibility with the complex structure $\mathrm{Sym}^g(\mathfrak{j})$ induced by $\mathfrak{j}$. First, recall the definition of a totally real submanifold of a complex manifold. If $L \subset Z$ is a submanifold of the complex manifold (Z, J), then we say that L is totally real if $T_pL \cap JT_pL = 0$ for all $p \in L$, i.e. the tangent spaces of L do not contain any J-complex lines. It is clear that the tori $\mathbb{T}_\alpha, \mathbb{T}_\beta$ are totally real when viewed as submanifolds of $(\Sigma^{\times g}, \mathfrak{j}^{\times g})$. Since the projection $\pi : \Sigma^{\times g} \to \mathrm{Sym}^g(\Sigma)$ is holomorphic and a $g!$-fold cover away from the diagonal D in $\mathrm{Sym}^g(\Sigma)$, it follows that $\mathbb{T}_\alpha$ and $\mathbb{T}_\beta$ are totally real submanifolds of $\mathrm{Sym}^g(\Sigma)$.

Whitney Discs in the Symmetric Product

There is an obstruction to the existence of Whitney discs between two intersection points $\mathbf{x}, \mathbf{y} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$. By renaming the $\boldsymbol{\beta}$ curves, if necessary, we can assume that $x_i \in \alpha_i \cap \beta_i$ and $y_i \in \alpha_i \cap \beta_{\sigma(i)}$, for some permutation $\sigma \in S_g$. Take a pair of paths $a : [0,1] \to \mathbb{T}_\alpha$, $b : [0,1] \to \mathbb{T}_\beta$ that start at $\mathbf{x}$ and end at $\mathbf{y}$. The difference $a - b$ is a loop in $\mathrm{Sym}^g(\Sigma)$, and hence determines a homology class. This element vanishes if and only if there is a Whitney disc connecting $\mathbf{x}$ to $\mathbf{y}$. Unfortunately, this homology class depends on the paths a and b, but we can remove this dependence in the following way. The path a can be written as $a(t) = (a_1(t), \ldots, a_g(t))$, where $a_i : [0,1] \to \alpha_i$, hence any other paths $[0,1] \to \mathbb{T}_\alpha$ differs from a by winding around the $\boldsymbol{\alpha}$ curves some number of times. The ways this winding can occur is parameterized by $H_1(\mathbb{T}_\alpha) \cong \mathbb{Z}^g$. Similar statements hold for the curve b and the $\boldsymbol{\beta}$ curves. So, by modding the subgroups $H_1(\mathbb{T}_\alpha)$ and $H_1(\mathbb{T}_\beta)$ out of $H_1(\mathrm{Sym}^g(\Sigma))$ the homology class of $a - b$ does not depend on a or b, and so the image of this homology class under the isomorphisms in lemma 8.4 in $H_1(Y)$ is denoted by $\epsilon(\mathbf{x}, \mathbf{y})$.

There is also a way of computing this homology class on Σ. Note that a curve $a : [0,1] \to \mathbb{T}_\alpha$ can be thought of as a g-tuple of sub-arcs of the $\boldsymbol{\alpha}$ curves such that the boundary (as a 1-chain) is $\partial a = y_1 + \cdots + y_g - x_1 - \cdots - x_g$. There is a similar description for a path $b : [0,1] \to \mathbb{T}_\beta$ and the $\boldsymbol{\beta}$ curves, hence the difference $a - b$ is a 1-cycle in Σ. This cycle depends on the paths used to define it, but only up to additional windings around $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ curves, hence it is a well-defined element of $H_1(\Sigma) / \langle [\alpha_1], \ldots, [\alpha_g], [\beta_1], \ldots, [\beta_g] \rangle$. The image of this homology class in $H_1(Y)$ under the isomorphism in lemma 8.4 is equal to $\epsilon(\mathbf{x}, \mathbf{y})$ that we defined above.

Clearly $\pi_2(\mathbf{x}, \mathbf{y})$ is empty if and only if $\epsilon(\mathbf{x}, \mathbf{y}) \neq 0$. Further, for $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$ we have the additive relation $\epsilon(\mathbf{x}, \mathbf{y}) + \epsilon(\mathbf{y}, \mathbf{z}) = \epsilon(\mathbf{x}, \mathbf{z})$, so we can define an equivalence relation on $\mathbb{T}_\alpha \cap \mathbb{T}_\beta$ by $\mathbf{x} \sim \mathbf{y}$ if $\epsilon(\mathbf{x}, \mathbf{y}) = 0$.

The set $\pi_2(\mathbf{x}, \mathbf{y})$ is not a group, but has a natural $\pi_2'(\mathrm{Sym}^g(\Sigma))$ action given by splicing a sphere onto a Whitney disc:

$$* : \pi_2'(\mathrm{Sym}^g(\Sigma)) \times \pi_2(\mathbf{x}, \mathbf{y}) \longrightarrow \pi_2(\mathbf{x}, \mathbf{y})$$

We can also splice a disc from $\mathbf{x}$ to $\mathbf{y}$ with a disc from $\mathbf{y}$ to $\mathbf{z}$ to get a disc from $\mathbf{x}$ to $\mathbf{z}$, giving a map

$$* : \pi_2(\mathbf{y}, \mathbf{z}) \times \pi_2(\mathbf{x}, \mathbf{y}) \longrightarrow \pi_2(\mathbf{x}, \mathbf{z})$$

This splicing action makes $\pi(\mathbf{x}, \mathbf{x})$ into a group. The algebraic intersection map n_z is clearly additive with respect to these splicing actions: $n_z(\psi * \phi) = n_z(\psi) + n_z(\phi)$.

Since it is difficult to visualize discs in $\mathrm{Sym}^g(\Sigma)$ we would like a way of representing them on the surface Σ. Recall that the function $n_w : \pi_2(\mathbf{x}, \mathbf{y}) \to \mathbb{Z}$ counts the intersection of the subvariety $\{w\} \times \mathrm{Sym}^g(\Sigma)$ with a disk $\phi \in \pi_2(\mathbf{x}, \mathbf{y})$. Let $D_1, \ldots, D_m$ denote the closures of the components of $\Sigma - \alpha_1 - \cdots - \alpha_g - \beta_1 - \cdots - \beta_g$, and choose a point z_i in the interior of each D_i. For a Whitney disc $u : \mathbb{D} \to \mathrm{Sym}^g(\Sigma)$ we define the **domain** of ϕ to be the form linear combination

$$\mathcal{D}(u) = \sum_{i=1}^m n_{z_i}(u) D_i$$

If $n_{z_i}(\phi) \geq 0$ for all i, then we write $\mathcal{D}(\phi) \geq 0$. This association does not depend on the points z_i, and clearly depends only on the homotopy class of u, hence we get a well-defined map on $\pi_2(\mathbf{x}, \mathbf{y})$. Domains are additive with respect to the splicing action defined above:

$$\mathcal{D}(\psi * \phi) = \mathcal{D}(\psi) + \mathcal{D}(\phi)$$

$$\mathcal{D}(S * \phi) = \mathcal{D}(\phi) + \sum D_i$$

For $\phi \in \pi_2(\mathbf{x}, \mathbf{y})$, we can think of $\mathcal{D}(\phi)$ as a 2-chain on Σ in the following way. Let $u \in \phi$ be a representative of ϕ, then the restriction $u|_{e_1}$ can be thought of as a curve from $\mathbf{x}$ to $\mathbf{y}$ contained in $\mathbb{T}_\alpha$, and $u|_{e_2}$ can be thought of as a curve from $\mathbf{y}$ to $\mathbf{x}$ contained in $\mathbb{T}_\beta$. As noted before, curves of this type determine 1-chains in Σ with boundaries $\partial u|_{e_1} = y_1 + \cdots + y_g - x_1 - \cdots - x_g$ and $\partial u|_{e_2} = x_1 + \cdots + x_g - y_1 - \cdots - y_g$. So, the boundary of the 2-chain $\mathcal{D}(u)$ is precisely

$$\partial \mathcal{D}(u) = \sum_i n_{z_i}(u)(\partial u|_{e_1} + \partial u|_{e_2})$$

A 2-chain of the form $P = \sum a_i D_i$ is said to be a *periodic domain* if the boundary ∂P is a sum of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ curves and such that if D_i is the component containing z, then $a_i = 0$ (i.e. $n_z(P) = 0$). For $\mathbf{x} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$, a class $\phi \in \pi_2(\mathbf{x}, \mathbf{y})$ is called periodic if $n_z(\phi) = 0$. For such a class we have $\mathcal{D}(\phi)$ is periodic. The set of periodic classes at $\mathbf{x}$ will be denoted by $\mathcal{P}_{\mathbf{x}}$.

Proposition 8.1. *If $g > 1$, then there is an isomorphism*

$$\pi_2(\boldsymbol{x}, \boldsymbol{x}) \cong \mathbb{Z} \oplus H^1(Y)$$

where $\mathcal{P}_{\boldsymbol{x}}$ is identified with $H^1(Y)$, and the splitting is given by $\phi \mapsto n_z(\phi)$. If $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$ such that $\epsilon(\boldsymbol{x}, \boldsymbol{y}) = 0$, then

$$\pi_2(\boldsymbol{x}, \boldsymbol{y}) \cong \mathbb{Z} \oplus H^1(Y)$$

as $\pi_2'(\mathrm{Sym}^g(\Sigma)) \times \mathcal{P}_{\boldsymbol{x}}$-torsors.

Spin^c Structures

We now introduce Spin^c structures into our set up. We use Turaev's formulation of Spin^c structures on 3-manifolds in terms of homology classes of unit vector fields. We say that two unit vector fields V_1 and V_2 are *homologous* if they are homotopic through unit vector fields in the complement of finitely many 3-balls in Y. A Spin^c structure on Y is a choice of homology class of non-vanishing vector fields, and the set of Spin^c structures on Y is denoted by Spin$^c(Y)$. Recall the following fact about orientable 3-manifolds.

Lemma 8.5. *All closed, orientable 3-manifolds are parallelizable.*

Proof. Let X be a closed, orientable 3-manifold. If we can find two linearly independent sections of TX, then TX can be split as a 2-plane bundle plus a line bundle, which must be orientable and hence trivial. The first obstruction to finding two linearly independent sections is the second Stiefel-Whitney class, which by the Wu formula is given by $w_2 = w_1^2 = 0$. The only other obstruction lives in the cohomology of Y with coefficients in $\pi_2(V_2(\mathbb{R}^3)) = 0$, hence vanishes. $\square$

Since TY is trivial we now fix an orthonormal trivialization τ. This trivialization gives us a one to one correspondence between unit vector fields V on Y and maps $f_V : Y \to S^2$. Fix a generator $\mu \in H^2(S^2)$ and define the map $\delta^\tau : \mathrm{Spin}^c(Y) \to H^2(Y)$ by $\delta^\tau(V) = f_V^*(\mu)$.

Lemma 8.6. *δ^τ is well-defined and is a one-to-one correspondence.*

Proof. First we need to show that if V_1 is homologous to V_2, then $\delta^\tau(V_1) = \delta^\tau(V_2)$. Let $f_{V_1}, f_{V_2} : Y \to S^2$ be the maps corresponding to V_1 and V_2. By assumption these maps are homotopic on the complement of 3-ball in Y. **UNFINISHED** $\square$

If V_1 and V_2 are unit vector fields on Y, then we can look at the difference

$$\delta(V_1, V_2) = \delta^\tau(V_1) - \delta^\tau(V_2) \in H^2(Y)$$

We have removed the trivialization τ from the notation because this difference does not depend on it. For if τ_1, τ_2 are two trivializations of TY, then there is a map $g : Y \to SO(3)$ such that $\tau_1(V_p) = g(p) \cdot \tau_2(V_p)$ for all vectors V_p in T_pY. So, if we take two trivializations $\tau, g \cdot \tau$, then the difference satisfies

$$\delta^{g \cdot \tau}(V) - \delta^\tau(V) = g^*(w)$$

where w is the generator of $H^2(SO(3); \mathbb{Z}/2) = \mathbb{Z}/2$, so the difference does not depend on trivializations. So, for fixed $V \in \mathrm{Spin}^c(Y)$ the one-to-one correspondence $\delta(v, -) : \mathrm{Spin}^c(Y) \to H^2(Y)$ makes $\mathrm{Spin}^c(Y)$ into an affine space over $H^2(Y)$. For $V_1, V_2 \in \mathrm{Spin}^c(Y)$ we will write $V_1 - V_2$ as a short hand for $\delta(V_1, V_2)$.

Now let f be a Morse function on Y compatible with our Heegaard diagram Recall that an intersection point $\mathbf{x} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$ determined a g-tuple of flow lines from index 2 to index 1 critical points, which we denoted by $\gamma_{\mathbf{x}}$. The base point z has a unique flow line going through it from the index 3 critical point to the index 0 critical point. If we delete tubular neighborhoods of these $g + 1$ flow lines from Y we get a space with boundary consisting of $g + 1$ spheres such that $\nabla_g f$ does not vanish. The restriction of $\nabla_g f$ to a boundary component is an index 0 vector field on a sphere, and hence can be extended to the ball it bounds. Extending $\nabla_g f$ to the interior of these spheres, and normalizing so that the vector field is unit length, we define $s_z(\mathbf{x})$ to be the Spin^c structure this vector field defines. This gives us a map

$$s_z : \mathbb{T}_\alpha \cap \mathbb{T}_\beta \longrightarrow \mathrm{Spin}^c(Y)$$

Lemma 8.7. *For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$ we have $s_z(\boldsymbol{y}) - s_x(\boldsymbol{x}) = \mathrm{PD}(\epsilon(\boldsymbol{x}, \boldsymbol{y}))$.*

Therefore $s_z(\mathbf{x}) = s_z(\mathbf{y})$ if and only if $\mathbf{x}$ can be connected to $\mathbf{y}$ by a Whitney disc. If $\mathfrak{s} \in \mathrm{Spin}^c(Y)$ is represented by a vector field V, then we let $\bar{\mathfrak{s}}$ be the Spin^c structure represented by $-V$. The first Chern class of Spin^c structures is a map $c_1 : \mathrm{Spin}^c(Y) \to H^2(Y)$ defined by $c_1(\mathfrak{s}) = \mathfrak{s} - \bar{\mathfrak{s}}$. This is just the first Chern class of the 2-plane bundle over Y given as the orthogonal complement of a vector field V representing $\mathfrak{s}$.

Analytic Set-Up of the Moduli Spaces

Our definition of the moduli space of pseudoholomorphic representatives is not going to be sufficient in order to ensure that the spaces $\mathcal{M}(\phi)$ are smooth manifolds. We need to introduce suitable perturbations to the objects involved, which we do now. The quotient map

$$\pi : \Sigma^{\times g} \to \mathrm{Sym}^g(\Sigma)$$

is an honest covering map over $\mathrm{Sym}^g(\Sigma) - D$. Let us fix a Kähler structure (j, η) on the surface Σ; that is, j is complex structure on Σ and η is a non-degenerate 2-form on Σ such that j tames η. The 2-form $\omega_0 = \eta^{\times g}$ on $\Sigma^{\times g}$ induces a Kähler form $\pi_*(\omega_0)$ on $\mathrm{Sym}^g(\Sigma) - D$. Let us fix a collection of points $\{z_i\} \subset \Sigma - \alpha_1 - \cdots - \alpha_g - \beta_1 - \cdots - \beta_g$, with one z_i in each component, and an open set $V \subset \mathrm{Sym}^g(\Sigma)$ with

$$\big(\{z_i\} \times \mathrm{Sym}^{g-1}(\Sigma) \cup D\big) \subset V \subset \mathrm{Sym}^g(\Sigma)$$

and such that the closure $\overline{V}$ does not intersect the tori $\mathbb{T}_\alpha$ and $\mathbb{T}_\beta$. We say that an almost complex structure J on $\mathrm{Sym}^g(\Sigma)$ is (j, η, V)-*nearly symmetric* if J tames $\pi_*(\omega_0)$ over $\mathrm{Sym}^g(\Sigma) - \overline{V}$, and J

agrees with $\text{Sym}^g(\mathfrak{j})$ over V. The space of $(\mathfrak{j}, \eta, V)$-nearly symmetric, almost complex structures will be denoted by $\mathcal{J}(\mathfrak{j}, \eta, V)$. The tameness condition gives us that the tori $\mathbb{T}_\alpha$ and $\mathbb{T}_\beta$ are totally real, which is used for energy bounds and Gromov compactness.

We will now describe a more convenient model for Whitney discs. Since the unit disc in $\mathbb{C}$ with two boundary points removed can be holomorphically identified with the infinite strip $[0,1] \times i\mathbb{R}$, we redefine $\mathbb{D}$ to be this infinite strip. Then $\pi_2(\mathbf{x}, \mathbf{y})$ consists of the homotopy classes of maps u from $\mathbb{D}$ into $\text{Sym}^g(\Sigma)$ such that $u(1 \times \mathbb{R}) \subset \mathbb{T}_\alpha$, $u(0 \times \mathbb{R}) \subset \mathbb{T}_\beta$, and $u(s - i\infty) = \mathbf{x}$, $u(s + i\infty) = \mathbf{y}$. For a path $J : [0,1] \to \mathcal{J}(\mathfrak{j}, \eta, V)$ of nearly symmetric, almost complex structures on $\text{Sym}^g(\Sigma)$ and $\phi \in \pi_2(\mathbf{x}, \mathbf{y})$, we define the moduli space

$$\mathcal{M}_{J_s}(\phi) = \left\{ u \in \phi : \frac{\partial u}{\partial s} + J_s \frac{\partial u}{\partial t} = 0 \right\}$$

There is an $\mathbb{R}$ action on $\mathcal{M}_{j_s}(\phi)$ given by $(r \cdot u)(s,t) = u(s + i(t + r))$. We define the space of unparameterized pseudoholomorphic discs to be the quotient

$$\widehat{\mathcal{M}}_{J_s}(\phi) = \frac{\mathcal{M}_{J_s}(\phi)}{\mathbb{R}}$$

Lemma 8.8. *If $u \in \mathcal{M}_{J_s}(\phi)$ is a J_s-holomorphic disc, then $\mathcal{D}(u) \geq 0$.*

Proof. Since J_s is $(\mathfrak{j}, \eta, V)$-nearly symmetric, we have that J_s is an honest complex structure in a neighborhood of each $\{z_i\} \times \text{Sym}^{g-1}(\Sigma)$. Therefore the disc u must intersect these subvarieties non-negatively. $\qquad\square$

Now we can set up the necessary Banach manifolds and vector bundles to show that $\mathcal{M}_{J_s}(\phi)$ is a smooth manifold. For a vector bundle E over $[0,1] \times \mathbb{R}$ and let $p, \delta > 0$ be real numbers and $k \geq 0$ an integer. We can put a δ-weighted Sobolev norm on the space of sections of E defined by

$$\|\sigma\|_{W^{p,k}_\delta} = \sum_{j=0}^{k} \int_{[0,1] \times \mathbb{R}} \left\| \nabla^{(j)} \sigma(s + it) \right\| e^{\delta \tau(t)} \, ds \wedge dt$$

where $\tau : \mathbb{R} \to \mathbb{R}$ is some smooth function with $\tau(t) = |t|$ for $|t| \geq 1$. The space of sections for which this norm is finite is denoted by $W^{k,p}_\delta(E)$.

For $p > 2$ let $\mathcal{B}_\delta(\mathbf{x}, \mathbf{y})$ be the space of maps $u : [0,1] \times \mathbb{R} \to \text{Sym}^g(\Sigma)$ in $W^{p,1}_{loc}$ satisfying the following additional properties:

1. The curve u satisfies the boundary conditions $u(1 \times \mathbb{R}) \subset \mathbb{T}_\alpha, u(0 \times \mathbb{R}) \subset \mathbb{T}_\beta$.

2. The curve u converges to the points $\mathbf{x}$ and $\mathbf{y}$ as $t \mapsto \pm\infty$ exponentially in the sense that there is $T \gg 0$ and functions

$$\xi_- \in W^{p,1}_\delta([0,1] \times (-\infty, -T], T_\mathbf{x} \text{Sym}^g(\Sigma)) \quad \xi_+ \in W^{p,1}_\delta([0,1] \times [T, \infty), T_\mathbf{y} \text{Sym}^g(\Sigma))$$

   such that for all $t > T$ we have

$$u(s - it) = \exp_\mathbf{x}(\xi_-(s - it)) \quad u(s + it) = \exp_\mathbf{y}(\xi_+(s + it))$$

The space $\mathcal{B}_\delta(\mathbf{x}, \mathbf{y})$ is a Banach manifold modeled on $W^{p,1}_\delta([0,1] \times \mathbb{R}, \mathbb{R}^{2g})$, and the tangent space at $u \in \mathcal{B}_\delta(\mathbf{x}, \mathbf{y})$ is precisely

$$T_u \mathcal{B}_\delta(\mathbf{x}, \mathbf{y}) = \left\{ \xi \in W^{p,1}_\delta(u^* T \text{Sym}^g(\Sigma)) : \xi(1, t) \in \mathbb{T}_\alpha, \xi(0, t) \in \mathbb{T}_\beta \right\}$$

For form the Banach vector bundle $\pi_{\mathcal{E}} : \mathcal{E} \to \mathcal{B}_\delta(\mathbf{x}, \mathbf{y})$ defined by (at least as a set)

$$\mathcal{E} = \bigcup_{u \in \mathcal{B}_\delta(\mathbf{x}, \mathbf{y})} \{u\} \times L^p_\delta(u^* T \operatorname{Sym}^g(\Sigma))$$

which has characteristic fiber $L^p_\delta([0, 1] \times \mathbb{R}, \mathbb{R}^{2g})$. We define a section $F : \mathcal{B}_\delta(\mathbf{x}, \mathbf{y}) \to \mathcal{E}$ of this vector bundle by

$$F(u) = \left(u, \frac{\partial u}{\partial s} + J_s \frac{\partial u}{\partial t} \right)$$

Now we have $\mathcal{M}_{J_s}(\mathbf{x}, \mathbf{y})$ is $F^{-1}(0_{\mathcal{E}})$, where $0_{\mathcal{E}}$ is the zero section of the vector bundle. The vertical differential of this section at a curve $u \in F^{-1}(0_{\mathcal{E}})$ is of the form (in coordinates)

$$DF_u : W^{p,1}_\delta([0, 1] \times \mathbb{R}, \mathbb{R}^{2g}) \longrightarrow L^p_\delta([0, 1] \times \mathbb{R}, \mathbb{R}^{2g})$$

$$DF_u(v) = \frac{\partial v}{\partial s} + J_s \frac{\partial v}{\partial t} + (\nabla_v J(s)) \frac{\partial u}{\partial t}$$

This is a Fredholm operator, and so $\mathcal{M}_{J_s}(\mathbf{x}, \mathbf{y})$ is a finite dimensional submanifold of $\mathcal{B}_\delta(\mathbf{x}, \mathbf{y})$. In fact, the Fredholm index of DF_u acting on the unweighted space ($\delta = 0$) is precisely the Maslov index of u. Clearly the components of $\mathcal{B}_\delta(\mathbf{x}, \mathbf{y})$ are in correspondence with the homotopy classes in $\pi_2(\mathbf{x}, \mathbf{y})$, and the Fredholm index of DF_u is constant on each of these components.

Theorem 8.1. *Fix a Heegaard diagram $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta})$, with $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ curves intersecting transversally, and fix data $(\mathfrak{j}, \eta, V)$. Then for generic path J_s of $(\mathfrak{j}, \eta, V)$-nearly symmetric, almost complex structures we have $\mathcal{M}_{J_s}(\boldsymbol{x}, \boldsymbol{y})$ is a smooth manifold.*

Definition of $\widehat{HF}$

For simplicity we will restrict to the case $b_1(Y) = 0$ so that $H_1(Y) \cong H^2(Y)$ are finite.

Fix a pointed Heegaard diagram $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, z)$ for Y such that the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ curves intersect transversally, and choose a Spin^c structure $\mathfrak{s} \in \operatorname{Spin}^c(Y)$. We also fix a generic complex structure $\mathfrak{j}$ on Σ and a generic path J_s of nearly-symmetric almost complex structures on $\operatorname{Sym}^g(\Sigma)$. Let $\widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathfrak{s})$ be the $\mathbb{Z}/2$ vector spaces generated by all $\mathbf{x} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$ such that $s_z(\mathbf{x}) = \mathfrak{s}$. We give this vector space a relative $\mathbb{Z}$-grading by

$$\operatorname{gr}(\mathbf{x}, \mathbf{y}) = \mu(\phi) - 2n_z(\phi)$$

where ϕ is any homotopy class in $\pi_2(\mathbf{x}, \mathbf{y})$. Since any other disc $\phi' \in \pi_2(\mathbf{x}, \mathbf{y})$ must be of the form $\phi' = \phi * kS$ we have

$$\mu(\phi * kS) - 2n_z(\phi * kS) = \mu(\phi) + k\mu(S) - 2\left(n_z(\phi) + kn_z(S) \right) = \mu(\phi) - 2n_z(\phi) + 2k - 2k$$

therefore $\operatorname{gr}(\mathbf{x}, \mathbf{y})$ is independent of the ϕ chosen. We define the differential $\partial : \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathfrak{s}) \to \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathfrak{s})$ by

$$\partial \mathbf{x} = \sum_{\substack{\mathbf{y} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta \\ s_z(\mathbf{y}) = \mathfrak{s}}} \sum_{\substack{\phi \in \pi_2(\mathbf{x}, \mathbf{y}) \\ \mu(\phi) = 1 \\ n_z(\phi) = 0}} \#\widehat{\mathcal{M}}(\phi) \cdot \mathbf{y}$$

The second sum actually consists of at most one term since there is at most one Whitney disc $\phi \in \pi_2(\mathbf{x}, \mathbf{y})$ such that $\mu(\phi) = 1$ and $n_z(\phi) = 0$. Technically ∂ depends on J_s, and so we may sometimes write ∂_{J_s}. Note that if $\mathbf{y}$ has non-zero coefficient in $\partial \mathbf{x}$, then

$$\operatorname{gr}(\mathbf{x}, \mathbf{y}) = \mu(\phi) - 2n_z(\phi) = 1$$

so ∂ lowers the grading by 1.

Theorem 8.2. $\partial^2 = 0$

Now we have that $(\widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathfrak{s}), \partial)$ is a chain complex, and so we can define its homology groups, which we denote by $\widehat{HF}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathfrak{s})$. We also define

$$\widehat{HF}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \bigoplus_{\mathfrak{s} \in \mathrm{Spin}^c(Y)} \widehat{HF}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathfrak{s})$$

8.2 Invariance of Heegaard Floer Homology

8.2.1 Independence of Complex Structures

Theorem 8.3. *Let Y be a closed oriented 3-manifold with fixed $Spin^c$ structure $\mathfrak{s}$. Then the homology group $\widehat{HF}(Y, \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathfrak{s})$ is independent of the complex structure $\mathfrak{j}$ on Σ and the path J_s.*

Proof. Let us fix a complex structure $\mathfrak{j}$ on Σ and take two paths $J_{s,0}, J_{s,1}$ of $(\eta, \mathfrak{j}, V)$-nearly symmetric almost complex structures, and connect them via a path $J : [0,1] \times \mathbb{R} \to \mathcal{J}(\eta, \mathfrak{j}, V)$ such that $J_{s,t} = J_{s,0}$ for $t \leq 0$ and $J_{s,t} = J_{s,1}$ for $t \geq 1$. For a Whitney disc $\phi \in \pi_2(\mathbf{x}, \mathbf{y})$ we define the moduli space of time-dependent holomorphic representatives by

$$\mathcal{M}_{J_{s,t}}(\phi) = \left\{ u : [0,1] \times \mathbb{R} \to \mathrm{Sym}^g(\Sigma) : \begin{array}{l} u \in \phi \\ \frac{du}{ds} + J_{s,t}\frac{du}{dt} = 0 \end{array} \right\}$$

We define the map

$$\widehat{\Phi}_{J_{s,t}} : \left(\widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathfrak{s}), \widehat{\partial}_{J_{s,0}} \right) \longrightarrow \left(\widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathfrak{s}), \widehat{\partial}_{J_{s,1}} \right)$$

by

$$\widehat{\Phi}_{J_{s,t}}(\mathbf{x}) = \sum_{\substack{\mathbf{y} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta \\ s_z(\mathbf{y}) = \mathfrak{s}}} \sum_{\substack{\phi \in \pi_2(\mathbf{x}, \mathbf{y}) \\ \mu(\phi) = 0 \\ n_z(\phi) = 0}} \#\mathcal{M}_{J_{s,t}}(\phi) \cdot \mathbf{y}$$

This map is well-defined, so we have to show that it is a chain map and induces an isomorphism on homology. To see that $\widehat{\Phi}_{J_{s,t}}$ is a chain map we look at the boundary components of a compact 1-dimensional space, namely $\mathcal{M}_{J_{s,t}}(\psi)$ when $\mu(\psi) = 1$. Composing $\widehat{\Phi}_{J_{s,t}}$ with the boundary operators we see that

$$\widehat{\Phi}_{J_{s,t}} \circ \widehat{\partial}_{J_{s,0}}(\mathbf{x}) = \widehat{\Phi}_{J_{s,t}} \sum_{\substack{\mathbf{y} \in \mathbb{T}_\alpha \cap T_\beta \\ s_z(\mathbf{y}) = \mathfrak{s}}} \sum_{\substack{\phi \in \pi_2(\mathbf{x}, \mathbf{y}) \\ \mu(\phi) = 1 \\ n_z(\phi) = 0}} \#\widehat{\mathcal{M}}_{J_{s,0}}(\phi) \cdot \mathbf{y}$$

$$= \sum_{\substack{\mathbf{y}, \mathbf{z} \in \mathbb{T}_\alpha \mathbb{T}_\beta \\ s_z(\mathbf{y}) = s_z(\mathbf{z}) = \mathfrak{s}}} \sum_{\substack{\phi \in \pi_2(\mathbf{x}, \mathbf{y}) \\ \phi' \in \pi_2(\mathbf{y}, \mathbf{z}) \\ \mu(\phi) = 1 \\ \mu(\phi') = 0 \\ n_z(\phi) = n_z(\phi') = 0}} \#\widehat{\mathcal{M}}_{J_{s,0}}(\phi) \cdot \#\mathcal{M}_{J_{s,t}}(\phi') \cdot \mathbf{z}$$

$$\widehat{\partial}_{J_{s,1}} \circ \widehat{\Phi}_{J_{s,t}}(\mathbf{x}) = \widehat{\partial}_{J_{s,1}} \sum_{\substack{\mathbf{y} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta \\ s_z(\mathbf{y}) = \mathfrak{s}}} \sum_{\substack{\phi' \in \pi_2(\mathbf{x}, \mathbf{y}) \\ \mu(\phi') = 0 \\ n_z(\phi') = 0}} \#\mathcal{M}_{J_{s,t}}(\phi') \cdot \mathbf{y}$$

$$= \sum_{\substack{\mathbf{y}, \mathbf{z} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta \\ s_z(\mathbf{y}) = s_z(\mathbf{z}) = \mathfrak{s}}} \sum_{\substack{\phi' \in \pi_2(\mathbf{x}, \mathbf{y}) \\ \phi \in \pi_2(\mathbf{y}, \mathbf{z}) \\ \mu(\phi') = 0 \\ \mu(\phi) = 1 \\ n_z(\phi') = n_z(\phi) = 0}} \#\mathcal{M}_{J_{s,t}}(\phi') \cdot \#\widehat{\mathcal{M}}_{J_{s,1}}(\phi) \cdot \mathbf{z}$$

By Gromov's compactness we see that for $\psi \in \pi_2(\mathbf{x}, \mathbf{z})$ with $\mu(\psi) = 1$ we have that $\mathcal{M}_{J_{s,t}}(\psi)$ can be compactified to a 1-dimensional manifold $\overline{\mathcal{M}_{J_{s,t}}}(\psi)$ with boundary

$$\partial \overline{\mathcal{M}_{J_{s,t}}}(\psi) = \left(\coprod_{\phi' * \phi = \psi} \widehat{\mathcal{M}}_{J_{s,0}}(\phi) \times \mathcal{M}_{J_{s,t}}(\phi') \right) \coprod \left(\coprod_{\phi * \phi' = \psi} \mathcal{M}_{J_{s,t}}(\phi') \times \widehat{\mathcal{M}}_{J_{s,1}}(\phi) \right)$$

where the unions are over those homotopy classes of discs such that $\mu(\phi') = 0$ and $\mu(\phi) = 1$. Now it is clear that the number of boundary components of $\overline{\mathcal{M}_{J_{s,t}}}(\psi)$ (modulo 2) is precisely the coefficient of $\mathbf{z}$ in $(\widehat{\Phi}_{J_{s,t}} \circ \widehat{\partial}_{J_{s,0}} - \widehat{\partial}_{J_{s,1}} \circ \widehat{\Phi}_{J_{s,t}})(\mathbf{x})$. Therefore $\widehat{\Phi}_{J_{s,t}}$ is a chain map.

To show that the induced map on homology is an isomorphism we show that $\widehat{\Phi}_{J_{s,t}} \circ \widehat{\Phi}_{J_{s,1-t}}$ is chain homotopic to the identity map. To define the chain homotopy we take a path $J_{s,t,\tau}$ between the juxtaposition (in the t coordinate) $J_{s,1-t} * J_{s,t}$ path and the path J_s (constant in t). We define the moduli space of this path to be

$$\mathcal{M}_{J_{s,t,\tau}}(\phi) = \bigcup_{\tau \in [0,1]} \mathcal{M}_{J_{s,t}(\tau)}(\phi)$$

This is a manifold of dimension $\mu(\phi) + 1$, and it already has some boundary components corresponding to the $\tau = 0$ and $\tau = 1$ ends. In particular,

$$\partial \mathcal{M}_{J_{s,t,\tau}}(\phi) = \mathcal{M}_{J_{s,1-t}}(\phi) \times \mathcal{M}_{J_{s,t}}(\phi) \cup \mathcal{M}_{J_s}(\phi) \tag{8.1}$$

We define the proposed chain homotopy $\widehat{H}_{J_{s,t,\tau}} : \left(\widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathfrak{s}), \widehat{\partial}_{J_{s,0}} \right) \to \left(\widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathfrak{s}), \widehat{\partial}_{J_{s,0}} \right)$ by

$$\widehat{H}_{J_{s,t,\tau}}(\mathbf{x}) = \sum_{\substack{\mathbf{y} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta \\ s_z(\mathbf{y}) = \mathfrak{s}}} \sum_{\substack{\phi \in \pi_2(\mathbf{x}, \mathbf{y}) \\ \mu(\phi) = -1 \\ n_z(\phi) = 0}} \#\mathcal{M}_{J_{s,t,\tau}}(\phi) \cdot \mathbf{y}$$

Clearly this map is of degree -1. We want to check that the coefficient of $\mathbf{z}$ in $(\widehat{\partial}_{J_{s,0}} \circ \widehat{H}_{J_{s,t,\tau}} + \widehat{H}_{J_{s,t,\tau}} \circ \widehat{\partial}_{J_{s,0}} + \widehat{\Phi}_{J_{s,t}} \circ \widehat{\Phi}_{J_{s,1-t}} + \mathrm{id})(\mathbf{x})$ is zero by relating it to the number of boundary components in a compact 1-dimensional manifold. The 1-dimensional manifold $\mathcal{M}_{J_{s,t,\tau}}(\psi)$ (with $\psi \in \pi_2(\mathbf{x}, \mathbf{z})$ and $\mu(\psi) = 0$) can be compactified by adding broken flows. The boundary of this compactification has three forms: two of these correspond to the boundary components of $\mathcal{M}_{J_{s,t,\tau}}(\psi)$ as shown in (8.1), and now we get an additional form from once broken flows. These broken flows are of the form

$$\left(\coprod_{\phi' * \phi = \psi} \widehat{\mathcal{M}}_{J_{s,0}}(\phi) \times \mathcal{M}_{J_{s,t,\tau}}(\phi') \right) \coprod \left(\coprod_{\phi * \phi' = \psi} \mathcal{M}_{J_{s,t,\tau}}(\phi') \times \widehat{\mathcal{M}}_{J_{s,0}}(\phi) \right)$$

96

where the union is over homotopy classes of discs with $\mu(\phi) = 1$ and $\mu(\phi') = -1$. It is now clearly that the boundary components of the compactification of $\mathcal{M}_{J_{s,t,\tau}}(\psi)$ correspond to the coefficient of $\mathbf{z}$ in the chain homotopy formula, and hence is zero. $\qquad\square$

8.2.2 Independence of Isotopies of Attaching Circles

Now we need to check that the homology groups $\widehat{HF}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathfrak{s})$ are invariant under Heegaard moves in our diagram $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta})$. Due to the construction of the chain complex $\widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathfrak{s})$ we can only consider *pointed* moves (those moves which never cross the fixed base point $z \in \Sigma$), so we need to make sure that pointed moves generate all moves.

Proposition 8.2. *Any two Heegaard diagrams $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, z)$ and $(\Sigma', \boldsymbol{\alpha}', \boldsymbol{\beta}', z')$ that determine the same 3-manifold can be transformed into each other via a finite sequence of pointed Heegaard moves.*

Proof. We know that we can connect the two diagrams with regular Heegaard moves, so we claim that each of these moves can be slightly altered to be a pointed move. Clearly stabilization can be made to miss the base point, and we can slightly isotope the attaching band used in making a handle slide to also miss the base point. So, we really need to only look at isotopies of the attaching circles.

Suppose we slightly isotoped the attaching circle α_1 to $\overline{\alpha}_1$, and in the process it crossed the base point z. If we surger out the rest of the $\boldsymbol{\alpha}$ curves we obtain a torus with $2g - 2$ marked points. In this torus α_1 can clearly be isotoped to $\overline{\alpha}_1$ without crossing the base point. However, each marked point α_1 crosses during this isotopy corresponds to a handleslide across the corresponding handle in Σ. Therefore we can replace the isotopy of α_1 to $\overline{\alpha}_1$ that crossed z with a sequence of handle slides and isotopies that missed z. $\qquad\square$

Theorem 8.4. *Let Y be a closed, oriented 3-manifold equipped with two Heegaard diagrams $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, z)$ and $(\Sigma, \boldsymbol{\alpha}', \boldsymbol{\beta}', z)$ such that one can be obtained from the other via a pointed isotopy of the attaching circles. Then*
$$\widehat{HF}(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}) \cong \widehat{HF}(\Sigma, \boldsymbol{\alpha}', \boldsymbol{\beta}')$$

Proof. If the two Heegaard diagrams are isotopic via pointed isotopies that keep the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ curves transverse at all times, then this is equivalent to varying the metric (or equivalently the complex structure) on Σ, which we does not change the homology groups by theorem 8.3.

So, we need to consider isotopies that introduce new intersection points between the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ curves. We need to do this in a controlled manner, and the way we do this is via exact Hamiltonian flows, so let us recall this machinery. Fix a Hamiltonian $H_t : X \to \mathbb{R}$ such that $H_t = 0$ for $t \notin [0, 1]$. This gives rise to a 1-parameter family of Hamiltonian vector fields X_t defined by

$$\omega(X_t, \cdot) = dH_t$$

Now this generates a 1-parameter family of diffeomorphisms $\Psi_t : X \to X$ such that

$$\frac{d\Psi_t}{dt} = X_t$$

Let us apply this to a special Hamiltonian. Take a smooth bump function $h : \Sigma \to \mathbb{R}$ such that $\operatorname{supp} h$ is a small neighborhood of a point on α_1 and let $f : \mathbb{R} \to [0, 1]$ be smooth with $\operatorname{supp} f \subseteq (0, 1)$. The define the Hamiltonian $H_t(x) = f(t)h(x)$ so that the corresponding 1-parameter family of diffeomorphisms Ψ_t moves a small piece of α_1 a little bit, but leaves everything else fixed.

We now use the isotopies to introduce a pair of canceling intersection points. Suppose α_1 is slightly isotoped via an exact Hamiltonian flow to intersect with β_1 transversally in an extra pair of canceling points. We can assume this isotopy is supported in $\Sigma \backslash (\alpha_2 \cup \cdots \cup \alpha_g \cup z)$. The isotopy Ψ_t of Σ induces an isotopy of the torus $\mathbb{T}_\alpha$ in $\mathrm{Sym}^g(\Sigma)$ in the obvious way.

We can now define Whitney discs with dynamic boundary conditions. In particular, a Ψ_t-Whitney disc from $\mathbf{x} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$ to $\mathbf{y} \in \Psi_1(\mathbb{T}_\alpha) \cap \mathbb{T}_\beta$ is a map $u : \mathbb{D} \to \mathrm{Sym}^g(\Sigma)$ satisfying (here we are thinking of $\mathbb{D}$ as $[0,1] \times \mathbb{R}$)

$$u(1 + it) \in \Psi_t(\mathbb{T}_\alpha), \text{ for all } t \in \mathbb{R}$$
$$u(0 + it) \in \mathbb{T}_\beta, \text{ for all } t \in \mathbb{R}$$
$$u(s - i\infty) = \mathbf{x}$$
$$u(s + i\infty) = \mathbf{y}$$

Let $\pi_2^{\Psi_t}(\mathbf{x}, \mathbf{y})$ denote the set of homotopy classes of such discs, and let $\mathcal{M}^{\Psi_t}(\phi)$ be the moduli space of pseudoholomorphic representatives of $\phi \in \pi_2^{\Psi_t}(\mathbf{x}, \mathbf{y})$. The reason we use these exact Hamiltonian flows to move the $\boldsymbol{\alpha}$ curves is that one can prove that we still have energy bounds on holomorphic discs, and so we can still form the Gromov compactifications.

We can now define a chain map

$$\Gamma^{\Psi_t} : \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \longrightarrow \widehat{CF}(\boldsymbol{\alpha}', \boldsymbol{\beta})$$

by the formula

$$\Gamma^{\Psi_t}(\mathbf{x}) = \sum_{\substack{\mathbf{y} \in \Psi_1(\mathbb{T}_\alpha) \cap \mathbb{T}_\beta}} \sum_{\substack{\phi \in \pi_2^{\Psi_t}(\mathbf{x}, \mathbf{y}) \\ n_z(\phi) = 0 \\ \mu(\phi) = 0}} \#\mathcal{M}^{\Psi_t}(\phi) \cdot \mathbf{y}$$

UNFINISHED $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

8.2.3 Independence of Handle Slides

In order to prove invariance under handle slides we need to introduce Whitney triangles and generalize all of our results about Whitney discs to triangles. A *Heegaard triple diagram of genus g* is a genus g surface Σ with three g-tuples $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$ of homologically independent curves in Σ. These curves determine genus g handlebodies $U_\alpha, U_\beta, U_\gamma$ and 3-manifolds $Y_{\alpha\beta} = U_\alpha \cup U_\beta, Y_{\beta\gamma} = U_\beta \cup U_\gamma, Y_{\alpha\gamma} = U_\alpha \cup U_\gamma$. This information specifies a 4-manifold $X_{\alpha\beta\gamma}$ with boundary $\partial X_{\alpha\beta} = Y_{\alpha\beta} \cup Y_{\beta\gamma} \cup Y_{\alpha\gamma}$ in the following way. Let Δ be the standard 2-simplex with vertices $v_\alpha, v_\beta, v_\gamma$ and edges $e_\alpha, e_\beta, e_\gamma$ where e_i connects v_j to v_k and $i \neq j, i \neq k$. If we thicken up Σ by $\Delta \times \Sigma$, then we can glue on the pieces $e_i \times U_i$ via

$$X_{\alpha\beta\gamma} = \frac{(\Delta \times \Sigma) \coprod (e_\alpha \times U_\alpha) \coprod (e_\beta \times U_\beta) \coprod (e_\gamma \times U_\gamma)}{(e_\alpha \times \Sigma) \sim (e_\alpha \times U_\alpha), (e_\beta \times \Sigma) \sim (e_\beta \times U_\beta), (e_\gamma \times \Sigma) \sim (e_\gamma \times U_\gamma)}$$

The corners of this manifold can naturally be smoothed out.

For $\mathbf{x} \in \mathbb{T}_\alpha \cap T_\beta, \mathbf{y} \in \mathbb{T}_\beta \cap \mathbb{T}_\gamma, \mathbf{z} \in \mathbb{T}_\alpha \cap \mathbb{T}_\gamma$, we say that a map $u : \Delta \to \mathrm{Sym}^g(\Sigma)$ is a Whitney triangle connecting $\mathbf{x}, \mathbf{y}, \mathbf{z}$ if $u(v_\gamma) = \mathbf{x}, u(v_\alpha) = \mathbf{y}, u(v_\beta) = \mathbf{z}$ and $u(e_\alpha) \subset \mathbb{T}_\alpha, u(e_\beta) \subset \mathbb{T}_\beta, u(e_\gamma) \subset \mathbb{T}_\gamma$. Let $\pi_2(\mathbf{x}, \mathbf{y}, \mathbf{z})$ denote the homotopy classes of such maps.

There is an obstruction in the first homology group of $X_{\alpha\beta\gamma}$ to the existence of Whitney triangles. Let $a, b, c : [0,1] \to \mathrm{Sym}^g(\Sigma)$ be any paths in $\mathbb{T}_\alpha, \mathbb{T}_\beta, \mathbb{T}_\gamma$ from $\mathbf{x}$ to $\mathbf{y}$, $\mathbf{y}$ to $\mathbf{z}$ and $\mathbf{z}$ to $\mathbf{z}$, respectively.

Then $a + b + c$ is a 1-cycle in $\text{Sym}^g(\Sigma)$. Its homology is not well-defined, but its image in the quotient

$$\frac{H_1(\text{Sym}^g(\Sigma))}{H_1(\mathbb{T}_\alpha) + H_1(\mathbb{T}_\beta) + H_1(\mathbb{T}_\gamma)} \cong H_1(X_{\alpha\beta\gamma})$$

is well-defined, and we denote it by $\epsilon(\mathbf{x}, \mathbf{y}, \mathbf{z})$. We have that $\pi_2(\mathbf{x}, \mathbf{y}, \mathbf{z})$ is non-empty if and only if $\epsilon(\mathbf{x}, \mathbf{y}, \mathbf{z}) = 0$. We can also define the map $n_z : \pi_2(\mathbf{x}, \mathbf{y}, \mathbf{z}) \to \mathbb{Z}$ to be the algebraic intersection number of a triangle with the subvariety $\{z\} \times \text{Sym}^{g-1}(\Sigma)$. The splicing action of discs can be extended to splice a disc and a triangle together. In particular, if $\phi \in \pi_2(\mathbf{x}, \mathbf{y})$ and $\psi \in \pi_2(\mathbf{y}, \mathbf{z}, \mathbf{w})$, then we can splice these together to get a triangle $\psi * \phi \in \pi_2(\mathbf{x}, \mathbf{z}, \mathbf{w})$. Clearly n_z is additive with respect to this splicing action: $n_z(\psi * \phi) = n_z(\psi) + n_z(\phi)$.

We can also extend the definition of domain of a Whitney disc to Whitney triangles. In particular, $\mathcal{D}(\psi)$ is a 2-chain in Σ consisting of the formal linear combinations of the regions $\Sigma \backslash (\boldsymbol{\alpha} \cup \boldsymbol{\beta} \cup \boldsymbol{\gamma})$. Such a linear combination is said to be a *triply-periodic* domain if the coefficient of the region containing z is zero and its boundary is a sum of $\boldsymbol{\alpha}, \boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ curves. The triple $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, z)$ is said to be *weakly admissible* if every non-zero triply-periodic domain has both positive and negative coefficients.

(**todo:** blah blah blah ... more to do on the above)

If $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})$ is a weakly-admissible Heegaard triple, then we construct the map

$$\widehat{f}_{\alpha\beta\gamma} : \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \otimes \widehat{CF}(\boldsymbol{\beta}, \boldsymbol{\gamma}) \longrightarrow \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\gamma})$$

by counting pseudoholomorphic triangles:

$$\widehat{f}_{\alpha\beta\gamma}(\mathbf{x} \otimes \mathbf{y}) = \sum_{\substack{\mathbf{z} \in \mathbb{T}_\alpha \cap \mathbb{T}_\gamma}} \sum_{\substack{\psi \in \pi_2(\mathbf{x}, \mathbf{y}, \mathbf{z}) \\ n_z(\psi) = 0 \\ \mu(\psi) = 0}} \#\mathcal{M}(\psi) \cdot \mathbf{z} \tag{8.2}$$

If we give $\widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \otimes \widehat{CF}(\boldsymbol{\beta}, \boldsymbol{\gamma})$ the differential $\partial(\mathbf{x} \otimes \mathbf{y}) = \partial\mathbf{x} \otimes \mathbf{y} + \mathbf{x} \otimes \partial\mathbf{y}$, then $\widehat{f}_{\alpha\beta\gamma}$ is a chain map. To verify this we need to show the following

$$\partial\widehat{f}_{\alpha\beta\gamma}(\mathbf{x} \otimes \mathbf{y}) + \widehat{f}_{\alpha\beta\gamma}(\partial\mathbf{x} \otimes \mathbf{y}) + \widehat{f}_{\alpha\beta\gamma}(\mathbf{x} \otimes \partial\mathbf{y}) = 0$$

for all $\mathbf{x} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$ and $\mathbf{y} \in \mathbb{T}_\beta \cap \mathbb{T}_\gamma$. This holds because there are three types of ends in the moduli space $\mathcal{M}(\psi)$ when $\mu(\psi) = 1$ that correspond to splicing a disc onto one of the three vertices of the triangle.

Let $\widehat{F}_{\alpha\beta\gamma}$ denote the induced map on homology. These maps satisfy an important associative law, which we state now. For this we take a Heegaard quadruple $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}, z)$ and so we get four maps $\widehat{f}_{\alpha\beta\gamma}, \widehat{f}_{\alpha\gamma\delta}, \widehat{f}_{\alpha\beta\delta}, \widehat{f}_{\beta\gamma\delta}$ from counting pseudoholomorphic triangles and their induced maps on homology $\widehat{F}_{\alpha\beta\gamma}, \widehat{F}_{\alpha\gamma\delta}, \widehat{F}_{\alpha\beta\delta}, \widehat{F}_{\beta\gamma\delta}$. These maps satisfy

$$\widehat{F}_{\alpha\gamma\delta}(\widehat{F}_{\alpha\beta\gamma}(- \otimes -) \otimes -) = \widehat{F}_{\alpha\beta\delta}(- \otimes \widehat{F}_{\beta\gamma\delta}(- \otimes -))$$

This is shown by constructing a chain homotopy between the various compositions via counting pseudoholomorphic rectangles. In particular, we can develop the theory of Whitney rectangles $u : \square \to \text{Sym}^g(\Sigma)$ in the obvious way, and define the set of homotopy classes of such rectangles by $\pi_2(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{w})$. If φ is a homotopy class of Whitney rectangles, then $\mathcal{M}(\varphi)$ will denote the space of pseudoholomorphic representatives. We can now define the map

$$\widehat{h}_{\alpha\beta\gamma\delta} : \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \otimes \widehat{CF}(\boldsymbol{\beta}, \boldsymbol{\gamma}) \otimes \widehat{CF}(\boldsymbol{\gamma}, \boldsymbol{\delta}) \longrightarrow \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\delta})$$

by counting pseudoholomorphic rectangles:

$$\widehat{h}_{\alpha\beta\gamma\delta}(\mathbf{x} \otimes \mathbf{y} \otimes \mathbf{z}) = \sum_{\substack{\mathbf{w} \in \mathbb{T}_\alpha \cap \mathbb{T}_\delta}} \sum_{\substack{\varphi \in \pi_2(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{w}) \\ n_z(\varphi)=0 \\ \mu(\varphi)=0}} \#\mathcal{M}(\varphi) \cdot \mathbf{w}$$

We claim this map determines a chain homotopy between the maps $\widehat{f}_{\alpha\gamma\delta}(\widehat{f}_{\alpha\beta\gamma}(- \otimes -) \otimes -)$ and $\widehat{f}_{\alpha\beta\delta}(- \otimes \widehat{f}_{\beta\gamma\delta}(- \otimes -))$; that is, we need to show that

$$\partial \widehat{h}_{\alpha\beta\gamma\delta}(\mathbf{x} \otimes \mathbf{y} \otimes \mathbf{z}) + \widehat{h}_{\alpha\beta\gamma\delta}(\partial(\mathbf{x} \otimes \mathbf{y} \otimes \mathbf{z})) + \widehat{f}_{\alpha\gamma\delta}(\widehat{f}_{\alpha\beta\gamma}(\mathbf{x} \otimes \mathbf{y}) \otimes \mathbf{z}) + \widehat{f}_{\alpha\beta\delta}(\mathbf{x} \otimes \widehat{f}_{\beta\gamma\delta}(\mathbf{y} \otimes \mathbf{z})) = 0$$

Note that $\partial(\mathbf{x} \otimes \mathbf{y} \otimes \mathbf{z}) = \partial\mathbf{x} \otimes \mathbf{y} \otimes \mathbf{z} + \mathbf{x} \otimes \partial\mathbf{y} \otimes \mathbf{z} + \mathbf{x} \otimes \mathbf{y} \otimes \partial\mathbf{z}$, so the above expands to

$$\partial \widehat{h}_{\alpha\beta\gamma\delta}(\mathbf{x} \otimes \mathbf{y} \otimes \mathbf{z}) + \widehat{h}_{\alpha\beta\gamma\delta}(\partial\mathbf{x} \otimes \mathbf{y} \otimes \mathbf{z}) +$$
$$\widehat{h}_{\alpha\beta\gamma\delta}(\mathbf{x} \otimes \partial\mathbf{y} \otimes \mathbf{z}) + \widehat{h}_{\alpha\beta\gamma\delta}(\mathbf{x} \otimes \mathbf{y} \otimes \partial\mathbf{z}) +$$
$$\widehat{f}_{\alpha\gamma\delta}(\widehat{f}_{\alpha\beta\gamma}(\mathbf{x} \otimes \mathbf{y}) \otimes \mathbf{z}) + \widehat{f}_{\alpha\beta\delta}(\mathbf{x} \otimes \widehat{f}_{\beta\gamma\delta}(\mathbf{y} \otimes \mathbf{z})) = 0$$

The coefficient of $\mathbf{w}$ in these six terms correspond to the six types of ends of the moduli space $\mathcal{M}(\varphi)$, where $\varphi \in \pi_2(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{w})$ is a rectangle with $\mu(\varphi) = 1$. In particular, the first four terms correspond to splicing a disc onto one of the four corners of the rectangle, and the last two terms correspond to the rectangle degenerating into two triangles meeting at a vertex in two different ways.

We could continue these constructions to discuss pseudoholomorphic m-gons, and their associative laws, but now we will discuss the handle slide invariance of Floer homology. First we look at two model calculations that will be used in constructing the isomorphisms.

Example 8.1. . Take the standard Heegaard diagram for $\#gS^1 \times S^2$: $\Sigma = E\# \cdots \#E$ is the g-fold connected sum of the torus E, and α_i, β_i are curves supported in the ith factor that bound discs in the solid genus g handlebody (i.e. they are meridians) and intersect transversally in a pair of canceling points. Let $z \in \Sigma$ be a base point far from the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ curves. Orient these curves in the same direction and label the positive and negative intersection points of α_i and β_i by $w_i^{\pm}$.

We claim that, up to grading shifts, $\widehat{HF}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \cong H_*(T^g)$. There are 2^g intersect points in $\mathbb{T}_\alpha \cap \mathbb{T}_\beta$, and we can write these as ordered pairs $\mathbf{x} = (x_1, \ldots, x_g)$ such that $x_i = w_i^+$ or w_i^-. In particular, for an integer $1 \leq i \leq g$, take $\mathbf{x}, \mathbf{y} \in \mathbb{T}_\alpha \cap T_\beta$ such that $x_i = w_i^+, y_i = w_i^-$ and $x_j = y_j$ for all $j \neq i$. The curves α_i and β_i bound two discs in Σ, call them D_{1i} and D_{2i}, and let $\phi_1, \phi_2 \in \pi_2(\mathbf{x}, \mathbf{y})$ be discs with $\mathcal{D}(\phi_1) = D_{1i}$ and $\mathcal{D}(\phi_2) = D_{2i}$. It is clear that $\mu(\phi_1) = \mu(\phi_2) = 1$. Clearly, if $\phi_1', \phi_2' \in \pi_2(\mathbf{y}, \mathbf{x})$ are discs with $\mathcal{D}(\phi_1') = -D_{1i}$ and $\mathcal{D}(\phi_2') = -D_{2i}$, then $\mu(\phi_1') = \mu(\phi_2') = -1$.

Now for any $\mathbf{x}, \mathbf{y} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$ and $\phi \in \pi_2(\mathbf{x}, \mathbf{y})$ with $n_z(\phi) =$ we can find the precise value of $\mu(\phi)$. Recall that we can think of these discs as a products of the standard discs from x_i to y_i in Σ. Let $p(\mathbf{x})$ denote the number of w_i^+'s in $\mathbf{x}$. Then of course we can assume $x_i = w_i^+$ and $y_i = w_i^+$ for $i \leq p(\mathbf{x})$, hence we have $\mu(\phi) = p(\mathbf{x}) - p(\mathbf{y})$. This gives us the relative gradings, $\mathrm{gr}(\mathbf{x}, \mathbf{y}) = p(\mathbf{x}) - p(\mathbf{y})$, hence the generators are partitioned into g gradings. If we give the lowest relative grading an absolute grading of 0, then we see that there are precisely $\binom{g}{i}$ generators in grading i.

Finally, we claim that $\partial = 0$, which completes the proof of our claim. For any $\mathbf{x} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$ let $\mathbf{y}$ be a generator with $\mathrm{gr}(\mathbf{x}, \mathbf{y}) = 1$. If $\phi \in \pi_2(\mathbf{x}, \mathbf{y})$ with $n_z(\phi) = 0$, then $1 = \mathrm{gr}(\mathbf{x}, \mathbf{y}) = \mu(\phi) = p(\mathbf{x}) - p(\mathbf{y})$, hence $\mathbf{x}$ has one more $+$ coordinate than $\mathbf{y}$. This means there is an i for which

$x_i = w_i^+, y_i = w_i^-$ and $x_j = y_j$ for $j \neq i$. As we saw before there are exact two discs connecting these points, and $\#\widehat{\mathcal{M}} = 1$ for both, hence $\mathbf{y}$ has zero coefficient in $\partial \mathbf{x}$.

The generator of $\widehat{HF}(\boldsymbol{\alpha}, \boldsymbol{\beta}, z)$ with highest grading is represented by the cycle $\widehat{\Theta}_{\alpha\beta} = \{w_1^+, \ldots, w_g^+\}$. We will use $\widehat{\Theta}_{\alpha\beta}$ to denote both the cycle and the homology class. $\qquad \square$

Example 8.2. . We take another Heegaard diagram of $\#gS^1 \times S^2$, but this time with one handle slid across another. In particular, let $\Sigma = E_1 \# \cdots \# E_g$, with each E_i a torus, and let α_i be a curve supported in E_i going once around the meridian of E_i. For $i > 1$ let γ_i be a curve supported in E_i going once around the meridian and intersecting α_i in a pair of canceling points. Finally, let γ_1 be the curve supported in $E_1 \# E_2$ that is the result of sliding α_1 across γ_2, and perturbed so that it intersects α_1 in a pair of canceling points and does not intersect either α_2 or γ_2. Again we can orient the intersection points of the $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ curves with $w_i^{\pm}$.

We claim that, up to grading shifts, $\widehat{HF}(\boldsymbol{\alpha}, \boldsymbol{\gamma}) \cong H_*(T^g)$.

UNFINISHED $\qquad \square$

With these model calculations out of the way we can now prove handle slide invariance. Fix a Heegaard diagram $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta})$ for Y, let $\boldsymbol{\gamma}$ be a g-tuple of attaching circles such that γ_1 is a handle slide of β_1 across β_2 and γ_i $(i > 1)$ is a small perturbation of β_i intersecting in a pair of canceling points, and $\boldsymbol{\delta}$ is another g-tuple of attaching circles that are small perturbations of the $\boldsymbol{\beta}$ curves such that they intersect at a pair of canceling points. Then the Heegaard diagrams $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}), (\Sigma, \boldsymbol{\alpha}, \boldsymbol{\gamma}), (\Sigma, \boldsymbol{\alpha}, \boldsymbol{\delta})$ all describe Y, and the diagrams $(\Sigma, \boldsymbol{\beta}, \boldsymbol{\gamma}), (\Sigma, \boldsymbol{\gamma}, \boldsymbol{\delta}), (\Sigma, \boldsymbol{\beta}, \boldsymbol{\delta})$ all describe $\#gS^1 \times S^2$. Let $\widehat{\Theta}_{\beta\gamma}, \widehat{\Theta}_{\gamma\delta}, \widehat{\Theta}_{\beta\delta}$ denote the top grading generators of the Floer homology of the latter three diagrams. We claim that the map $f : \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \to \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\gamma})$ defined by $f = f_{\alpha\beta\gamma}(- \otimes \widehat{\Theta}_{\beta\gamma})$ is a quasi-isomorphism (let F denote the induced map on homology). We already know that the map $g : \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \to \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\delta})$ defined by $g = f_{\alpha\beta\delta}(- \otimes \widehat{\Theta}_{\beta\delta})$ is a quasi-isomorphism, and the associativity law gives us

$$
\begin{aligned}
F(\xi) &= F_{\alpha\beta\gamma}(\xi \otimes \widehat{\Theta}_{\beta\gamma}) \\
&= F_{\alpha\beta\gamma}(\xi \otimes F_{\beta\delta\gamma}(\widehat{\Theta}_{\beta\delta} \otimes \widehat{\Theta}_{\delta\gamma})) \\
&= F_{\alpha\delta\gamma}(F_{\alpha\beta\delta}(\xi \otimes \widehat{\Theta}_{\beta\delta}) \otimes \widehat{\Theta}_{\delta\gamma}))
\end{aligned}
$$

The last line is a composition of two isomorphisms, $F_{\alpha\delta\gamma}(- \otimes \widehat{\Theta}_{\delta\gamma})$ and $F_{\alpha\beta\delta}(- \otimes \widehat{\Theta}_{\beta\delta})$, which was checked in our model calculations, hence F is an isomorphism. This concludes the proof of invariance under handle slides.

8.2.4 Independence of Stabilizations

Theorem 8.5. *Let* $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, z)$ *be a pointed Heegaard diagram for* Y*, and let* $(\Sigma \# E, \boldsymbol{\alpha}', \boldsymbol{\beta}', z)$ *be a stabilization, where* E *is a torus and* $\boldsymbol{\alpha}' = \boldsymbol{\alpha} \cup \{\alpha_{g+1}\}, \boldsymbol{\beta}' = \boldsymbol{\beta} \cup \{\beta_{g+1}\}$ *with* $\alpha_{g+1}, \beta_{g+1}$ *homologically independent curves on* E *intersecting once. Then*

$$
\widehat{HF}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \cong \widehat{HF}(\boldsymbol{\alpha}, \alpha_{g+1}, \boldsymbol{\beta}, \beta_{g+1})
$$

Proof. We can arrange this stabilization such that the torus E is glued to Σ along a region contained in the component of $\Sigma \backslash (\boldsymbol{\alpha} \cup \boldsymbol{\beta})$ containing the base point z. Let $c \in \Sigma \# E$ be the intersection point of α_{g+1} and β_{g+1}, and let $\boldsymbol{\alpha}' = \boldsymbol{\beta} \cup \beta_{g+1}$ and $\boldsymbol{\alpha}' = \boldsymbol{\alpha} \cup \alpha_{g+1}$. Let z' be the corresponding base point in $\Sigma \# E$. We clearly have $\mathbb{T}_{\alpha'} \cap \mathbb{T}_{\beta'} = \mathbb{T}_\alpha \cap \mathbb{T}_\beta \cup \{c\}$, so this gives us an identification of the chain groups $\widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}, z)$ and $\widehat{CF}(\boldsymbol{\alpha}', \boldsymbol{\beta}', z')$. If $\mathbf{x}, \mathbf{y} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$ and $\mathbf{x}' = \mathbf{x} \times \{c\}, \mathbf{y}' = \mathbf{y} \times \{c\}$, then for

$\phi \in \pi_2(\mathbf{x}, \mathbf{y})$ with $n_z(\phi) = 0$ and $\phi' \in \pi_2(\mathbf{x}', \mathbf{y}')$ with $n_{z'}(\phi')$ we have an identification of the moduli spaces $\mathcal{M}(\phi') = \mathcal{M}(\phi) \times \{c\}$. Therefore the chain complexes are identical. $\qquad \square$

8.3 The Surgery Exact Sequence

For this section we mostly follow the presentations in [2] and [3]. Let k be a knot in a closed, oriented, integral homology 3-sphere Y. The knot determines a meridian μ in $\partial \nu k$, uniquely determined (up to isotopy) by the property that μ is not null-homologous in $\partial \nu k$ but is null-homologous in νk, and a longitude ℓ in $\partial \nu k$, uniquely determiend (up to isotopy) by the property that ℓ is null-homologous in $Y \backslash k$ and $\#(\ell \cap \mu) = -1$. Let Y_0 and Y_1 denote the 0- and +1-surgeries on k, i.e. we fill the torus $Y \backslash \nu k$ with a torus whose meridian is mapped to ℓ and $\mu + \ell$, respectively. The theorem we will work to prove is the following.

Theorem 8.6. *There is an exact sequence relating the Heegaard Floer homologies of Y, Y_0 and Y_1:*

$$\widehat{HF}(Y_0) \xrightarrow{\widehat{F}_0} \widehat{HF}(Y_1) \qquad (8.3)$$

with $\widehat{F}$ and $\widehat{F}_1$ forming a triangle with $\widehat{HF}(Y)$.

There is a considerable amount of machinery we need to set up before we can prove this. First we need to construct special Heegaard diagrams for Y, Y_0 and Y_1. In fact, we claim there is a pointed Heegaard multi-diagram $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}, z)$ such that

1. The Heegaard diagrams $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}), (\Sigma, \boldsymbol{\alpha}, \boldsymbol{\gamma}), (\Sigma, \boldsymbol{\alpha}, \boldsymbol{\delta})$ describe Y, Y_0, Y_1 respectively.

2. The curves $\beta_i, \gamma_i, \delta_i$, for $i = 1, \ldots, g-1$, are small isotopic translates of each other that pairwise, transversally intersect in a pair of canceling points.

3. The curve δ_g is isotopic to the juxtaposition of β_g and γ_g.

Note that the last two conditions imply that $(\Sigma, \boldsymbol{\beta}, \boldsymbol{\gamma}), (\Sigma, \boldsymbol{\gamma}, \boldsymbol{\delta}), (\Sigma, \boldsymbol{\beta}, \boldsymbol{\delta})$ are all Heegaard diagrams of $\#(g-1)S^1 \times S^2$. To construct these special Heegaard diagrams take a self-indexing Morse function f on $Y \backslash \nu k$ with one index 0 critical point, g index 1 critical points and $(g-1)$ index 2 critical points, and set $\Sigma = f^{-1}(3/2)$. Recall that the attaching circles have the following description. The $\boldsymbol{\alpha}$ curves are precisely where the ascending manifolds of index 1 critical points intersect Σ, and the $\boldsymbol{\beta}$ curves are where the descending manifolds of index 2 critical points intersect Σ. We fill the missing torus in $Y \backslash \nu k$ by adding a 2-handle in three different ways, and then adding the 3-handle in the unique way to obtain a closed manifold. First we attaching a 2-handle with attaching circle μ, thus obtaining Y. Next we attach a 2-handle with attaching circle ℓ, obtaining Y_0. Finally, we attach a 2-handle with attaching circle $\mu + \ell$, obtaining Y_1. Flowing these attaching circles down to the surface Σ gives us the circles $\beta_g, \gamma_g, \delta_g$ respectively. Let γ_i and δ_i be small isotopies of β_i (for $i = 1, \ldots, g-1$) so that these curves pairwise intersect transversally at a pair of canceling points. We have now constructed the Heegaard multi-diagram $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta})$ with the proposed properties.

Label the intersection points of the $\boldsymbol{\beta}, \boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ curves as in example 8.1

$$y_i^{\pm} = \beta_i \cap \gamma_i \qquad v_i^{\pm} = \gamma_i \cap \delta_i \qquad w_i^{\pm} = \beta_i \cap \delta_i$$

Also, the curves $\beta_g, \gamma_g, \delta_g$ pairwise intersect in exactly one point, so let

$$y_g = \beta_g \cap \gamma_g \qquad v_g = \gamma_g \cap \delta_g \qquad w_g = \beta_g \cap \delta_g$$

Recall the intersection points $\widehat{\Theta}_{\beta\gamma} = \left\{ y_1^+, \ldots, y_{g-1}^+, y_g \right\}, \widehat{\Theta}_{\gamma\delta} = \left\{ v_1^+, \ldots, v_{g-1}^+, v_g \right\}$ and $\widehat{\Theta}_{\beta\delta} = \left\{ w_1^+, \ldots, w_{g-1}^+, w_g \right\}$ are cycles in the top grading of $\widehat{HF}(\boldsymbol{\beta}, \boldsymbol{\gamma}), \widehat{HF}(\boldsymbol{\gamma}, \boldsymbol{\delta})$ and $\widehat{HF}(\boldsymbol{\beta}, \boldsymbol{\delta})$ respectively.

Recall that we have the following chain maps from counting pseudoholomorphic triangles

$$\widehat{f}_{\alpha\beta\gamma} : \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \otimes \widehat{CF}(\boldsymbol{\beta}, \boldsymbol{\gamma}) \longrightarrow \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\gamma})$$

$$\widehat{f}_{\alpha\gamma\delta} : \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\gamma}) \otimes \widehat{CF}(\boldsymbol{\gamma}, \boldsymbol{\delta}) \longrightarrow \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\delta})$$

$$\widehat{f}_{\alpha\beta\delta} : \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \otimes \widehat{CF}(\boldsymbol{\beta}, \boldsymbol{\delta}) \longrightarrow \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\delta})$$

$$\widehat{f}_{\beta\gamma\delta} : \widehat{CF}(\boldsymbol{\beta}, \boldsymbol{\gamma}) \otimes \widehat{CF}(\boldsymbol{\gamma}, \boldsymbol{\delta}) \longrightarrow \widehat{CF}(\boldsymbol{\beta}, \boldsymbol{\delta})$$

and their induced maps $\widehat{F}_{\alpha\beta\gamma}, \widehat{F}_{\alpha\gamma\delta}, \widehat{F}_{\alpha\beta\delta}, \widehat{F}_{\beta\gamma\delta}$ on homology. Let $\widehat{h}_{\alpha\beta\gamma\delta}$ be the chain homotopy between the two compositions $\widehat{f}_{\alpha\gamma\delta}(\widehat{f}_{\alpha\beta\gamma}(- \otimes -) \otimes -)$ and $\widehat{f}_{\alpha\beta\delta}(- \otimes \widehat{f}_{\beta\gamma\delta}(- \otimes -))$, which is given by counting pseudoholomorphic rectangles.

Now we can define the maps in the surgery exact sequence. We define

$$\widehat{F} = \widehat{F}_{\alpha\beta\gamma}(- \otimes \widehat{\Theta}_{\beta\gamma})$$
$$\widehat{F}_0 = \widehat{F}_{\alpha\gamma\delta}(- \otimes \widehat{\Theta}_{\gamma\delta})$$
$$\widehat{F}_1 = \widehat{F}_{\alpha\beta\delta}(- \otimes \widehat{\Theta}_{\beta\delta})$$

First we show that any composition of consecutive maps in this sequence is zero. If $\xi \in \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta})$, then one composition looks like

$$\widehat{F}_0 \circ \widehat{F}(\xi) = \widehat{F}_{\alpha\gamma\delta}(\widehat{F}_{\alpha\beta\gamma}(\xi \otimes \widehat{\Theta}_{\beta\gamma}) \otimes \widehat{\Theta}_{\gamma\delta}) = \widehat{F}_{\alpha\beta\delta}(\xi \otimes \widehat{F}_{\beta\gamma\delta}(\widehat{\Theta}_{\beta\gamma} \otimes \widehat{\Theta}_{\gamma\delta}))$$

where we used the associativity law in the last equality. So, if we can show that $\widehat{F}_{\beta\gamma\delta}(\widehat{\Theta}_{\beta\gamma} \otimes \widehat{\Theta}_{\gamma\delta}) = 0$, then we will have shown that this first composition is zero. We can do this directly by examining the Heegaard triple $(\Sigma, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta})$.

We think of this diagram as the first $g - 1$ torus summands in Σ having three meridian intersecting pairwise in a pair of canceling points $y_i^{\pm}, v_i^{\pm}, w_i^{\pm}$ (the curves being $\beta_i, \gamma_i, \delta_i$, $i = 1, \ldots, g-1$, of course), and in the last summand we have three straight curves intersecting pairwise in one point (the curves being $\beta_g, \gamma_g, \delta_g$). Let us examine each of these summands as a model calculation.

Suppose E is a torus with three curves β, γ, δ isotopic to a meridian and such that they pairwise intersect in a pair of intersecting points $y^{\pm}, v^{\pm}, w^{\pm}$ and with a base point z far from these curves. Then there is a unique homotopy class of triangles connecting y^+, v^+ and w^+ such that its domain is non-negative and $n_z = 0$.

Now suppose E is a torus with three curves β, γ, δ such that β is isotopic to the meridian, γ is isotopic to the longitude, δ is isotopic to $\beta + \gamma$, and they pairwise intersect at one point, denoted by y, v, w. Then for each integer $k > 0$ there are exactly two homotopy classes $\psi_k^{\pm}$ of triangles connecting y, v and w such that $n_z(\psi_k^{\pm}) = \frac{k(k-1)}{2}$. To this we lift to the universal cover of E, which is $\mathbb{R}^2$. The lifts of β, γ, δ become the lines $\{x = i\}, \{y = i\}$, and $\left\{y = x + i + \frac{1}{2}\right\}$, for $i \in \mathbb{Z}$. Then we can take ψ_k^+ to be the triangle bounded by $\left\{x = 0, y = 0, y = x + k + \frac{1}{2}\right\}$ and ψ_k^- to be the triangle bounded by $\left\{x = 0, y = 0, y = x - k + \frac{1}{2}\right\}$. Further, for each of these triangles there is a unique holomorphic representative, and so $\mu(\psi_k^{\pm}) = 0$.

Now let us go back to our original situation. We can think of $\Sigma = E_1 \# \cdots \# E_g$ as a sum of tori with the curves $\beta_i, \gamma_i, \delta_i$ supported in E_i. Then the situations in the first $g - 1$ summands of Σ is identical to the first example we examined above, and the situation in the last summand of Σ is

103

identical to the last example we examined above. Since the base point $z \in \Sigma$ is placed away from the curves we can see that there are exactly two homotopy classes $\psi^\pm$ of triangles from $\widehat{\Theta}_{\beta\gamma}, \widehat{\Theta}_{\gamma\delta}$ and $\widehat{\Theta}_{\beta\delta}$ such that $n_z(\psi^\pm) = 0 = \mu(\psi^\pm)$ and $\mathcal{D}(\psi^\pm) \geq 0$. These homotopy classes are obtained by taking the product of the first $g-1$ triangles between the points y_i^+, v_i^+, w_i^+, and then taking the product with one of the homotopy classes between y_g, v_g, w_g. Further, each of these homotopy classes admit the same number of pseudoholomorphic representatives, so $\#\mathcal{M}(\psi^+) = \#\mathcal{M}(\psi^-)$.

Consider the quantity $\widehat{f}_{\beta\gamma\delta}(\widehat{\Theta}_{\beta\gamma} \otimes \widehat{\Theta}_{\gamma\delta})$, which we saw above that if this quantity is zero, then we have $\widehat{F}_0 \circ \widehat{F} = 0$. Referring to (8.2) for the definition of $\widehat{f}_{\beta\gamma\delta}$ we see that only the coefficient of $\widehat{\Theta}_{\beta\delta}$ can possibly have a non-zero coefficient in $\widehat{f}_{\beta\gamma\delta}(\widehat{\Theta}_{\beta\gamma} \otimes \widehat{\Theta}_{\gamma\delta})$. However, the coefficient of this generator is the sum of the number of pseudoholomorphic triangles between $\widehat{\Theta}_{\beta\gamma}, \widehat{\Theta}_{\gamma\delta}, \widehat{\Theta}_{\beta\delta}$, which is $\equiv 0 \mod 2$. Therefore $\widehat{f}_{\beta\gamma\delta}(\widehat{\Theta}_{\beta\gamma} \otimes \widehat{\Theta}_{\gamma\delta}) = 0$.

We can use completely symmetrical arguments to show that the other two compositions $\widehat{F}_1 \circ \widehat{F}_0$ and $\widehat{F}_0 \circ \widehat{F}$ are both zero, therefore the surgery triangle is a chain complex. In order to show that this chain complex has trivial homology we will use a slightly technical framework in homological algebra. So, we take a moment to provide the necessary homological background for these constructions.

Homological Algebra Background

Let (A_1, ∂_1) and (A_2, ∂_2) be chain complexes of $\mathbb{Z}/2$ vector spaces. We are not assuming these complexes are graded even though it would be easy to introduce the proper grading conventions. If $f_1 : A_1 \to A_2$ is a chain map, then we can construct the **mapping cone** of f_1, denoted by $(M(f_1), \partial)$. This is the chain complex with chain group equal to $A_1 \oplus A_2$ and differential defined by the matrix

$$\partial = \begin{pmatrix} \partial_1 & 0 \\ f_1 & \partial_2 \end{pmatrix}$$

It is easy to check that $\partial^2 = 0$.

Proposition 8.3. *The obvious maps in the sequence*

$$0 \longrightarrow A_2 \overset{\iota}{\longrightarrow} M(f_1) \overset{\pi}{\longrightarrow} A_1 \longrightarrow 0$$

are chain maps and this sequence is exact. Further, the connecting morphism in the associated long exact sequence is precisely f_1.

Proof. The maps in the sequence are defined by $\iota(b) = (0, b)$ and $\pi(a, b) = a$, which clearly makes the sequence exact. It is easy to see that these maps are also chain maps. For the second part of the proposition let us recall how the connecting morphism is constructed in general.

Suppose we have a short exact sequence of chain complexes

$$0 \longrightarrow (A, \partial_A) \overset{f}{\longrightarrow} (B, \partial_B) \overset{g}{\longrightarrow} (C, \partial_C) \longrightarrow 0$$

We define a map $\delta : H(C, \partial_C) \to H(A, \partial_A)$ in the following way. If $c \in C$ such that $\partial_C c = 0$, then since g is surjective we can find $b \in B$ such that $g(b) = c$. But then $0 = \partial_C c = \partial_C g(b) = g(\partial_B b)$, and since $\operatorname{im} f = \ker g$ there is an element $a \in A$ such that $f(a) = \partial_B b$. Finally we see that $0 = \partial_B^2 b = \partial_B f(a) = f(\partial_A a)$, and since f is injective this implies that $\partial_A a = 0$. So, we define

$\delta([c]) = [a]$. It is easy to see that this is a well defined map and gives us an exact sequence

$$
\begin{array}{ccc}
H(A, \partial_A) & \xrightarrow{\quad f_* \quad} & H(B, \partial_B) \\
& & \\
& \delta \nwarrow \quad \swarrow g_* & \\
& H(C, \partial_C) &
\end{array}
$$

Now applying this construction to our mapping cone short exact sequence we see that for $a \in A_1$, $\delta([a])$ is defined by finding an element $(b, c) \in M(f_1)$ such that $a = \pi(b, c) = b$ (so $(a, 0)$ will do just fine), then finding an element $d \in A_2$ such that $(0, d) = \iota(d) = \partial(a, 0) = (\partial_1 a, f_1(a)) = (0, f_1(a))$ (so $d = f_1(a)$ will do just fine). Therefore $\delta([a]) = [f_1(a)]$, and this proves the proposition. $\square$

Proposition 8.4. *The mapping cone is functorial and satisfies a naturality condition. Specifically, if $f_1 : A_1 \to A_2$ and $g_1 : B_1 \to B_2$ are two chain maps between chain complexes, and we have maps ψ_1, ψ_2 such that the following diagram*

$$
\begin{array}{ccc}
A_1 & \xrightarrow{\ f_1\ } & A_2 \\
\psi_1 \downarrow & & \downarrow \psi_2 \\
B_1 & \xrightarrow[\ g_1\ ]{} & B_2
\end{array}
\tag{8.4}
$$

commutes up to homotopy, then there is an induced map $M(\psi_1, \psi_2) : M(f_1) \to M(g_1)$. Further, the short exact sequences associated to the cones $M(f_1)$ and $M(f_2)$ fit into a commutative (up to homotopy) diagram

$$
\begin{array}{ccccccccc}
0 & \longrightarrow & A_2 & \longrightarrow & M(f_1) & \longrightarrow & A_1 & \longrightarrow & 0 \\
& & \psi_2 \downarrow & & M(\psi_1, \psi_2) \downarrow & & \psi_1 \downarrow & & \\
0 & \longrightarrow & B_2 & \longrightarrow & M(g_1) & \longrightarrow & B_1 & \longrightarrow & 0
\end{array}
$$

Proof. We clearly want to define the induced map $M(\psi_1, \psi_2)$ by $M(\psi_1, \psi_2)(a_1, a_2) = (\psi_1(a_1), \psi_2(a_2))$. It is easy to see that this map has the necessary properties. $\square$

Proposition 8.5. *Let $\{(A_i, \partial_i)\}_{i \in \mathbb{Z}}$ be a collection of chain complexes and let $f_i : A_i \to A_{i+1}$ be a chain map for each $i \in \mathbb{Z}$ satisfying*

1. *The composition $f_{i+1} \circ f_i$ is chain nullhomotopic via $H_i : A_i \to A_{i+2}$*

2. *the map.*
$$\psi_i := f_{i+2} \circ H_i + H_{i+1} \circ f_i : A_i \to A_{i+3}$$

   is a quasi-isomorphism.

Then, $H(M(f_i)) \cong H(A_{i+2})$.

Proof. First we see that ψ_i is a chain map; that is, $\partial_{i+3} \circ \psi_i = \psi_i \circ \partial_i$. Messing with the compositions we easily see

$$
\begin{aligned}
\psi_i \partial_i &= f_{i+2} H_i \partial_i + H_{i+1} f_i \partial_i \\
&= f_{i+2} \partial_{i+2} H_i + f_{i+2} f_{i+2} f_i + H_{i+1} \partial_{i+1} f_i \\
&= \partial_{i+3} f_{i+2} H_i + \partial_{i+3} H_{i+1} f_i \\
&= \partial_{i+3} \psi_i
\end{aligned}
$$

We define a map $\varphi_i : M(f_i) \to A_{i+2}$ by $\varphi_i(a,b) = H_i(a) + f_{i+1}(b)$ and a map $\rho_i : A_i \to M(f_{i+1})$ by $\rho_i(a_i) = (f_i(a_i), H_i(a_i))$. It is easy to check that these maps are chain maps, and we have $\varphi_{i+1} \circ \rho_i = \psi_i$. Consider the following diagram

$$
\begin{array}{ccccccccc}
A_i & \xrightarrow{f_i} & A_{i+1} & \xrightarrow{\iota_{i+1}} & M(f_i) & \xrightarrow{\pi_{i+1}} & A_i & \xrightarrow{f_i} & A_{i+1} \\
\Big\| & & \Big\| & & \Big\downarrow{\varphi_i} & & \Big\downarrow{\psi_i} & & \Big\downarrow{\psi_{i+1}} \\
A_i & \xrightarrow{f_i} & A_{i+1} & \xrightarrow{f_{i+1}} & A_{i+2} & \xrightarrow{f_{i+2}} & A_{i+3} & \xrightarrow{f_{i+3}} & A_{i+4} \\
\Big\downarrow{\psi_i} & & \Big\downarrow{\psi_{i+1}} & & \Big\downarrow{\rho_{i+2}} & & \Big\| & & \Big\| \\
A_{i+3} & \xrightarrow{f_{i+3}} & A_{i+4} & \xrightarrow{\iota_{i+4}} & M(f_{i+3}) & \xrightarrow{\pi_{i+3}} & A_{i+3} & \xrightarrow{f_{i+3}} & A_{i+4}
\end{array}
$$

All the maps in this diagram are chain maps, and all but two squares are immediately seen to commute. For the other two we can easily define chain homotopies between the two different compositions, so all the squares commute up to homotopy. If we pass two homology we see that the top and bottom rows are exact, and the compositions of the columns (except for the middle composition) are isomorphisms, so it follows by the five lemma that $\rho_{i+2} \circ \varphi_i$ is also an isomorphism. The fact that $\varphi_i \circ \rho_{i-1} = \psi_{i-1}$ and $\rho_{i+2} \circ \varphi_i$ are quasi-isomorphisms implies that φ_i is an isomorphism, and hence ρ_i is also a quasi-isomorphism. Therefore $H(M(f_i)) \cong H(A_{i+2})$. $\qquad\square$

Back to the Surgery Exact Sequence

We will now see that proposition 8.5 implies the surgery exact sequence. To do this we take an infinite family of $\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}$ curves, denoted by $\boldsymbol{\beta}^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\delta}^{(i)}$, that are generic exact Hamiltonian perturbations of the $\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}$ curves, and assume $\boldsymbol{\beta}^{(0)} = \boldsymbol{\beta}, \boldsymbol{\gamma}^{(0)} = \boldsymbol{\gamma}, \boldsymbol{\alpha}^{(0)} = \boldsymbol{\alpha}$. Then let $A_{3i} = \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}^{(i)}), A_{3i+1} = \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\gamma}^{(i)})$ and $A_{3i+2} = \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\delta}^{(i)})$. Note that $A_{3i}, A_{3i+1}, A_{3i+2}$ are the Floer complexes for Y, Y_0 and Y_1 respectively. We define the maps $f_i : A_i \to A_{i+1}$ by counting pseudoholomorphic triangles as before. The fact that $f_{3i+1} \circ f_{3i}, f_{3i+2} \circ f_{3i+1}, f_{3i+3} \circ f_{3i+2}$ are chain null-homotopic comes from counting holomorphic rectangles, and the homotopies are given by

$$
H_{3i} := \widehat{h}_{\alpha\beta^{(i)}\gamma^{(i)}\delta^{(i)}}\left(- \otimes \widehat{\Theta}_{\beta^{(i)}\gamma^{(i)}} \otimes \widehat{\Theta}_{\gamma^{(i)}\delta^{(i)}}\right)
$$
$$
H_{3i+1} := \widehat{h}_{\alpha\gamma^{(i)}\delta^{(i)}\beta^{(i+1)}}\left(- \otimes \widehat{\Theta}_{\gamma^{(i)}\delta^{(i)}} \otimes \widehat{\Theta}_{\delta^{(i)}\beta^{(i+1)}}\right)
$$
$$
H_{3i+2} := \widehat{h}_{\alpha\delta^{(i)}\beta^{(i+1)}\gamma^{(i+1)}}\left(- \otimes \widehat{\Theta}_{\delta^{(i)}\beta^{(i+1)}} \otimes \widehat{\Theta}_{\beta^{(i+1)}\gamma^{(i+1)}}\right)
$$

So we need to verify the second hypothesis of proposition 8.5. First we show that A_i is quasi-isomorphic to A_{i+3} via a "closest point" map.

Proposition 8.6. *For a small perturbation $\boldsymbol{\beta}'$ of the $\boldsymbol{\beta}$ curves let $\widehat{\Theta}_{\beta\beta'}$ be the top-grading generator of $\widehat{HF}(\boldsymbol{\beta}, \boldsymbol{\beta}')$. Then the chain map*

$$
f : \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \longrightarrow \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}')
$$

defined by

$$
\xi \mapsto \widehat{f}_{\alpha\beta\beta'}(\xi \otimes \widehat{\Theta}_{\beta\beta'})
$$

induces an isomorphism on homology.

Proof. We assume that the perturbation is taken so that β_i and β_i' intersect in a pair of canceling points. The idea is to approximate $f = \widehat{f}_{\alpha\beta\beta'}(-\otimes\widehat{\Theta}_{\beta\beta'})$ with a map that is clearly an isomorphism. To do this we introduce $\mathbb{R}$-filtrations on $\widehat{CF}(\boldsymbol{\alpha},\boldsymbol{\beta})$ and $\widehat{CF}(\boldsymbol{\alpha},\boldsymbol{\beta}')$.

Let us fix an area for on Σ and an intersection point $\mathbf{x}_0 \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$, and define

$$\mathcal{F} : \mathbb{T}_\alpha \cap \mathbb{T}_\beta \longrightarrow \mathbb{R}$$

by

$$\mathcal{F}(\mathbf{x}) = \mathcal{A}(\mathcal{D}(\phi)) - n_z(\phi) \cdot \mathcal{A}(\Sigma)$$

where $\phi \in \pi_2(\mathbf{x},\mathbf{x}_0)$ and $\mathcal{A}(R)$ denotes the area of a region. This is clearly well-defined. If $b_1(Y_{\alpha\beta}) > 0$, then we need to require that $\mathcal{A}(P) = 0$ for any periodic domain, which is possibly to find such an area form for any weakly admissible diagram. We can also assume that the signed between the curves β_i and β_i' is zero. For $\xi, \xi' \in \widehat{CF}(\boldsymbol{\alpha},\boldsymbol{\beta})$ we will write $\xi < \xi'$ if

$$\max_{\substack{\mathbf{x}\in\mathbb{T}_\alpha\cap\mathbb{T}_\beta \\ \mathbf{x} \text{ has non-zero coefficient in } \xi}} \mathcal{F}(\mathbf{x}) < \max_{\substack{\mathbf{x}'\in\mathbb{T}_\alpha\cap\mathbb{T}_\beta \\ \mathbf{x}' \text{ has non-zero coefficient in } \xi'}} \mathcal{F}(\mathbf{x}')$$

We can repeat this construction to get a filtration on $\widehat{CF}(\boldsymbol{\alpha},\boldsymbol{\beta}')$ relative to a reference point $\mathbf{x}_0'$.

If the β_i' is sufficiently close to β_i, then for each $\mathbf{x} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$ there is a unique closest intersection point $\mathbf{x}' \in \mathbb{T}_\alpha \cap \mathbb{T}_{\beta'}$. This association sets up a bijection between the sets $\mathbb{T}_\alpha \cap \mathbb{T}_\beta$ and $\mathbb{T}_\alpha \cap \mathbb{T}_{\beta'}$, which we can extend to a linear isomorphism

$$\iota : \widehat{CF}(\boldsymbol{\alpha},\boldsymbol{\beta}) \longrightarrow \widehat{CF}(\boldsymbol{\alpha},\boldsymbol{\beta}')$$

Of course this map is not necessarily a chain map, but we claim that the chain map f can be written as $f = \iota + \ell$, where $\ell : \widehat{CF}(\boldsymbol{\alpha},\boldsymbol{\beta}) \to \widehat{CF}(\boldsymbol{\alpha},\boldsymbol{\beta}')$ is a linear map such that $\ell < \iota$. Since ι is an isomorphism, this implies that f is an isomorphism on the chain level, and since this map is a chain map we have that the chain complexes $\widehat{CF}(\boldsymbol{\alpha},\boldsymbol{\beta})$ and $\widehat{CF}(\boldsymbol{\alpha},\boldsymbol{\beta}')$ are isomorphic.

There is a unique triangle $\psi_0 \in \pi_2(\mathbf{x},\widehat{\Theta}_{\beta\beta'},\iota(\mathbf{x}))$ such that $\mathcal{D}(\psi_0) \geq 0$ and is supported in the region between β_i and β_i'. This triangle admits a unique holomorphic representative, so $\#\mathcal{M}(\psi_0) = 0$. We now define the map $\ell : \widehat{CF}(\boldsymbol{\alpha},\boldsymbol{\beta}) \to \widehat{CF}(\boldsymbol{\alpha},\boldsymbol{\beta}')$ by

$$\ell(\mathbf{x}) = \sum_\psi \#\mathcal{M}(\psi) \cdot \iota(\mathbf{x})$$

where the sum is over all $\psi \in \pi_2(\mathbf{x},\widehat{\Theta}_{\beta\beta'},\iota(\mathbf{x}))$ such that $n_z(\psi) = 0$ and $\mathcal{D}(\psi) \neq k\mathcal{D}(\psi_0)$ for some integer k (i.e. ψ is not supported in the domain of ψ_0. We clearly have $f = \iota + \ell$.
UNFINISHED $\square$

Proposition 8.6 gives us a quasi-isomorphism $\theta_i : A_i \to A_{i+3}$ obtained by counting triangles. We claim that θ_i is chain homotopic to $f_{i+2} \circ H_i + H_{i+1} \circ f_i : A_i \to A_{i+3}$. Let us see this for $i = 0$, and the general situation will follow easily. Let us write $\boldsymbol{\beta}'$ for the $\boldsymbol{\beta}^{(0)}$ curves for what follows. We have the following maps

$$\theta = f_{\alpha\beta\beta'}(-\otimes\widehat{\Theta}_{\beta\beta'}) : \widehat{CF}(\boldsymbol{\alpha},\boldsymbol{\beta}) \to \widehat{CF}(\boldsymbol{\alpha},\boldsymbol{\beta}')$$
$$f_0 = f_{\alpha\beta\gamma}(-\otimes\widehat{\Theta}_{\beta\gamma}) : \widehat{CF}(\boldsymbol{\alpha},\boldsymbol{\beta}) \to \widehat{CF}(\boldsymbol{\alpha},\boldsymbol{\gamma})$$
$$f_2 = f_{\alpha\delta\beta'}(-\otimes\widehat{\Theta}_{\delta\beta'}) : \widehat{CF}(\boldsymbol{\alpha},\boldsymbol{\delta}) \to \widehat{CF}(\boldsymbol{\alpha},\boldsymbol{\beta}')$$
$$H_0 = h_{\alpha\beta\gamma\delta}(-\otimes\widehat{\Theta}_{\beta\gamma}\otimes\widehat{\Theta}_{\gamma\delta}) : \widehat{CF}(\boldsymbol{\alpha},\boldsymbol{\beta}) \to \widehat{CF}(\boldsymbol{\alpha},\boldsymbol{\delta})$$
$$H_1 = h_{\alpha\gamma\delta\beta'}(-\otimes\widehat{\Theta}_{\gamma\delta}\otimes\widehat{\Theta}_{\delta\beta'}) : \widehat{CF}(\boldsymbol{\alpha},\boldsymbol{\gamma}) \to \widehat{CF}(\boldsymbol{\alpha},\boldsymbol{\beta}')$$

We want to show that $f_2 \circ H_0 + H_1 \circ f_0$ is chain homotopic to θ_0. This comes from the associative law for pseudoholomorphic rectangles. The associative law for rectangles states that for a Heegaard multi-diagram $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\beta}')$ the following map

$$
\begin{aligned}
G = \ & \widehat{h}_{\alpha\gamma\delta\beta'}(\widehat{f}_{\alpha\beta\gamma}(-\otimes-)\otimes-\otimes-) + \widehat{f}_{\alpha\delta\beta'}(\widehat{h}_{\alpha\beta\gamma\delta}(-\otimes-\otimes-)\otimes-) + \\
& \widehat{h}_{\alpha\beta\delta\beta'}(-\otimes f_{\beta\gamma\delta}(-\otimes-)\otimes-) + \widehat{f}_{\alpha\beta\beta'}(-\otimes\widehat{h}_{\beta\gamma\delta\beta'}(-\otimes-\otimes-)) + \\
& \widehat{h}_{\alpha\beta\gamma\beta'}(-\otimes-\otimes\widehat{f}_{\gamma\delta\beta'}(-\otimes-))
\end{aligned}
\tag{8.5}
$$

is null-homotopic as a map

$$
G : \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \otimes \widehat{CF}(\boldsymbol{\beta}, \boldsymbol{\gamma}) \otimes \widehat{CF}(\boldsymbol{\gamma}, \boldsymbol{\delta}) \otimes \widehat{CF}(\boldsymbol{\delta}, \boldsymbol{\beta}') \longrightarrow \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}')
$$

The null-homotopy is given by counting pseudoholomorphic pentagons. In particular, for the Heegaard multi-diagram $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\beta}')$ we define the map

$$
P_{\alpha\beta\gamma\delta\beta'} : \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \otimes \widehat{CF}(\boldsymbol{\beta}, \boldsymbol{\gamma}) \otimes \widehat{CF}(\boldsymbol{\gamma}, \boldsymbol{\delta}) \otimes \widehat{CF}(\boldsymbol{\delta}, \boldsymbol{\beta}') \longrightarrow \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}')
$$

by counting pseudoholomorphic pentagons

$$
P_{\alpha\beta\gamma\delta\beta'}(\mathbf{x} \otimes \mathbf{y} \otimes \mathbf{z} \otimes \mathbf{w}) = \sum_{\substack{\mathbf{p}\in\mathbb{T}_\alpha\cap\mathbb{T}_{\beta'}}} \sum_{\substack{\varphi\in\pi_2(\mathbf{x},\mathbf{y},\mathbf{z},\mathbf{w},\mathbf{p}) \\ n_z(\varphi)=0 \\ \mu(\varphi)=0}} \#\mathcal{M}(\varphi) \cdot \mathbf{p}
$$

We claim that $\partial \circ P_{\alpha\beta\gamma\delta\beta'} + P_{\alpha\beta\gamma\delta\beta'} \circ \partial$ is equal to the map G in (8.5). This follows from examining the ends of $\mathcal{M}(\varphi)$ for $\varphi \in \pi_2(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{w}, \mathbf{p})$ and $\mu(\varphi) = 1$. In particular there are two kinds of degenerations of a pentagon: a disc can break off from one of the five vertices making a pentagon spliced with a disc, or a triangle can break off from one of the vertices making a rectangle spliced with a triangle.

We can now construct the chain homotopy between $f_2 \circ H_0 + H_1 \circ f_0$ and θ_0. It is the map

$$
K : \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \longrightarrow \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}')
$$

defined by

$$
K = P_{\alpha\beta\gamma\delta\beta'}(-\otimes\widehat{\Theta}_{\beta\gamma}\otimes\widehat{\Theta}_{\gamma\delta}\otimes\widehat{\Theta}_{\delta\beta'})
$$

For $\xi \in \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ we have

$$
\partial P_{\alpha\beta\gamma\delta\beta'}(\xi\otimes\widehat{\Theta}_{\beta\gamma}\otimes\widehat{\Theta}_{\gamma\delta}\otimes\widehat{\Theta}_{\delta\beta'}) + P_{\alpha\beta\gamma\delta\beta'}(\xi\otimes\widehat{\Theta}_{\beta\gamma}\otimes\widehat{\Theta}_{\gamma\delta}\otimes\widehat{\Theta}_{\delta\beta'}) = G(\xi\otimes\widehat{\Theta}_{\beta\gamma}\otimes\widehat{\Theta}_{\gamma\delta}\otimes\widehat{\Theta}_{\delta\beta'})
$$

and by definition of G we have

$$
\begin{aligned}
G(\xi\otimes\widehat{\Theta}_{\beta\gamma}\otimes\widehat{\Theta}_{\gamma\delta}\otimes\widehat{\Theta}_{\delta\beta'}) = & \widehat{h}_{\alpha\gamma\delta\beta'}(\widehat{f}_{\alpha\beta\gamma}(\xi\otimes\widehat{\Theta}_{\beta\gamma})\otimes\widehat{\Theta}_{\gamma\delta}\otimes\widehat{\Theta}_{\delta\beta'}) + \widehat{f}_{\alpha\delta\beta'}(\widehat{h}_{\alpha\beta\gamma\delta}(\xi\otimes\widehat{\Theta}_{\beta\gamma}\otimes\widehat{\Theta}_{\gamma\delta})\otimes\widehat{\Theta}_{\delta\beta'}) + \\
& \widehat{h}_{\alpha\beta\delta\beta'}(\xi\otimes f_{\beta\gamma\delta}(\widehat{\Theta}_{\beta\gamma}\otimes\widehat{\Theta}_{\gamma\delta})\otimes\widehat{\Theta}_{\delta\beta'}) + \widehat{f}_{\alpha\beta\beta'}(\xi\otimes\widehat{h}_{\beta\gamma\delta\beta'}(\widehat{\Theta}_{\beta\gamma}\otimes\widehat{\Theta}_{\gamma\delta}\otimes\widehat{\Theta}_{\delta\beta'})) + \\
& \widehat{h}_{\alpha\beta\gamma\beta'}(\xi\otimes\widehat{\Theta}_{\beta\gamma}\otimes\widehat{f}_{\gamma\delta\beta'}(\widehat{\Theta}_{\gamma\delta}\otimes\widehat{\Theta}_{\delta\beta'}))
\end{aligned}
$$

We can easily see that the first term in this expansion is precisely $H_1 \circ f_0(\xi)$, the second term is precisely $f_2 \circ H_0(\xi)$, the third and fifth terms are zero, and finally we claim the fourth term is identified with $\theta_0(\xi)$. It clearly suffices to show that

$$
\widehat{h}_{\beta\gamma\delta\beta'}(\widehat{\Theta}_{\beta\gamma}\otimes\widehat{\Theta}_{\gamma\delta}\otimes\widehat{\Theta}_{\delta\beta'}) = \widehat{\Theta}_{\beta\beta'}
$$

This follows from a model calculation.

Proof of theorem 8.6. We now have all of the necessary machinery to prove the surgery exact sequence. By proposition 8.5 we have that the chain complex $M(f_0)$ is quasi-isomorphic to A_2. But, $M(f_0) = A_0 \oplus A_1 = \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \oplus \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\gamma})$ and $A_2 = \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\delta})$. The short exact sequence

$$0 \longrightarrow \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\gamma}) \longrightarrow M(f_0) \longrightarrow \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \longrightarrow 0$$

leads to a long exact sequence

$$\widehat{HF}(\boldsymbol{\alpha}, \boldsymbol{\gamma}) \longrightarrow H(M(f_0))$$
$$\searrow \qquad \swarrow$$
$$\widehat{HF}(\boldsymbol{\alpha}, \boldsymbol{\beta})$$

This is the surgery exact sequence. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

$\mathbb{Z}/2$-Gradings

We can introduce $\mathbb{Z}/2$-gradings in $\widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and the homology groups, and the surgery exact sequence is homogenous with respect to these gradings.

A Few Simple Applications

8.4 The Absolute $\mathbb{Q}$ Grading

If Y is a rational homology sphere, then the relative $\mathbb{Z}$ grading on $\widehat{HF}(Y)$ can be lifted to an absolute $\mathbb{Q}$ grading. Let W be the cobordism from S^3 to Y by adding m 2-handles to D^4.
 UNFINISHED

8.5 Heegaard Floer Knot Homology

The Construction

We can also construct an invariant of knots in 3-manifolds from Heegaard Floer homology. Let k be a knot in a closed, oriented 3-manifold Y. First we need to come up with a Heegaard diagram for Y that somehow encodes information about our knot k. To do this let us first fix a Heegaard diagram $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}_0)$ for the knot complement $Y \backslash \nu k$, where $\boldsymbol{\alpha}$ is a g-tuple of curves and $\boldsymbol{\beta}_0$ is a $(g-1)$-tuple of curves. If $\mu \subset \partial \nu k$ is a meridian of k, then $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta})$ is a Heegaard diagram for Y, where $\boldsymbol{\beta} = \boldsymbol{\beta}_0 \cup \{\mu\}$. Let us place to base points $z, w \in \Sigma$ very close to μ, but on opposite sides, and away from the rest of the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ curves. We can connect w to z with a small arc δ_1 that crosses μ transversally in one point, and we can connect z to w with an arc δ_2 supported in the complement of the $\boldsymbol{\alpha}$ curves in Σ (which is topologically just a disc). Then the curve $\delta_1 \cup \delta_2$ (as it sits in S^3) is ambiently isotopic to k, and if k is oriented it will inherit an orientation.

We will use the base point w to define the regular Floer chain complex $\widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}, w)$ (i.e. the differential counts discs that miss $w \times \text{Sym}^{g-1}(\Sigma)$), whose homology we know consists of a single copy of $\mathbb{Z}/2$. Recall that we can lift the relative grading on these Floer chain groups to *absolute* gradings on homology. In this absolutely graded theory, the Floer homology of S^3 is supported in degree 0 and the Floer homology of $S^1 \times S^2$ is supported in degrees $\pm\frac{1}{2}$. Let $\widehat{CF}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}, w)$ denote the subgroup of $\widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}, w)$ generated by homogeneous elements of $\widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}, w)$ of degree i.

Next, we use the base point z to define a filtration on the graded group $\widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}, w)$. Define the function $\mathcal{F} : (\mathbb{T}_\alpha \cap \mathbb{T}_\beta)^2 \to \mathbb{Z}$ acting on $\mathbf{x}, \mathbf{y} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta$ by

$$\mathcal{F}(\mathbf{x}, \mathbf{y}) = n_z(\phi) - n_w(\phi)$$

for any disc $\phi \in \pi_2(\mathbf{x}, \mathbf{y})$. This function can be lifted to a *unique* function $\mathcal{F} : \mathbb{T}_\alpha \cap \mathbb{T}_\beta \to \mathbb{Z}$ with the following two properties:

1. $\mathcal{F}(\mathbf{x}, \mathbf{y}) = \mathcal{F}(\mathbf{x}) - \mathcal{F}(\mathbf{y})$

2. $\#\left\{\mathbf{x} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta : \mathcal{F}(\mathbf{x}) = i\right\} \equiv \#\left\{\mathbf{x} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta : \mathcal{F}(\mathbf{x}) = -i\right\} \mod 2$

Finally we define the filtration $\mathcal{F}^*$ on $\widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}, w)$ by letting $\mathcal{F}^j \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}, w) \subseteq \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}, w)$ be the subgroup generated by $\mathbf{x}$ with $\mathcal{F}(\mathbf{x}) \leq j$. This filtration has the following nice properties:

1. The filtration is decreasing: $\mathcal{F}^j \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}, w) \subseteq \mathcal{F}^{j+1} \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}, w)$.

2. $\mathcal{F}^*$ is bounded: for large enough j we have $\mathcal{F}^j \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}, w) = \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}, w)$ and for small enough j we have $\mathcal{F}^j \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}, w) = 0$.

3. The filtration respects the absolute grading: $\mathcal{F}^j \widehat{CF}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}, w) \subseteq \widehat{CF}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}, w)$.

4. The differential $\partial : \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}, w) \to \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}, w)$ respects the filtration: $\partial\left(\mathcal{F}^j \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}, w)\right) \subseteq \mathcal{F}^j \widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}, w)$.

All of this information implies numerous things. First of all, we can construct the associated bigraded vector space

$$\widehat{CFK}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}, z, w, j) := \mathcal{F}^j \widehat{CF}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}, w) / \mathcal{F}^{j-1} \widehat{CF}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}, w)$$

The differential on $\widehat{CF}_*(\boldsymbol{\alpha}, \boldsymbol{\beta}, w)$ descends to a differential on $\widehat{CFK}_*(\boldsymbol{\alpha}, \boldsymbol{\beta}, z, w, *)$, and so we can define the homology of this complex, which is a bigraded vector space $\widehat{HFK}_*(\boldsymbol{\alpha}, \boldsymbol{\beta}, z, w, *)$. Further, by *the* fundamental theorem of spectral sequences, we have that there is a spectral sequence $\left\{E^r_{p,q}, d_r\right\}$ of homological type with $E^2 = \widehat{HFK}(\boldsymbol{\alpha}, \boldsymbol{\beta}, z, w)$ and converging to $E^\infty = \widehat{HF}(S^3)$. It turns out this entire spectral sequence is an invariant of the knot, and will be useful later for defining additional scalar invariants.

This bigraded homology group turns out to be a topological invariant of the knot k, and so we write $\widehat{HFK}_i(Y, k, j)$ to denote the homology groups.

Proposition 8.7. *If $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}_0, \mu)$ and $(\Sigma', \boldsymbol{\alpha}', \boldsymbol{\beta}'_0, \mu')$ represent the same knot complement, then one representation can be obtained from the other via a finite sequence of the following moves (and their inverses):*

1. *Handle slides and isotopies amongst the $\boldsymbol{\alpha}$ or $\boldsymbol{\beta}_0$.*

2. *Isotopy of μ.*

3. *Handle slides of μ across the $\boldsymbol{\beta}_0$ curves.*

4. *Stabilizations.*

Proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

Theorem 8.7. *For an oriented knot k in S^3 with compatible Heegaard diagram $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, z, w)$, the filtered chain homotopy type of $\widehat{CFK}(\boldsymbol{\alpha}, \boldsymbol{\beta}_0, z, w)$ is a topological invariant of k, i.e. it does not depend on the choice of Heegaard diagram and complex structures used in its construction.*

Proof. We proceed like in the proof of invariance of $\widehat{HF}$, except now we have to make sure that all of our maps respect the filtration on $\widehat{CFK}$. First consider a perturbation of the path of nearly-symmetric, almost complex structures $J_{s,0}$ and $J_{s,1}$. Early we defined a continuation map $\Phi_{J_{s,t}}$: $(\widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}), \partial_{J_{s,0}}) \to (\widehat{CF}(\boldsymbol{\alpha}, \boldsymbol{\beta}), \partial_{s,1})$ by counting time-dependent pseudoholomorphic discs. Since the filtrations on both of these complexes are precisely the same ....

$$\Phi_{J_{s,t}}(\mathbf{x}) = \sum_{\substack{\mathbf{y} \in \mathbb{T}_\alpha \cap \mathbb{T}_\beta}} \sum_{\substack{\phi \in \pi_2(\mathbf{x}, \mathbf{y}) \\ \mu(\phi) = 0 \\ n_z(\phi) = 0}} \#\mathcal{M}_{J_{s,t}}(\phi) \cdot \mathbf{y}$$

UNFINISHED

$\square$

Extension to Links

Now we explain how to extend the above construction to links. Let L be an m-component link in Y, and let $\{p_i, q_i\}_{i=1}^{m-1}$ be a set of points in L such that if we identify p_i with q_i in L we get a connected graph. We can think of these points as $(m-1)$-embedded 0-spheres. Let W be any null-cobordism of Y, and let $\kappa(Y)$ be the boundary of the 4-manifold obtained by attaching $(m-1)$ 1-handles whose attaching spheres are precisely $\{p_i, q_i\}$. Clearly $\kappa(Y)$ is diffeomorphic to $Y \# (m-1) S^1 \times S^2$. Then we can connect $(m-1)$ bands to L in $\kappa(Y)$ that start at p_i, end at q_i and run through the core of the i-th 1-handle, thus obtaining a knot $\kappa(L)$ in $\kappa(Y)$. The isotopy class of this knot in $\kappa(Y)$ depends only on the isotopy class of L in Y.

We now define the Floer homology of a link L in Y as the Floer homology of $\kappa(L)$ in $\kappa(Y)$, i.e. $\widehat{HFK}(Y, L) := \widehat{HFK}(\kappa(Y), \kappa(L))$. In order to get absolute gradings in $\widehat{HFK}$ for a link, recall the following absolute grading of the Floer homology of $\# m S^1 \times S^2$ (with $i \in \mathbb{Z}$):

$$\widehat{HF}_i(\# m S^1 \times S^2) = \begin{cases} (\mathbb{Z}/2)^{\binom{m}{i+m/2}} & m \text{ even and } -m/2 \leq i \leq m/2 \\ (\mathbb{Z}/2)^{\binom{m}{i+m/2-1/2}} & m \text{ odd and } -m/2 \leq i - \frac{1}{2} \leq m/2 \\ 0 & \text{otherwise} \end{cases}$$

That is, the Floer homology of $\# m S^1 \times S^2$ is the homology of the m-torus T^m with gradings shifted so that the homology is symmetric with respect to grading 0.

Exact Sequences for $\widehat{HFK}$

We can use the surgery exact sequence discussed from theorem 8.6 to derive exact sequences for the knot homology invariant. Fix an oriented knot k in S^3, and let γ be a knot in S^3 disjoint from k such that $\operatorname{lk}(k, \gamma) = 0$. The surgery exact sequence provides us with maps

$$\widehat{F} : \widehat{HF}(S^3) \to \widehat{HF}(S^3_0(\gamma))$$

$$\widehat{F}_0 : \widehat{HF}(S^3_0(\gamma)) \to \widehat{HF}(S^3_1(\gamma))$$

$$\widehat{F}_1 : \widehat{HF}(S^3_1(\gamma)) \to \widehat{HF}(S^3)$$

111

that fit into an exact triangle. Further, the map $\widehat{F}_0$ and $\widehat{F}_1$ are homogenous maps of degree $-1/2$, whereas the map $\widehat{F}$ is a sum of homogenous maps which do not increase the grading. It can be shown that the chain maps on $\widehat{CF}$ that induce the above maps respect the filtrations on $\widehat{CF}$, and hence the surgery exact sequence passes to an exact sequence in knot Floer homology:

$$\widehat{HFK}(S_0^3(\gamma),k) \xrightarrow{\widehat{F}_0} \widehat{HFK}(S_1^3(\gamma),k)$$

$$\widehat{F} \searrow \qquad \swarrow \widehat{F}_1$$

$$\widehat{HFK}(S^3,k)$$

Example 8.3. We can use the above surgery sequence for knot Floer homology to make some simple computations. Let B be the Borromean link in S^3, and let $B(m,n)$ be the knot in $L(m,1)\#L(n,1)$ obtained as one of the components of B after doing surgery on the other two components with framing m and n. For example, $B(m,\infty)$ is the unknot in $L(m,1)$ for any m, and $B(1,1)$ is the right-handed trefoil in S^3. We will compute later that the knot Floer homology of the right-handed trefoil is precisely

$$\widehat{HFK}_i(S^3, B(1,1), j) = \begin{cases} \mathbb{Z}/2 & (i,j) = (-2,-1), (-1,0), (0,1) \\ 0 & \text{otherwise} \end{cases}$$

Now we will compute $\widehat{HFK}(S^1 \times S^2, B(0,1))$. **UNFINISHED**

For an oriented link L in S^3, let L_+, L_-, L_0 denote the oriented links that are the same as L except in the vicinity of a crossing, in which case L_+, L_-, L_0 have that crossing changed to a positive, negative or oriented smoothing, respectively. If L has m components, then L_+ and L_- will also have m components, but L_0 will have $m+1$ components. In order to talk about the Floer homology of the links L_+ and L_- we look at their associated knots in the space $\#(m-1)S^1 \times S^2$, whereas the associated knot of the link L_0 lives in $\#mS^1 \times S^2$. Let γ be an unknot placed around a crossing of L such that L intersects the spanning disc of γ in a pair of canceling points, and both strands at this crossing belong to the same link component. The link L_- sitting inside the space $S_1^3(\gamma)$ is isotopic to L_+ sitting in S^3. We claim that L_- sitting in $S_0^3(\gamma) \cong S^1 \times S^2$ is isotopic to the link in $S^1 \times S^2$ obtained by attaching a 1-handle to S^3 near the smoothing in L_0, and then connecting the two components of L_0 near the smoothing with a compatibly oriented band going through the 1-handle.

An Euler Characteristic Calculation

The skein exact sequences for Floer knot homology allow us to compute the Euler characteristic of $\widehat{HFK}_{**}$, and compare it to a classical knot invariant.

Getting Heegaard Diagrams for a Certain Class of Knots

We will now explain how to genus 1 Heegaard diagrams for 2-bridge knots, and see that their Floer homology can be computed combinatorially. A 2-bridge knot is a knot k i S^3 such that there is a height function $f : S^3 \to \mathbb{R}$ such that $f|_k$ has precisely two maxima.
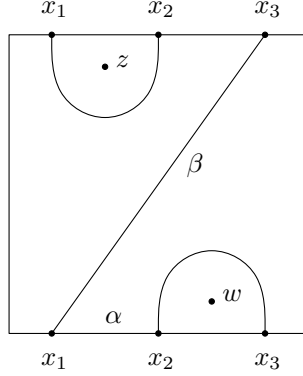
Figure 8.1: A Heegaard diagram for the trefoil knot

The Trefoil Knot

Figure 8.1 shows a Heegaard diagram for S^3 that is compatible with the left-handed trefoil. The intersection points of the α and β curves are labelled x_1, x_2 and x_3. Let us figure out the relative grading of these generators, computed with respect to the base point w. For $\mathrm{gr}(x_1, x_2)$ note that we have the obvious disc ϕ_{12} which does not intersect w and has a unique holomorphic representative (up to translation), hence $\mathrm{gr}(x_1, x_2) = 1$. For $\mathrm{gr}(x_2, x_3)$ we also have the obvious disc ϕ_{23}, but now the orientations are reversed, so $\mu(\phi_{23}) = -1$, and this disc intersects w negatively, hence $\mathrm{gr}(x_2, x_3) = -1 - 2(-1) = +1$. By additivity of gradings we have $\mathrm{gr}(x_1, x_3) = 2$, hence the gradings of the generators differ by one with x_1 in the top grading and x_3 in the bottom grading. To pin down the translation indeterminacy of this grading we compute the homology of the complex $(\widehat{CF}(\alpha\beta, w)$, which we know is $\mathbb{Z}/2$ supported in grading zero. In ∂x_1 we clearly have the coefficient of x_2 is one (we do not need to consider x_3 since its grading is too low and ∂ lowers gradings by only one). Next we have $\partial x_2 = 0$ since the disc connecting x_2 to x_3 does not support a holomorphic representative. Finally we have $\partial x_3 = 0$ since x_3 in the lowest grading. We have now computed $\ker \partial = \langle x_2, x_3 \rangle$ and $\mathrm{im}\, \partial \langle x_2 \rangle$, hence $\widehat{HF}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \langle x_3 \rangle$, which implies $\mathrm{gr}(x_1) = 2, \mathrm{gr}(x_2) = 1$ and $\mathrm{gr}(x_3) = 0$.

Now let us compute the relative Alexander gradings. Recall that this is defined by $f(\mathbf{x}, \mathbf{y}) = n_z(\phi) - n_w(\phi)$. For the disc ϕ_{12} connecting x_1 to x_2 we clearly have $f(x_1, x_2) = 1$. The disc ϕ_{23} connecting x_2 to x_3 intersects w negatively, hence $f(x_2, x_3) = 1$. And by additivity of this grading we have $f(x_1, x_3) = 2$. The symmetry property this grading is supposed to have implies that $f(x_1) = 1, f(x_2) = 0, f(x_3) = -1$. Finally, we compute the homology of $(\widehat{CFK}(\alpha, \beta, z, w), \partial_k)$. However, none of the discs connecting the intersection points miss the points z and w, hence the differentials are zero. Therefore $\widehat{HFK}(\alpha, \beta, z, w) = \langle x_1, x_2, x_3 \rangle$, and the gradings look like

$\widehat{HFK}_{ij}$	-1	0	1
2			$\langle x_1 \rangle$
1		$\langle x_2 \rangle$	
0	$\langle x_3 \rangle$		

It is clear that the Euler characteristic of this bigraded vector space is just the Alexander polynomial: $\Delta_k(t) = t^{-1} - 1 + t$.
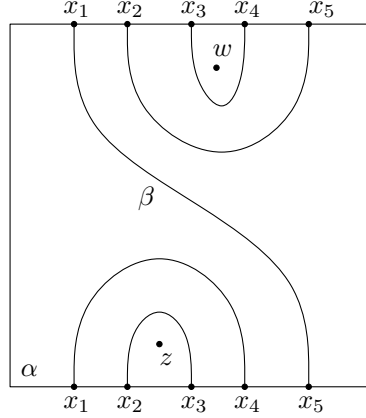
Figure 8.2: A Heegaard diagram for the figure eight knot

The Figure Eight Knot

Figure 8.2 shows a Heegaard diagram for S^3 that is compatible with the figure eight knot. Let us first figure out the relative gradings on these generators with respect to the base point w. There is a disc connecting x_1 to x_2 and it intersects w positively, so $\text{gr}(x_1, x_2) = -1$. The obvious disc connecting x_2 to x_3 does not intersect w, and so $\text{gr}(x_2, x_3) = 1$. The obvious disc connecting x_3 to x_4 has orientations reversed, and so its Maslov index is -1 and intersects w negatively, hence $\text{gr}(x_3, x_4) = 1$. Finally, the disc from x_4 to x_5 has orientations reversed and does not intersect w, so $\text{gr}(x_4, x_5) = -1$. To pin down the absolute grading we compute the homology of $(\widehat{CF}(\alpha, \beta, w), \partial)$. The differentials can easily be computed as

$$\partial x_1 = x_4$$
$$\partial x_2 = x_3$$
$$\partial x_3 = 0$$
$$\partial x_4 = 0$$
$$\partial x_5 = x_5$$

Therefore $\ker \partial = \langle x_3, x_4, x_1 + x_5 \rangle$ and $\text{im} \, \partial \, \langle x_3, x_4 \rangle$, hence $\widehat{HF}(\alpha, \beta, w) = \langle x_1 + x_5 \rangle$. This means that x_1 and x_5 are in grading zero, and so $\text{gr}(x_2) = 1, \text{gr}(x_3) = 0$ and $\text{gr}(x_4) = -1$.

Next we compute the relative Alexander gradings. If we just use the discs we used in computing the Maslov grading then we see that the relative Alexander gradings are the same as the relative Maslov gradings. In order for the symmetry property to hold for the maslov gradings we must have $f(x_1) = f(x_3) = f(x_5) = 0, f(x_2) = 1$ and $f(x_4)$, hence the Alexander gradings are precisely the Maslov gradings. Further, the differential ∂_k on $\widehat{CFK}(\alpha, \beta, z, w)$ is zero, so we have computed $\widehat{HFK}$:

$\widehat{HFK}_{ij}$	-1	0	1
1			$\langle x_2 \rangle$
0		$\langle x_1, x_3, x_5 \rangle$	
-1	$\langle x_4 \rangle$		

It is clear that the Euler characteristic of this bigraded vector space is the Alexander polynomial of the figure eight: $\Delta_k(t) = t^{-1} - 3 + t$.
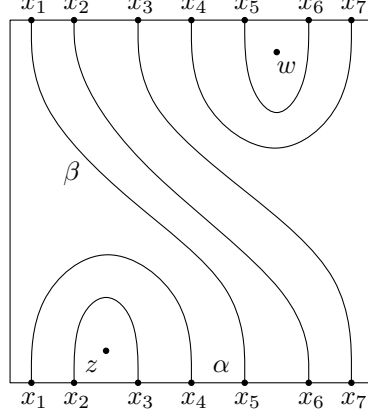
Figure 8.3: A Heegaard diagram for the knot 5_2

A Knot with 5 Crossings

Figure 8.3 shows a Heegaard diagram for S^3 that is compatible with the 5_2 knot (a figure eight knot with an extra twist). Let us first figure out the relative gradings on these generators with respect to the base point w. There is a disc connecting x_1 to x_2 that does not intersect w, so we have $\text{gr}(x_1, x_2) = 1$, and similarly $\text{gr}(x_2, x_3) = 1$. However, the disc connecting x_3 to x_4 passes through w positively, and so $\text{gr}(x_3, x_4) = -1$. Continuing this we can easily compute

$$\begin{array}{ccc} \text{gr}(x_1, x_2) = 1 & \text{gr}(x_2, x_3) = 1 & \text{gr}(x_3, x_4) = -1 \\ \text{gr}(x_4, x_5) = -1 & \text{gr}(x_5, x_6) = 1 & \text{gr}(x_6, x_7) = 1 \end{array}$$

Next we compute the homology of $(\widehat{CF}(\alpha, \beta, w), \partial)$ so that we can figure out the absolute gradings of our generators. We can compute the differentials to be:

$$\begin{aligned} \partial x_1 &= x_2 + x_4 + x_6 \\ \partial x_2 &= x_3 \\ \partial x_3 &= 0 \\ \partial x_4 &= x_3 + x_7 \\ \partial x_5 &= x_2 + x_4 + x_6 \\ \partial x_6 &= x_7 \\ \partial x_7 &= 0 \end{aligned}$$

It is easy to see that the kernel of ∂ is 4 dimensional, generated by $x_3, x_7, x_1 + x_5, x_2 + x_4 + x_6$, while its image is 3 dimensional, generated by $x_3, x_7, x_2 + x_4 + x_6$. Therefore $\widehat{HF}(\alpha, \beta, w)$ is generated by $x_1 + x_5$, and so has absolute grading 0. As a consequence we have $\text{gr}(x_1) = \text{gr}(x_5) = 0, \text{gr}(x_2) = \text{gr}(x_4) = \text{gr}(x_6) = -1$ and $\text{gr}(x_3) = \text{gr}(x_7) = -2$.

Now let us compute the Alexander gradings. We can use the same discs we used to compute the Maslov gradings. In particular, the disc from x_1 to x_2 intersects z positively but misses w, hence $f(x_1, x_2) = 1$, and similarly $f(x_2, x_3) = 1$. The disc from x_3 to x_4 intersects w positively, but misses z, hence $f(x_3, x_4) = -1$. The disc from x_4 to x_5 intersects z negative and misses w, so $f(x_4, x_5) = -1$. The disc from x_5 to x_6 intersects w negatively and misses z, so $f(x_5, x_6) = 1$. Finally, the disc from x_6 to x_7 intersects z positively, and so $f(x_6, x_7) = 1$. So, the relative Alexander gradings are equal to the relative Maslov gradings:

$$\begin{array}{ccc} f(x_1, x_2) = 1 & f(x_2, x_3) = 1 & f(x_3, x_4) = -1 \\ f(x_4, x_5) = -1 & f(x_5, x_6) = 1 & f(x_6, x_7) = 1 \end{array}$$

115

In particular, x_1, x_5 are in the same grading, x_2, x_4, x_6 are in the same grading, and x_3, x_7 are in the same grading. By the symmetry property of the Alexander grading we are forced to have $f(x_1) = f(x_5) = 1$, $f(x_2) = f(x_4) = f(x_6) = 0$ and $f(x_3) = f(x_7) = -1$. Further, the differential ∂_k on $\widehat{CFK}(\alpha, \beta, z, w)$ is trivial since all discs intersect either z or w, and so we have computed $\widehat{HFK}$:

$\widehat{HFK}_{ij}$	-1	0	1
0			$\langle x_1, x_5 \rangle$
-1		$\langle x_2, x_4, x_6 \rangle$	
-2	$\langle x_3, x_7 \rangle$		

It is clear that the Euler characteristic of this bigraded vector space is the Alexander polynomial of the 5_2 knot: $\Delta_k(t) = 2t^{-1} - 3 + 2t$.

UNFINISHED

A Homotopy Groups

Let us assume that we are working in the "nice" category of Hausdorff, compactly generated, non-degenerately based spaces. For based spaces (X, x_0) and (Y, y_0), let $f_1, f_2 : X \to Y$ be maps (not necessarily based) and $u : I \to Y$ a path. We say that f_1 and f_2 are **freely homotopic along** u if there is a homotopy $F : X \times I \to Y$ from f_1 to f_2 such that $F(x_0, t) = u(t)$ for all $t \in I$, and we denote this by $f_1 \underset{u}{\sim} f_2$. That is, although we are allowing the base point of X to move during the homotopy, we are restricting it to move along a specific path. If f_1 and f_2 are based maps, then u is necessarily a loop. Here is a simple lemma to show this definition makes sense.

Lemma A.1.

1. For any map $f_1 : X \to Y$ and path $u : (I, 0) \to (Y, f_1(x_0))$ there is a map $f_2 : X \to Y$ such that $f_1 \underset{u}{\sim} f_2$.

2. If $f_1 \underset{u}{\sim} f_2$, $f_1 \underset{v}{\sim} f_3$ and $u \sim v \operatorname{rel} \{0, 1\}$, then $f_1 \underset{const}{\sim} f_3$.

3. If $f_1 \underset{u}{\sim} f_2$ and $f_2 \underset{v}{\sim} f_3$, then $f_1 \underset{v*u}{\sim} f_3$.

Proof.

1. By definition of non-degenerately based we have that (X, x_0) is a fibration. Therefore the homotopy extension problem $f_1 \cup u : X \times \{0\} \cup \{x_0\} \times I \to Y$ can be solved to give a map $F : X \times I \to Y$ with the desired properties.

2.

$\square$

This lemma gives us an action of $\pi_1(Y, y_0)$ on the based homotopy classes $[X, Y]_*$: for $[f] \in [X, Y]_*$ and $[u] \in \pi_1(Y, y_0)$, let $[u] \cdot [f] = [g]$, where g is any map such that $f \underset{u}{\sim} g$. This is well-defined by the above lemma.

Proposition A.1. *If Y is path connected, then the forgetful map $[X, Y]_* \to [X, Y]$ descends to a bijection $[X, Y]_* / \pi_1(Y, y_0) \to [X, Y]$. If Y is simply connected, then the forgetful map is bijective.*

B   Duality Theorems

Theorem B.1 (Poincaré Duality)**.** *If M a closed, oriented n-manifold with fundamental class $[M] \in H_n(M; \mathbb{Z})$, then the homomorphism*

$$PD : H^p(M; \mathbb{Z}) \to H_{n-p}(M; \mathbb{Z})$$

defined by $\alpha \mapsto \alpha \frown [M]$ is an isomorphism.

This isomorphism is called the **Poincaré isomorphism**. It is natural in the sense that if $f : M \to N$ is a map of closed, oriented manifolds such that $f_*[M] = [N]$, then the following diagram commutes

$$
\begin{array}{ccc}
H^p(M; \mathbb{Z}) & \xrightarrow{\ PD\ } & H_{n-p}(M; \mathbb{Z}) \\
\big\uparrow{\scriptstyle f^*} & & \big\downarrow{\scriptstyle f_*} \\
H^p(N; \mathbb{Z}) & \xrightarrow[\ PD\ ]{} & H_{n-p}(N; \mathbb{Z})
\end{array}
$$

The duality isomorphism leads to a bilinear pairing

$$H^p(M; \mathbb{Z}) \otimes H^{n-p}(M; \mathbb{Z}) \to \mathbb{Z}$$

which is non-degenerate

We can drop the orientability assumption and get a $\mathbb{Z}/2$-form of Poincaré duality.

Theorem B.2. *If M is a closed n-manifold with a $\mathbb{Z}/2$-orientation given by $[M] \in H_n(M; \mathbb{Z}/2)$, then the homomorphism*

$$PD : H^p(M; \mathbb{Z}/2) \to H_{n-p}(M; \mathbb{Z}/2)$$

defined by $\alpha \mapsto \alpha \frown [M]$ is an isomorphism.

We can also extend these duality theorems to manifolds with boundary.

Theorem B.3 (Poincaré-Lefschetz Duality)**.** *Let M be a compact, oriented n-manifold with boundary with fundamental class $[M, \partial M] \in H_n(M, \partial M; \mathbb{Z})$. Then the homomorphisms*

$$PD : H^p(M, \partial M; \mathbb{Z}) \to H_{n-p}(M; \mathbb{Z})$$

$$PD : H^p(M; \mathbb{Z}) \to H_{n-p}(M, \partial M; \mathbb{Z})$$

given by capping homology classes with $[M, \partial M]$ are isomorphisms.

Finally, a more general form of duality is Alexander duality.

Theorem B.4 (Alexander Duality)**.** *Let X be a closed subset of S^n which is a CW complex. Then*

$$\tilde{H}^i(S^n \backslash X) \cong \tilde{H}_{n-i-1}(X)$$

C   Cohomology with Homotopy

There is a way of dealing with cohomology classes in terms of homotopy classes of maps. Recall that an **Eilenberg-MacLane** space of type (G, n) (where G is abelian if $G \geq 2$) is a CW-complex $K(G, n)$ such that $\pi_i(K(G, n)) = G$ if $i = n$ and 0 otherwise. Hurewicz's theorem implies that $H_i(K(G, n)) = 0$ for all $i < n$ and $H_n(K(G, n)) = G$, and so the exact sequence from the universal coefficient theorem gives a *natural* isomorphism

$$H^n(K(G, n); G) \cong \mathrm{Hom}(H_n(K(G, n)), G) \cong \mathrm{Hom}(G, G)$$

Let $\iota \in H^n(K(G, n); G)$ denote the cohomology class associated to $\mathrm{id} : G \to G$ in the above isomorphisms. Then each homotopy class of map $f : X \to K(G, n)$ gives a cohomology class $f^* \iota \in H^n(X; G)$. This association turns out to be a *natural* isomorphism of abelian groups

$$[X, K(G, n)] \cong H^n(X; G)$$

where $[X, K(G, n)]$ is given the canonical group structure from the H-group structure on $K(G, n) = \Omega K(G, n - 1)$. Naturality of the isomorphism means that for any map $f : X \to Y$ the following diagram commutes

$$
\begin{array}{ccc}
[X, K(G, n)] & \xleftarrow{\;f^*\;} & [Y, K(G, n)] \\
\cong \downarrow & & \downarrow \cong \\
H^n(X; G) & \xleftarrow[f^*]{} & H^n(Y; G)
\end{array}
$$

where f^* denotes the induced map both on the homotopy class of maps and on cohomology.

In the special case of $n = 2$ and $G = \mathbb{Z}$ we get another view of cohomology. The 2nd degree cohomology $H^2(X; \mathbb{Z})$ is isomorphic to $[X, K(\mathbb{Z})] = [X, \mathbb{C}P^\infty]$. But, at the same time, the classifying space of $U(1)$ is $BU(1) = \mathbb{C}P^\infty$, and so $[X, \mathbb{C}P^\infty]$ is in one-to-one correspondence with complex line bundles over X. So each cohomology class $\alpha \in H^2(X; \mathbb{Z})$ corresponds to a complex line bundle $L_\alpha \to X$.

D Homeomorphisms of Surfaces

When constructing combinatorial pictures of 3-manifolds it will be useful to know something about self-homeomorphisms of surfaces. We will be gluing together manifolds with boundary equal to a surface of genus g along a self-homeomorphism of that surface, so in particular we only need to know self-homeomorphisms of surfaces *up to isotopy*. This motivates the definition of the mapping class group. For any manifold M, let $\mathrm{Homeo}(M)$ denote the group of orientation preserving, self-homeomorphisms of M, and let $\mathrm{Homeo}_0(M)$ denote the normal subgroup of $\mathrm{Homeo}(M)$ that are isotopic to the identity. The quotient $H(M) = \mathrm{Homeo}(M)/\mathrm{Homeo}_0(M)$ is called the mapping class group of M.

The structure of $H(M)$ is usually quite complicated, but in the case of the torus we have a simple description of its mapping class group, and for a general surface we have an explicit set of generators. Let Σ be a surface of genus g, and let c be an embedded circle in Σ. Associated to this curve there are 2 elements of $H(\Sigma)$, which are inverses of each other, called the Dehn twists along c. Take a small tubular neighborhood $N(c)$ of c in Σ. Since $N(c)$ is an annulus, let $\varphi : N(c) \to A$ be a homeomorphism onto the standard annulus $A = \{z : 1 \leq |z| \leq 2\}$ such that $\varphi(c) = \{z : |z| = 2\}$. Then define the homeomorphism $\tau_c : \Sigma \to \Sigma$ by

$$\tau(x) = \begin{cases} x & x \notin N(c) \\ \varphi^{-1}\left(\varphi(x) \cdot e^{2\pi i(|\varphi(x)|-1)}\right) & x \in N(c) \end{cases}$$

This determines an element of $H(\Sigma)$, and had we twisted in the other direction we would have gotten the inverse to the above element. Further Dehn twists around isotopic curves are isotopic.

There is a standard family of $3g$ embedded curves on a surface Σ of genus g, label them by $\alpha_i, \beta_i, \gamma_j$ for $i = 1, \ldots, g$ and $j = 1, \ldots, g-1$.

Theorem D.1 (Dehn-Lickorish)**.** *The group $H(M)$ is generated by Dehn twists along the curves $\alpha_i, \beta_i, \gamma_i$.*

Theorem D.1 gives us an explicit description of the mapping class group of the torus. Let $T = S^1 \times S^1$ be the 2-torus. Let $\mu = S^1 \times \{1\}$ and $\lambda = \{1\} \times S^1$ be the standard meridian and longitude on T. A homeomorphism $f : T \to T$ induces an isomorphism $f_* : \pi_1(T) \to \pi_1(T)$, where $\pi_1(T) = \mathbb{Z} \oplus \mathbb{Z}$ is generated by μ and λ. Every automorphism of $\mathbb{Z} \oplus \mathbb{Z}$ is given by an integer-valued matrix whose determinant is ± 1. If f is orientation preserving, then the matrix representation of f_* has determinant $+1$. Further, isotopic homeomorphisms induce the same map on homology, hence we have a homomorphism

$$\Phi : H(T) \longrightarrow SL(2, \mathbb{Z}) \tag{D.1}$$

Proposition D.1. *The map Φ is an isomorphism.*

Proof. We will only show the surjective part of this proposition. Any matrix $A \in SL(2, \mathbb{Z})$ can be reduced to the identity by composing with elementary transformations. Over the integers these transformations are of the form

$$\begin{pmatrix} 1 & \pm 1 \\ 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 \\ \pm 1 & 1 \end{pmatrix}$$

Under the isomorphism (D.1) these matrices correspond to Dehn twists around μ and λ, which we know generates $H(T)$ by theorem D.1. Therefore Φ is surjective. $\qquad\square$

E Some Analysis

Let V and W be Banach spaces and $f : V \to W$ a map. We say that f is a differentiable at a point $v \in V$ if there is a bounded linear map $df_v : V \to W$ such that

$$\lim_{h \to 0} \frac{||f(v+h) - f(v) - df_v(h)||}{||h||} = 0$$

We call df_v the derivative of f at v. If df_v exists for all $v \in V$ we obtain a map $df : V \to L(V, W)$, where $L(V, W)$ is the Banach space of bounded linear maps $V \to W$ given the operator norm: for $L \in L(V, W)$ we set $||L|| = \sup\{||L(v)|| : v \in V, ||v|| \leq 1\}$. Since $L(V, W)$ is another Banach space we can discuss the differentiable properties of df. In general, we say that f is of class C^1 if f is differentiable and df is continuous (i.e. bounded), and inductively say that f is of class C^k if df is of class C^{k-1}.

A C^k Banach manifold modeled on a Banach space V is a topological space X with a cover $\{U_\alpha\}$ and collection of charts $x_\alpha : U_\alpha \to E$ which are bijections onto the open set $x_\alpha(U_\alpha)$ such that for any two charts (x_α, U_α) and (x_β, U_β) we have that $x_\beta \circ x_\alpha^{-1} : x_\alpha(U_\alpha \cap U_\beta) \to x_\beta(U_\alpha \cap x_\beta)$ is C^k as a map between open sets of the Banach space E. If we speak of just *Banach manifolds*, and leave out its C^k class, then we take the manifold to be C^∞. We can develop the notions of tangent vectors and differentials in the same way as finite dimensional manifolds.

Theorem E.1 (Implicit Function Theorem). *Let $f : X \to Y$ be a C^k map between Banach manifolds. If $y \in Y$ is a regular value, then $f^{-1}(y)$ is a C^k submanifold of X and $T_x f^{-1}(y) = \ker df_x$ for all $x \in f^{-1}(y)$.*

A linear map $T : V \to W$ between Banach spaces is said to be **Fredholm** if the kernel and cokernel of T are finite dimensional, and the image of T is closed in W. Such maps are our best chance at transferring problems in the infinite dimensional setting to the finite dimensional world. The **index** of a Fredholm operator is defined to be $\operatorname{ind}(T) = \dim \ker T - \dim \operatorname{coker} T$. The space of Fredholm operators $F(V, W)$ forms an open set in $L(V, W)$, and the index function $\operatorname{ind} : F(V, W) \to \mathbb{Z}$ is locally constant.

Let X be a topological space and $P(x)$ a statement which evaluates to true or false for each $x \in X$. We say that $P(x)$ is true for **generic** $x \in X$ if the solution set $\{x \in X : P(x)\}$ contains a countable intersection of open dense sets. The infinite dimensional version of Sard's theorem can now be stated.

Theorem E.2 (Sard-Smale). *Let X and Y be separable Banach manifolds. Let $f : X \to Y$ be a C^k map such that $df_x : T_x X \to T_{f(x)} Y$ is Fredholm of index l for all $x \in X$. Assume $k \geq 1$ and $k \geq l + 1$. Then a generic $y \in Y$ is a regular value of f, i.e. df_x is onto for all $x \in f^{-1}(y)$.*

Let X be a Banach manifold modeled on V and E another Banach manifold. A smooth surjective map $\pi : E \to X$ of Banach manifolds is said to be a Banach W-bundle (with W a Banach space) if X is equipped with a covering $\{U_\alpha\}$ and smooth charts $\varphi_\alpha : \pi^{-1}(U_\alpha) \to U_\alpha \times U_\alpha \times W$ such that for any two $\varphi_\alpha, \varphi_\beta$ we have $\varphi_\beta \circ \varphi_\beta : (U_\alpha \cap U_\beta) \times W \to (U_\alpha \cap U_\beta) \times W$ is of the form $(p, w) \mapsto (p, g(p)(w))$, where $g : U_\alpha \cap U_\beta \to GL(W)$ is smooth (here $GL(W)$ is the group of topological linear isomorphisms $W \to W$). This implies that each fiber $E_x := \pi^{-1}(x)$ has the structure of a Banach space, which is isomorphic to W but not in any canonical way. We denote the zero section (i.e. all of the zero vectors) of E by 0_E.

Let $\pi : E \to X$ be a Banach W-bundle, and let us write the elements of E as pairs $(x, e) \in X \times E$ where $\pi(e) = x$. The tangent space of E at a point $(x, 0)$ in the zero section naturally splits as

$$T_{(x,0)} E = T_x X \oplus E_x$$

For a point $x \in X$ let $\pi_x : T_{(x,0)}E \to E_x$ be the projection onto the second factor. If $s : X \to E$ is a section ($\pi \circ s = \text{id}$) and let $x \in X$ be a point such that $s(x) = 0$. Then using the natural identification above we define the *vertical part* of the differential ds to be the composition $Ds_x := \pi_x \circ ds_x : T_x X \to E_x$ (we use a capital D for the vertical differential and a lower case d for the regular differential). The following result is very useful.

Theorem E.3. *Let X and Y be separable Banach manifolds, $E \to X \times Y$ a Banach bundle, and $s : X \times Y \to E$ a smooth section. If for each $(x,y) \in s^{-1}(0_E)$ we have*

1. *The vertical differential $Ds_{(x,y)} : T_{(x,y)}(X \times Y) \to E_{(x,y)}$ is surjective.*

2. *The vertical differential $Ds_{(x,y)}$ restricted to $T_y Y$ is Fredholm of index k.*

Then for generic $x \in Y$, the set $\{y \in Y : s(x,y) = 0\}$ is a k-dimensional submanifold of Y. Further, the restriction of $Ds_{(x,y)}$ to $T_y Y$ is surjective.

Proof. By the first condition on s we already have that $s^{-1}(0_E)$ is a Banach submanifold of $X \times Y$ by the inverse function theorem. Let $\pi : X \times Y \to X$ be the projection onto the first factor. Let $(x,y) \in X \times Y$ such that $s(x,y) = 0$. Then, as we remarked before, the tangent space at $(x,y,0)$ in E splits as $T_{(x,y,0)}E = T_x X \oplus T_y Y \oplus E_{(x,y)}$, and so the vertical differential is given by $Ds_{(x,y)} = \pi_{(x,y)} \circ ds_{(x,y)}$, where $\pi_{(x,y)}$ is the projection of $T_{(x,y,0)}$ onto $E_{(x,y)}$.
   UNFINISHED

$\square$

For a region $U \subseteq \mathbb{R}^n$ we define $L^p(U)$ to be the space of functions $f : U \to \mathbb{R}$ such that

$$||f||_p := \left(\int_U |f|^p \right)^{1/p} < \infty$$

If $f \in L^p([a,b])$ we say that $g \in L^p([a,b])$ is the **weak derivative** of f if for every $\phi \in C^1([a,b])$ with $\phi(a) = \phi(b) = 0$ we have

$$\int_a^b f \phi' = - \int_a^b g \phi$$

More generally, if $f \in L^p(U)$ and α is a multi-index, then we say that $g \in L^p(U)$ is the α-th weak derivative of f if for every $\phi \in C_c^1(U)$ we have

$$\int_U f D^\alpha \phi = (-1)^{|\alpha|} \int_U g \phi$$

where $D^\alpha = \frac{\partial^{|\alpha|}}{\partial x^{\alpha_1} \cdots \partial x^{\alpha_k}}$ is the α-th differential operator. Weak derivatives are unique (up to a set of measure zero), and if a function is already differentiable then its weak derivative is equal to its regular derivative. Therefore we also denote the weak derivative by $D^\alpha f$, or just f' for $U \subset \mathbb{R}$. We can further generalize this ideal to functions in $L^p(U, \mathbb{R}^n)$ by defining weak derivatives for each component. The subset of $L^p(U, \mathbb{R}^n)$ with k weak derivatives in $L^p(U, \mathbb{R}^n)$ is denoted by $W^{k,p}(U, \mathbb{R}^n)$.

Proposition E.1. *If $f \in W^{1,2}(U, \mathbb{R}^n)$ ($U \subseteq \mathbb{R}$) such that f' is continuous, then $f \in C^1(U)$.*

122

Proof. For a fixed $a \in U$ define $g : U \to \mathbb{R}^n$ by

$$g(x) = \int_a^x f'(t)\, dt$$

Since f' is continuous we have that g is C^1. But, then $g' = f'$ and so $g = f$ almost everywhere. Since $W^{1,2}$ embeds into C^0 we have that f is continuous, hence $g = f$ *everywhere*, and so f is C^1. $\square$

(**todo:** Sobolev norms, completions and embeddings)

Lemma E.1. *Let X, Y, Z be Banach spaces, $D : X \to Y$ a bounded linear operator and $K : X \to Z$ a compact linear operator. If there is a constant $C > 0$ such that*

$$||x||_X \leq C \left(||Dx||_Y + ||Kx||_Z \right)$$

for all $x \in X$, then D has closed range and finite dimensional kernel.

Proof. $\square$

F   Some Exercises

Exercise F.1. Let k_1, k_2 be oriented, disjoint knots in S^3. The four definitions of $\mathrm{lk}(k_1, k_2)$ presented in these notes are equivalent.

Solution.

1. Let μ_1 be a meridian of k_1 and F_2 be an oriented Seifert surface of k_2. Perturb F_2 slightly, if necessary, so that k_1 and F_2 intersect transversally in the interior of F_2. We claim that the integer L such that $[k_2] = L[\mu_1]$ is precisely $k_1 \cdot F_2$.

$\square$

Exercise F.2. If $\overline{k}$ denotes the mirror of k, then $\Delta_k(t) = \Delta_{\overline{k}}(t)$.

Solution. Suppose $\overline{k}$ is the reflection of k through a plane Π in $\mathbb{R}^3$. If F is a Seifert surface of k, then let $\overline{F}$ denote the Seifert surface of $\overline{k}$ obtained by reflecting through Π. If $x_1, \ldots, x_{2g}$ are generators for $H_1(F)$, then let $\overline{x}_1, \ldots, \overline{x}_{2g}$ denote their reflections, and these generate $H_1(\overline{F})$. We have the relation

$$\mathrm{lk}(x_i, x_j^+) = -\mathrm{lk}(\overline{x}_i, \overline{x}_j^+)$$

therefore $\overline{S} = -S$. So,

$$\Delta_{\overline{k}}(t) = \det(t^{1/2}\overline{S} - t^{-1/2}\overline{S}^T) = \det(-(t^{1/2}S - t^{-1/2}S^T)) = \Delta_k(t)$$

$\square$

Exercise F.3. $\Delta_{k_1 \# k_2}(t) = \Delta_{k_1}(t) \cdot \Delta_{k_2}(t)$

Proof. If F_1, F_2 are Seifert surfaces for k_1, k_2 respectively, then $F_1 \natural F_2$ is a Seifert surface for $k_1 \# k_2$. We clearly have $H_1(F_1 \natural F_2) \cong H_1(F_1) \oplus H_1(F_2)$, so if we pick bases $\alpha = \{x_1, \ldots, x_{2g}\}$ and $\beta = \{y_1, \ldots, y_{2g}\}$ for $H_1(F_1)$ and $H_1(F_2)$ we can take their union to get a basis for $H_1(F_1 \natural F_2)$. Further, the push-offs of the α basis elements do not link the the β elements, and vice-versa, so the Seifert matrix S with respect to this basis is $S_1 \oplus S_2$, where S_i is the Seifert matrix of F_i. Therefore $\Delta_{k_1 \# k_2}(t) = \Delta_{k_1}(t) \cdot \Delta_{k_2}(t)$. $\square$

Exercise F.4. If k is a knot then $\Delta_k(1) = 1$, and if L is a link with more than one component then $\Delta_L(1) = 0$.

Proof. If S is a Seifert matrix of some Seifert surface of k with respect to some basis, then $\Delta_k(1) = \det(S - S^T)$. The (i, j)-entry of $S - S^T$ is $\mathrm{lk}(x_i, x_j^+) - \mathrm{lk}(x_j, x_i^+) = \mathrm{lk}(x_i, x_j^+) + \mathrm{lk}(x_i, x_j^-) = x_i \cdot x_j$. So, $S - S^T$ is the matrix of the intersection pairing $\cdot : H_1(F) \otimes H_1(F) \to \mathbb{Z}$ in the basis $\{x_i\}$, which we know is a non-degenerate, anti-symmetric form, and so it is a direct sum of matrices $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ in some basis. The determinant of this matrix is clearly 1, and so it follows $\Delta_k(1) = 1$.

We can repeat the same arguments for a link L with more than one component, except now we claim the intersection pairing $\cdot : H_1(F) \otimes H_1(F) \to \mathbb{Z}$ is degenerate. If F is a Seifert surface for L, then there is a non-trivial 1-cycle α going once around one of the boundary components of F (since ∂L has more than one connected components). However, every embedded curve in F can be isotoped to completely miss the boundary of F, hence the intersection product of α with any other cycle is zero, hence the intersection pairing is degenerate. It follows that so the determinant of any matrix representation is 0, hence $\Delta_L(t) = 0$. $\square$

Exercise F.5. The Alexander polynomial is zero for any split link.

Proof. If k_1 and k_2 are knots in S^3 that can be separated by a closed 3-ball, then $F_1 \# F_2$ is a Seifert surface for $k_1 \cup k_2$, where F_1, F_2 are Seifert surfaces for k_1, k_2 respectively. If $\{x_i\}$ and $\{y_i\}$ are bases for F_1 and F_2 respectively, then $\{x_i, y_j, \mu\}$ is a basis for $F_1 \# F_2$, where μ is a meridian of the tube used to connected F_1 and F_2. This meridian does not link with any of the push-offs of the x_i's and y_i's, and vice-versa, therefore there is an entire column of zeros in $t^{1/2} S - t^{-1/2} S^T$, and so $\Delta_{k_1 \cup k_2}(t) = 0$. $\qquad\square$

Exercise F.6. Let L be an oriented link, and let L_+, L_-, L_0 denote the links obtained from L be changing a fixed crossing to a positive crossing, a negative crossing, or resolving the crossing in an oriented fashion. Then we have the skein relation

$$\Delta_{L_+}(t) - \Delta_{L_-}(t) = (t^{1/2} - t^{-1/2})\Delta_{L_0}(t)$$

Proof. $\qquad\square$

Exercise F.7.

1. $\sigma(k) = -\sigma(\overline{k})$

2. $\sigma(k_1 \# k_2) = \sigma(k_1) + \sigma(k_2)$

3. $\sigma(k)$ is even for knots k

4. If k_1 and k_2 are concordant, then $\sigma(k_1) = \sigma(k_2)$.

Proof.

1. If S is a Seifert matrix for k, then $-S$ is a Seifert matrix for $\overline{k}$, so

$$\overline{k} = \sigma(-S - S^T) = \sigma(-(S + S^T)) = \sigma(S + S^T) = \sigma(k)$$

2. If S_1, S_2 are Seifert matrices for k_1, k_2, then $S_1 \oplus S_2$ is a Seifert matrix for $k_1 \# k_2$, and so the result follows.

3. First we claim that $\det(S + S^T) \neq 0$ so that $S + S^T$ has no zero eigenvalues. The Alexander polynomial can be written as $\Delta_k(t) = a_0 + a_1(t + t^{-1}) + \cdots + a_n(t^n + t^{-n})$, and since $\Delta_k(1) = 1$ we must have a_0 is odd, and hence $\det(k) = \det(S + S^T)$ is odd. Since the number of positive eigenvalues plus the number of eigenvalues of $S + S^T$ is equal to $2g$, it follows that their difference is even.

4. Recall that k_1 and k_2 are said to be concordant if .....

$\qquad\square$

Exercise F.8. For a knot k we have $\text{sign}(\Delta_k(-1)) = (-1)^{\sigma(k)/2}$.

Proof. We have $\Delta_k(-1) = \det(S + S^T)$ for some Seifert matrix S of k, and the sign of this is $\qquad\square$

References

[1] R. Bott. Morse theory indomitable. *Publications mathématiques de l'I.H.É.S.*, (68):99–114, 1988.

[2] Peter Ozsváth and Zoltán Szabó. On the heegaard floer homology of branched double-covers. *arXiv:math/0309170v1 (math.GT)*, 2003.

[3] Peter Ozsváth and Zoltán Szabó. Lectures on heegaard floer homology. *Clay Mathematics Proceedings*, 2005.

[4] J. Robbin and D. Salamon. The spectral flow and the maslov index. *Bull. London Math Soc.*, 27:1–33, 1995.