Low-Dimensional Topology and Knot Theory

András Juhász

Contents

Introduction

This course gives an introduction to low-dimensional topology. This is the study of manifolds of dimensions two, three, and four, and includes knot theory. Roughly speaking, an n-dimensional manifold is a topological space that locally looks like $\mathbb{R}^n$. They arise as configurations spaces of mechanical systems or solution sets of equations. They also appear in theoretical physics, for example as spacetimes in different versions of string theory.

To understand the distinction between low (≤ 4) and high (> 4) dimensions, we first introduce the general theory of topological and smooth manifolds, and give an overview of high dimensions. A particularly fruitful tool is to look at the critical points of smooth functions on a smooth manifold, which is called Morse theory. This allows one to decompose manifolds into handles, which can be thought of as thickened cells of a cell decomposition. The Whitney trick, which is a method to eliminate pairs of intersection points of complementary-dimensional submanifolds and only works in high dimensions, allows us to prove the celebrated h-cobordism theorem. As a corollary, we obtain the generalised Poincaré conjecture, which states that, if a smooth n-manifold for $n \geq 5$ is simply-connected and has the same homology as the n-sphere, then it is homeomorphic to it.

We then cover some results from the classical theory of 3-manifolds. This includes the Schönflies theorem (every smoothly embedded 2-sphere in $\mathbb{R}^3$ bounds a disk), Dehn's lemma and the loop theorem, unique prime decomposition of 3-manifolds via normal surface theory, and the classification of lens spaces (3-manifolds that can be obtained by gluing together two solid tori).

3- and 4-manifolds can be described using links in the 3-sphere whose components are labelled by integers. Hence, knot theory forms a fundamental part of low-dimensional topology. We introduce some classical knot invariants, such as the Alexander polynomial and the signature. These can be derived from surfaces that the knot bounds. We introduce several important classes of knots, then study the fundamental group of the knot complement. We give a brief overview of a more modern knot invariant called the Jones polynomial, whose topological meaning is still mysterious. Moving towards dimension four, we study surfaces that a knot bounds in the 4-ball. Dehn surgery is a tool for removing a neighbourhood of a knot and regluing it in order to obtain a new 3-manifold. One can obtain any 3-manifold using a sequence of such surgeries on knots in the 3-sphere. Another way to obtain new 3-manifolds from a link in the 3-sphere is via branched covers (meromorphic maps are analogues in dimension 2).

We conclude with looking at 4-manifolds. Kirby calculus gives us a diagrammatic way to define and manipulate 4-manifolds using links. We then overview the classification of simply-connected topological 4-manifolds in terms of their intersection forms on second homology due to Freedman, and the obstructions to topological 4-manifolds admitting smooth structures due to Donaldson.

It will become apparent by the end of this course that low-dimensional topology uses tools from many areas of mathematics, including geometry, analysis, and algebra.

CHAPTER 1

Higher dimensional manifolds

Manifolds form a particularly nice class of topological spaces that appear in many areas of mathematics. We first give a brief introduction to topological and smooth manifolds, but will mostly focus on smooth manifolds thereafter. We do assume some basic familiarity with 1-manifolds and surfaces. We then give an overview of Morse theory, which allows one to study smooth manifolds using the critical points of smooth functions they admit. This leads us to handle decompositions and the h-cobordism theorem, a key result that underlies the classification of higher dimensional manifolds using surgery theory.

1.1. Topological manifolds

For a non-negative integer n, an n-dimensional manifold (or n-manifold, in short) is a topological space that locally looks like Euclidean space $\mathbb{R}^n$. We now give a rigorous definition.

Definition 1.1.1. An *n-dimensional (topological) manifold M* is a topological space such that

(1) each point of M has a neighbourhood homeomorphic to $\mathbb{R}^n$,
(2) M is second countable (i.e., it has a countable basis of open sets), and
(3) M is Hausdorff (i.e., different points have disjoint neighbourhoods).

Exercise 1.1.2. Give examples of topological spaces that satisfy exactly two of conditions (1)–(3).

Topological manifolds form a subcategory of the category of topological spaces, where the morphisms are continuous maps. As for topological spaces, isomorphisms are the homeomorphisms.

We now look at some examples. Clearly, $\mathbb{R}^n$ is an n-manifold. So is the n-sphere

$$S^n = \{x \in \mathbb{R}^{n+1} : |x| = 1\}.$$

The real projective space

$$\mathbb{RP}^n = (\mathbb{R}^{n+1} \setminus \{0\})/\mathbb{R}^* = S^n/\{\pm 1\}$$

is the set of lines in $\mathbb{R}^{n+1}$. The complex projective space

$$\mathbb{CP}^n = (\mathbb{C}^{n+1} \setminus \{0\})/\mathbb{C}^* = S^{2n+1}/S^1$$

is the set of complex lines in $\mathbb{C}^{n+1}$, and has (real) dimension $2n$. All the familiar matrix groups $\mathrm{GL}(n)$, $\mathrm{O}(n)$, $\mathrm{SO}(n)$ etc. are manifolds (in fact, they are Lie groups).

Manifolds with boundary will also play an important role. We write

$$\mathbb{R}^n_+ := \{(x_1, \ldots, x_n) \in \mathbb{R}^n : x_n \geq 0\}.$$

DEFINITION 1.1.3. An n-dimensional *manifold with boundary* is a topological space M such that

(1) each point of M has a neighbourhood homeomorphic to $\mathbb{R}^n$ or $\mathbb{R}^n_+$,
(2) M is second countable, and
(3) M is Hausdorff.

The *interior* $\mathrm{Int}(M)$ of M consists of those points of M that have neighbourhoods homeomorphic to $\mathbb{R}^n$, and its *boundary* is $\partial M := M \setminus \mathrm{Int}(M)$.

Note that, if a point of M has a neighbourhood homeomorphic to $\mathbb{R}^n$, then it also has a neighbourhood homeomorphic to $\mathbb{R}^n_+$, so we cannot define ∂M as the set of points of M having a neighbourhood homeomorphic to $\mathbb{R}^n_+$. The subspace ∂M is a manifold without boundary.

For example, the n-disk

$$D^n = \{x \in \mathbb{R}^n : |x| \leq 1\}$$

is a manifold with boundary S^{n-1}. We will write B^n for the open ball $\mathrm{Int}(D^n)$. Note that every manifold is a manifold with (empty) boundary. We say that a manifold is *closed* if it is compact and has no boundary.

DEFINITION 1.1.4. A subset N of an m-manifold M is called a *submanifold* if it is a manifold with the subspace topology. We say that N is a *locally flat submanifold of* M if each $p \in N$ has a neighbourhood $U_p \subset M$ such that the pair $(U_p, U_p \cap N)$ is homeomorphic to $(\mathbb{R}^m, \mathbb{R}^n)$ for some n.

EXAMPLE 1.1.5. Let K be a non-trivial knot in S^3; i.e., a connected, locally flat 1-dimensional submanifold of S^3 such that the pair (S^3, K) is not homeomorphic to (S^3, S^1). Then the cone on K from the centre $\underline{0}$ of D^4 is a submanifold of D^4, but it is not locally flat at $\underline{0}$.

DEFINITION 1.1.6. A closed, connected n-manifold is called *orientable* if

$$H_n(M) \cong \mathbb{Z}.$$

If M is orientable, an *orientation* of M is a choice of generator of $H_n(M)$. This generator is called a *fundamental class*.

Note that the above isomorphism is not canonical, and hence there is no preferred orientation. The above homological definition of orientation captures the intuition of having a coherent system of local orientations in the following sense: Let $\alpha \in H_n(M) \cong \mathbb{Z}$ be a generator. Then, for every point $p \in M$, the image of α under the map $H_n(M) \to H_n(M, M \setminus \{p\})$ induced by the embedding $(M, \emptyset) \to (M, M \setminus \{p\})$ is a generator of $H_n(M, M \setminus \{p\}) \cong \mathbb{Z}$.

More generally, if M is a compact, connected manifold with boundary, then M is called orientable when $H_n(M, \partial M) \cong \mathbb{Z}$, and an orientation is a generator of this group. A fundamental example of a non-orientable manifold is the Möbius band. A 2-manifold is orientable if and only if it does not contain a Möbius band.

EXERCISE 1.1.7. Construct a cell decomposition of $\mathbb{RP}^n$, and use this to compute the homology groups $H_*(\mathbb{RP}^n; \mathbb{Z})$, $H_*(\mathbb{RP}^n; \mathbb{Z}_2)$, and the cohomology ring $H^*(\mathbb{RP}^n; \mathbb{Z}_2)$. For what n is $\mathbb{RP}^n$ orientable?

One of the main problems of manifold topology is the classification of manifolds.

EXERCISE 1.1.8. Show that a 0-manifold is a discrete, countable topological space.

We shall focus on connected manifolds, as every other manifold is a countable disjoint union of such. We state the classification of 1-manifolds. For a proof, see for example Fuks–Rokhlin [**13**] or Milnor [**44**].

THEOREM 1.1.9. *Every connected* 1*-manifold with boundary is homeomorphic to one of* S^1*,* $\mathbb{R}$*,* $[0, \infty)$*, or* $I := [0, 1]$.

If M and N are manifolds with boundary, we can form their *product* $M \times N$, which has dimension $\dim(M)+\dim(N)$. For example, the *n-torus* T^n is the product of n copies of S^1. We call $S^1 \times D^2$ a *solid torus*.

EXERCISE 1.1.10. Show that $\partial(M \times N) = (\partial M \times N) \cup (M \times \partial N)$.

REMARK 1.1.11. Clearly, $(\partial M \times N) \cap (M \times \partial N) = \partial M \times \partial N$. In particular, by applying the formula of Exercise 1.1.10 to $S^3 = \partial D^4 \approx \partial(D^2 \times D^2)$, we see that

$$S^3 = (S^1 \times D^2) \cup (D^2 \times S^1).$$

In other words, we can obtain the 3-sphere by gluing two solid tori along their boundaries. The gluing map interchanges the meridian $\{1\} \times S^1$ and the longitude $S^1 \times \{1\}$ of $T^2 = \partial(S^1 \times D^2)$.

Another important operation on manifolds is the *connected sum*. Suppose that M and N are manifolds of the same dimension n. Remove interiors of closed balls from M and N, and glue them along their sphere boundaries. The resulting manifold $M\#N$ only depends on which components of M and N we removed the balls from, up to homeomorphism. If $n > 2$, then $\pi_1(S^{n-1}) \cong 1$, and hence $\pi_1(M\#N)$ is the free product of $\pi_1(M)$ and $\pi_1(N)$ by the Seifert–van Kampen theorem.

A manifold of dimension two is also called a *surface*. For a positive integer g, let Σ_g be the connected sum of g copies of the 2-torus T^2, and we write $\Sigma_0 = S^2$. The surface Σ_g is orientable; i.e., two-sided. Here g is called the *genus* of the orientable surface Σ_g. We can obtain the non-orientable surface N_g by taking the connected sum of $g \geq 1$ copies of the real projective plane $\mathbb{RP}^2$. We now state the classification of compact surfaces without proof.

THEOREM 1.1.12. *Every closed, connected surface is homeomorphic to either* Σ_g *for some* $g \geq 0$*, or* N_g *for some* $g \geq 1$*. Compact surfaces with boundary are obtained from these by removing finitely many open balls.*

The proof for triangulated surfaces can be found in most introductory textbooks (see for example Munkres [**52**]) and was covered in the Part A Topology option. The fact that every topological surface can be triangulated is a non-trivial result of Radó [**57**]; see also Hatcher [**22**]. This is not true in higher dimensions, as there are topological manifolds in every dimension at least 5 that cannot be triangulated by the work of Manolescu [**39**]. Surprisingly, the proof of this result relies on a gauge-theoretic invariant of 3-manifolds called monopole Floer homology.

3-manifolds are much more complicated. While a lot is known about them, we do not have a complete classification as in dimension 2 and below. The most important 3-manifold invariant is the fundamental group. Perelman proved the famous Poincaré Conjecture only 100 years after it was first formulated. This states that the only closed simply-connected 3-manifold is S^3. In fact, he proved the much stronger Geometrisation Conjecture of Bill Thurston: Every closed 3-manifold can be cut along embedded 2-spheres and 2-tori such that each of the resulting pieces carries one of eight special geometric structures. These include spherical, Euclidean, and hyperbolic. The most difficult to understand are the hyperbolic pieces, which are determined by their fundamental groups by Mostow rigidity.

In dimensions 4 and higher, there is no hope for obtaining a complete classification. This is due to the following two results:

THEOREM 1.1.13. *For every finitely presented group G and integer $n \geq 4$, there exists a closed n-manifold M with $\pi_1(M) \cong G$.*

PROOF. Let $\langle\, x_1, \dots, x_k \,|\, r_1, \dots, r_l \,\rangle$ be a presentation of G. Consider the n-manifold X obtained by taking the connected sum of k copies of $S^1 \times S^{n-1}$. Then $\pi_1(X)$ is the free product of k copies of

$$\pi_1(S^1 \times S^{n-1}) \cong \pi_1(S^1) \times \pi_1(S^{n-1}) \cong \mathbb{Z};$$

i.e., $\pi_1(X) \cong \langle\, x_1, \dots, x_k \,\rangle$.

We are now going to change X to introduce the relations $r_1, \dots, r_l$ using an operation called surgery that we will discuss in more detail later. Let γ_i be a curve in X freely homotopic to the relation $r_i \in \pi_1(X)$ for $i \in \{1, \dots, l\}$. As $n \geq 3$, we can assume that the curves $\gamma_1, \dots, \gamma_l$ are embedded, pairwise disjoint, and have neighbourhoods $N(\gamma_i)$ homeomorphic to $S^1 \times D^{n-1}$ by representing each generator x_i using curves of the form $S^1 \times \{p\}$ for some $p \in S^{n-1}$.

An application of the Seifert–van Kampen theorem shows that

$$\pi_1(X \setminus N(\gamma_1 \cup \dots \cup \gamma_l)) \cong \pi_1(X).$$

Indeed, $\pi_1(N(\gamma_i)) \cong \mathbb{Z}$ is generated by $[\gamma_i]$, which is homotopic to the generator of

$$\pi_1(\partial N(\gamma_i)) \cong \pi_1(S^1 \times S^{n-2}) \cong \mathbb{Z},$$

where we have used that $\pi_1(S^{n-2}) \cong 1$ as $n \geq 4$. Hence, gluing $N(\gamma_i)$ to $X \setminus N(\gamma_1 \cup \dots \cup \gamma_l)$ does not change the fundamental group, as it introduces a new generator y_i and a new relation $y_i = w_i$ for an element $w_i \in \pi_1(X \setminus N(\gamma_1 \cup \dots \cup \gamma_l))$.

Finally, we glue a copy of $D^2 \times S^{n-2}$ to $\partial N(\gamma_i) \approx S^1 \times S^{n-2}$ such that ∂D^2 is identified with $S^1 \times \{p\}$ for each $i \in \{1, \dots, l\}$ for some $p \in S^{n-2}$. Again, by the Seifert–van Kampen theorem, this kills the homotopy class of γ_i. $\square$

REMARK 1.1.14. In contrast, note that the fundamental group of every closed 3-manifold has a presentation with the same number of generators as relations, and is hence not arbitrary. This holds since every closed 3-manifold admits a Heegaard decomposition; see Chapter 2.2.

THEOREM 1.1.15. *There is no algorithm to decide whether two finitely presented groups are isomorphic.*

Hence, to classify manifolds of dimension at least 4, one has to put restrictions on the fundamental group. A natural choice is the class of simply-connected manifolds. Simply-connected topological 4-manifolds were classified by Freedman; see Chapter 4.2.

1.2. Fibre bundles

Here, we review the necessary background on fibre bundles without proofs. For further details, see the books of Steenrod [**66**] and Milnor–Stasheff [**46**]. Intuitively, a fibre bundle is a space E that is locally a product, but might be twisted globally.

DEFINITION 1.2.1. A *fibre bundle* consists of a surjective, continuous map $\pi\colon E \to B$, where E is called the *total space* and B the *base space*, and a topological space F called the *fibre*. We assume that B is path-connected, and that each point of B has a neighbourhood U such that there is a homeomorphism $h\colon \pi^{-1}(U) \to U \times F$ that makes the following diagram commutative:

$$\begin{tikzcd} \pi^{-1}(U) \arrow[r, "h"] \arrow[d, "\pi"] & U \times F, \arrow[dl, "p_U"] \\ U & \end{tikzcd}$$

where p_U is projection onto U.

It immediately follows from the definition that $F_b := \pi^{-1}(\{b\})$ is homeomorphic to the fibre F for every $b \in B$. However, this homeomorphism is not canonical. Depending on the context, we will denote a fibre bundle using only the total space E, by $\pi\colon E \to B$, or by $F \to E \to B$, where $F \to E$ refers to a homeomorphism between F and $\pi^{-1}(\{b\})$ for some $b \in B$.

EXAMPLE 1.2.2. (1) A covering space is a fibre bundle where π is a local homeomorphism, and consequently F is discrete.
(2) If μ is the Möbius band, then the projection $\pi\colon \mu \to S^1$ onto its core circle is a fibre bundle with fibre $[-1, 1]$.
(3) Let $\varphi\colon F \to F$ be a homeomorphism. Then

$$M_\varphi := I \times F/_{(1,x)\sim(0,\varphi(x))},$$

called the *mapping torus* of φ, is a fibre bundle over S^1 with fibre F, where $\pi([(t, x)]) := t$ for $t \in I$ and $x \in F$. When $F = [-1, 1]$ and $\varphi(x) = -x$, we recover the previous example.

DEFINITION 1.2.3. Let $\pi\colon E \to B$ and $\pi'\colon E' \to B'$ be fibre bundles. A *bundle morphism* consists of continuous maps $\varepsilon\colon E \to E'$ and $\beta\colon B \to B'$ such that the following diagram commutes:

$$\begin{tikzcd} E \arrow[r, "\varepsilon"] \arrow[d, "\pi"] & E' \arrow[d, "\pi'"] \\ B \arrow[r, "\beta"] & B'. \end{tikzcd}$$

Bundles and morphisms between them form a category, and hence we can talk about bundle isomorphisms. If the base space B is contractible, then every bundle over it is *trivial*; i.e., equivalent to a product.

DEFINITION 1.2.4. We say that $s\colon B \to E$ is a *section* of the bundle $\pi\colon E \to B$ if $\pi \circ s = \mathrm{Id}_B$.

Not every bundle admits a section. For example, consider the nontrivial double cover of S^1.

Definition 1.2.5. Let $\pi\colon E \to B$ be a fibre bundle with fibre F. Given a continuous map $\phi\colon B' \to B$, we can form the pullback bundle $\pi'\colon \phi^*E \to B'$ by setting

$$\phi^*E := \{(b', e) \in B' \times E : \phi(b') = \pi(e)\}$$

and $\pi'(b', e) := b'$. Then this is also a fibre bundle with fibre F. Furthermore, if we set $\varepsilon(b', e) := e$, then (ε, ϕ) is a bundle morphism; i.e., the following diagram is commutative:

$$\begin{tikzcd} \phi^*E \arrow[r, "\varepsilon"] \arrow[d, "\pi'"] & E \arrow[d, "\pi"] \\ B' \arrow[r, "\phi"] & B. \end{tikzcd}$$

If s is a section of $\pi\colon E \to B$, then we can define its pullback as

$$\phi^*s(b') := (b', s(\phi(b'))).$$

Recall that a *topological group* is a group G that is endowed with a topology such that the product $G \times G \to G$ and the inverse $G \to G$ are both continuous.

Definition 1.2.6. Let $F \to E \xrightarrow{\pi} B$ be a fibre bundle, and let G be a topological group admitting a left homeomorphism action on F. A *G-atlas* on the bundle is a set of local trivialisations

$$\{(U_i, h_i\colon \pi^{-1}(U_i) \to U_i \times F)\colon i \in \mathcal{I}\},$$

where $\mathcal{I}$ is an index set and $\{U_i : i \in \mathcal{I}\}$ is an open cover of B, such that the transition maps

$$h_j \circ h_i^{-1}\colon (U_i \cap U_j) \times F \to (U_i \cap U_j) \times F$$

are of the form $h_j \circ h_i^{-1}(b, x) = (b, t_{ij}(b)x)$ for continuous *transition functions* $t_{ij}\colon U_i \cap U_j \to G$.

Two G-atlases are *equivalent* if their union is also a G-atlas. A *G-structure* on a fibre bundle is an equivalence class of G-atlases.

The transition functions satisfy the following conditions:

(1) $t_{ii} \equiv 1$,
(2) $t_{ij} = t_{ji}^{-1}$, and
(3) $t_{ij}t_{jk} = t_{ik}$.

Hence, the transition functions form a Čech 1-cocycle.

Closely related to the notion of G-bundles are principal bundles.

Definition 1.2.7. Let G be topological group. Then a *principal G-bundle* is a fibre bundle $\pi\colon P \to B$ such that P admits a continuous right action of G that preserves the fibres, and acts by homeomorphisms on each fibre freely and transitively.

In particular, the fibre of a principal G-bundle is homeomorphic to G, though not canonically. The homeomorphism becomes canonical once we fix the identity element of the fibre. A principal bundle is trivial if and only if it admits a section.

Suppose that $\pi\colon P \to B$ is a principal G-bundle, and that G acts on a topological space F on the left by homeomorphisms. Then we can form the *associated bundle*

$$P \times_G F := P \times F/_{(p,f)\sim(pg, g^{-1}f)},$$

which is a fibre bundle with a G-structure.

Conversely, given a fibre bundle $\pi\colon E \to B$ with a G-structure, we can associate to it a principal G-bundle over B by using the transition functions t_{ij} to glue $U_i \times G$ and $U_j \times G$.

DEFINITION 1.2.8. Let $\mathbb{F}$ be either $\mathbb{R}$ or $\mathbb{C}$. A rank n *vector bundle* over $\mathbb{F}$ is a fibre bundle $\mathbb{F}^n \to E \xrightarrow{\pi} B$ such that

(1) $\pi^{-1}(b)$ has the structure of an $\mathbb{F}$-vector space for every $b \in B$;
(2) every point of B has a neighbourhood U and a local trivialisation
$$h\colon \pi^{-1}(U) \to U \times \mathbb{F}^n$$
such that
$$h|_{\pi^{-1}(\{b\})}\colon \pi^{-1}(\{b\}) \to \{b\} \times \mathbb{F}^n$$
is a linear isomorphism for every $b \in U$.

Equivalently, a real vector bundle is a fibre bundle with fibre $\mathbb{R}^n$ and a $\mathrm{GL}(n,\mathbb{R})$-structure. One can construct the associated principal $\mathrm{GL}(n,\mathbb{R})$-bundle by considering the bundle of n-frames in the fibres of the vector bundle. Similarly, a complex vector bundle is associated to the principal $\mathrm{GL}(n,\mathbb{C})$-bundle of complex n-frames in its fibres.

A fundamental operation on vector bundles is the *direct sum* or *Whitney sum* operation. Given vector bundles ξ and η over the same base space B, we can define their direct sum $\xi \oplus \eta$ as a vector bundle over B whose fibre over $b \in B$ is $\xi_b \oplus \eta_b$. Its precise definition is the following:

DEFINITION 1.2.9. Let ξ and η be vector bundles over B. Then
$$\xi \oplus \eta := \Delta^*(\xi \times \eta),$$
where $\Delta\colon B \to B \times B$ is the diagonal map, and $\xi \times \eta$ is the product bundle over $B \times B$.

Given a real rank n vector bundle ξ, its structure group is $\mathrm{GL}(n,\mathbb{R})$. Fixing a Riemannian metric on ξ; i.e., a positive definite, symmetric bilinear form on each fibre of ξ that varies continuously, reduces the structure group to $O(n)$. Indeed, the associated principal $O(n)$-bundle consists of orthonormal n-frames in the fibres of ξ. Every vector bundle admits a Riemannian metric, and so we can reduce its structure group to $O(n)$. Note that $O(n)$ is a deformation retract of $\mathrm{GL}(n,\mathbb{R})$ by the Gram–Schmidt process.

Fixing a consistent orientation on the fibres of ξ reduces the structure group to $\mathrm{GL}_+(n,\mathbb{R})$. If we are also given a Riemannian metric and an orientation, the structure group is $\mathrm{SO}(n)$. Recall that $\mathrm{SO}(2) \approx S^1$ and $\pi_1(\mathrm{SO}(n)) \cong \mathbb{Z}_2$ for $n > 2$.

DEFINITION 1.2.10. For $n \geq 2$, the *spin group* $\mathrm{Spin}(n)$ is the connected double cover of $\mathrm{SO}(n)$. Let ξ be an oriented rank n vector bundle over B, and write $P(\xi)$ for the associated principal $\mathrm{SO}(n)$-bundle. A spin structure on ξ consists of a principal $\mathrm{Spin}(n)$-bundle $\tilde{P} \to B$, together with a bundle morphism $\tilde{P} \to P(\xi)$ that is fibrewise the nontrivial double cover $\mathrm{Spin}(n) \to \mathrm{SO}(n)$.

1.3. Smooth manifolds

We say that a function from $\mathbb{R}^n$ to $\mathbb{R}^m$ is *smooth* or C^∞ if it is infinitely differentiable; i.e., its coordinate functions have continuous partial derivatives of arbitrarily high order. Smooth manifolds are topological manifolds with some extra structure that allows one to define smooth maps between them. We can think of a smooth n-manifold as being glued together from open subsets of $\mathbb{R}^n$ such that the gluing maps are all smooth with a smooth inverse (i.e., diffeomorphisms). Alternatively, we can think of it as a topological manifold with a set of local charts that form the pages of an atlas, and when converting coordinates from one chart to the other, we apply a smooth map. For example, a usual atlas of the surface of the Earth, which is topologically S^2, consists of pages each showing a planar region.

Definition 1.3.1. Let M be a topological n-manifold. A *chart* on M is a homeomorphism ϕ from an open subset U of M to an open subset of $\mathbb{R}^n$. A *smooth atlas* of M is a set $\mathcal{A}$ of charts $\{(U_i, \phi_i) \colon i \in \mathcal{I}\}$ such that $M = \bigcup_{i \in \mathcal{I}} U_i$ and the *transition function*

$$\phi_j \circ \phi_i^{-1} \colon \phi_i(U_i \cap U_j) \to \phi_j(U_i \cap U_j)$$

is smooth for every $i, j \in \mathcal{I}$. We say that the smooth atlases $\mathcal{A}$ and $\mathcal{A}'$ are *equivalent* if $\mathcal{A} \cup \mathcal{A}'$ is also a smooth atlas. A *smooth structure* on the topological manifold M is an equivalence class of smooth atlases.

Remark 1.3.2. In the above definition, we could replace the word "smooth" with "C^k-differentiable" to define *C^k-differentiable manifolds*. However, there is a bijection between C^k-differentiable and smooth structures on any topological manifold for $k \geq 1$, so we will only study smooth manifolds. We will use the terms "local coordinates"and "chart" interchangeably.

If we require the transition functions to be analytic, we obtain the class of *real analytic manifolds*. If we are given charts to $\mathbb{C}^k$ with biholomorphic transition functions, we obtain *complex manifolds*. A 1-dimensional complex manifold is called a *Riemann surface*.

If the transition maps are piecewise linear, we obtain the notion of *PL manifolds*. A triangulation of a topological manifold is a homeomorphism with the topological realisation of a simplicial complex. Recall that the link of a vertex v of a triangulation is the boundary of the union of closed simplices incident to v. A PL structure corresponds to a triangulation where the link of every vertex is a sphere.

Every smooth manifold can be triangulated; see Munkres [**51**]. However, not every triangulable manifold admits a smooth structure, even if we assume that the boundary of the link of each vertex is a sphere.

Definition 1.3.3. Let M and N be smooth manifolds with smooth atlases $\{(U_i, \phi_i) : i \in \mathcal{I}\}$ and $\{(V_j, \psi_j) : j \in \mathcal{J}\}$. We say that the map $f \colon M \to N$ is C^r for $r \in \mathbb{N} \cup \{\infty\}$ if the map

$$\psi_j \circ f \circ \phi_i^{-1} \colon \phi_i(U_i \cap f^{-1}(V_j)) \to \psi_j(f(U_i) \cap V_j)$$

is r-times continuously differentiable for every $i \in \mathcal{I}$ and $j \in \mathcal{J}$. (This does not depend on the choice of atlases.) We denote the set of C^r maps from M to N by $C^r(M, N)$. We will often call C^∞ maps *smooth*.

We can endow $C^r(M, N)$ with several topologies, but we will not give the precise definition. Intuitively, a neighbourhood of a function f consists of functions whose partial derivatives up to order r are close to those of f in local coordinate systems.

Smooth manifolds with smooth maps between them form a category. An equivalence in this category is called a diffeomorphism:

Definition 1.3.4. The map $\phi\colon M \to N$ is a *diffeomorphism* if it is a homeomorphism such that both ϕ and ϕ^{-1} are smooth.

Two diffeomorphisms ϕ_0, $\phi_1\colon M \to N$ are *pseudo-isotopic* if there is a diffeomorphism $\Phi\colon M\times I \to N\times I$ such that $\Phi(x,0) = (\phi_0(x),0)$ and $\Phi(x,1) = (\phi_1(x),1)$ for every $x \in M$. If Φ is also *level preserving*; i.e., $\Phi(x,t) \in N\times\{t\}$ for every $x \in M$, then ϕ_0 and ϕ_1 are called *isotopic*.

Given an isotopy Φ, we will use the notation $\Phi_t(x) := \Phi(x,t)$ for $t \in I$. Then we can think of Φ as a smooth 1-parameter family of diffeomorphisms Φ_t connecting ϕ_0 and ϕ_1. In fact, if we endow the space of diffeomorphisms $\mathrm{Diff}(M,N)$ from M to N with the C^∞ topology, then two diffeomorphisms are isotopic if and only if they lie in the same path component of $\mathrm{Diff}(M,N)$. The proof of this involves deforming a continuous path of diffeomorphisms to a smooth one, which we omit.

We now define tangent vectors and tangent and cotangent spaces.

Definition 1.3.5. Let M be a smooth n-manifold and $p \in M$ a point. Given smooth curves γ_1, $\gamma_2\colon (-\varepsilon,\varepsilon) \to M$ such that $\gamma_1(0) = \gamma_2(0) = p$, we say that γ_1 and γ_2 are *equivalent* if their velocity vectors $(\phi\circ\gamma_1)'(0)$ and $(\phi\circ\gamma_2)'(0)$ are the same in any (and hence all) chart (U,ϕ) with $p \in U$. A *tangent vector* of M at p is an equivalence class of such curves.

We denote by T_pM the set of all tangent vectors of M at p, and call it the *tangent space* of M at p. This is a vector space, where the operations are the natural ones given in a coordinate chart; e.g., $[\gamma_1]+[\gamma_2] = [\gamma_3]$ if

$$(\phi\circ\gamma_1)'(0) + (\phi\circ\gamma_2)'(0) = (\phi\circ\gamma_3)'(0)$$

for some chart (U,ϕ) about p. Finally, we write $TM = \bigcup_{p\in M} T_pM$ for the *tangent bundle* of M. This is a vector bundle with base space M. A *vector field* on M is a section of the tangent bundle TM.

The *cotangent bundle* T^*M of M is obtained by taking the union of the dual spaces $T_p^*M = (T_pM)^*$ for $p \in M$.

To define the topology on TM, we choose an atlas $\mathcal{A} = \{(U_i,\phi_i) : i \in \mathcal{I}\}$ on M, and glue together the product bundles $U_i\times\mathbb{R}^n$ for $i \in \mathcal{I}$, as follows: For i, $j \in \mathcal{I}$, a point $x \in U_i\cap U_j$, and a vector $v \in \mathbb{R}^n$, we identify $(x,v) \in U_i\times\mathbb{R}^n$ with

$$\big(x, D(\phi_j\circ\phi_i^{-1})(\phi_i(x))(v)\big) \in U_j\times\mathbb{R}^n,$$

where $D(\phi_j\circ\phi_i^{-1})(\phi_i(x))$ is the Jacobian of $\phi_j\circ\phi_i^{-1}$ at $\phi_i(x)$.

A subset of a smooth n-manifold is called a submanifold if it is locally modelled on $(\mathbb{R}^n,\mathbb{R}^k)$ for some $k \le n$:

Definition 1.3.6. Let M be a smooth n-manifold and $k \le n$. We say that $N \subset M$ is a smooth k-dimensional *submanifold* if there is a chart (U,ϕ) about each $p \in N$ such that $\phi(N\cap U) = \phi(U)\cap\mathbb{R}^k$. The charts $(N\cap U, \phi|_{N\cap U})$ define an atlas of N, which becomes a smooth k-manifold.

If N is a submanifold of M, then TN is a subbundle of the restriction $TM|_N$. The *normal bundle* of N is defined as the quotient

$$\nu_{N\subset M} := (TM|_N)/TN.$$

Definition 1.3.7. Let S be a smooth submanifold of M. We write E for the total space of the normal bundle $\nu_{S\subset M}$ and $0_S\colon S \to E$ for the zero-section. Then we say that the open subset $N(S) \subset M$ is a *tubular* or *regular neighbourhood* of S if there is a diffeomorphism $\varphi\colon E \to N(S)$ such that $\varphi \circ 0_S = \mathrm{Id}_S$.

PROPOSITION 1.3.8. *Every smooth submanifold has a tubular neighbourhood.*

We omit the proof.

DEFINITION 1.3.9. Let N and N' be smooth submanifolds of the manifold M. We say that N and N' are *transverse* (or intersect *transversely*) if at each intersection point $p \in N \cap N'$, we have $T_pN + T_pN' = T_pM$.

If N and N' are transverse, then $N \cap N'$ is a smooth manifold. Furthermore,

$$\dim(N \cap N') = \dim(N) + \dim(N') - \dim(M),$$

which follows from the observation that $T_p(N \cap N') = T_pN \cap T_pN'$ for $p \in N \cap N'$, and the dimension formula for the intersection of two linear subspaces of a vector space.

DEFINITION 1.3.10. If N and N' are smooth submanifolds of M, then they are said to be *ambient isotopic* (or simply isotopic) if there is an isotopy $\Phi\colon M \times I \to M \times I$ such that $\Phi_0 = \mathrm{Id}_M$ and $\Phi_1(N) = N'$.

DEFINITION 1.3.11. Let $f\colon M \to N$ be a smooth map. The *differential* of f at $p \in M$ is a linear map $df_p\colon T_pM \to T_{f(p)}N$, defined as follows: Let γ be a curve representing a vector $v \in T_pM$. Then $df_p(v)$ is the tangent vector of N at $f(p)$ represented by $f \circ \gamma$. The maps df_p for $p \in M$ assemble to a morphism of vector bundles $df\colon TM \to TN$.

Let $C^\infty(M)$ denote the vector space of smooth functions $f\colon M \to \mathbb{R}$. Given a tangent vector $v \in T_pM$ represented by a curve $\gamma\colon (-\varepsilon, \varepsilon) \to M$ and a smooth function $f \in C^\infty(M)$, we can define the directional derivative vf as $(f \circ \gamma)'(0)$. This is independent of the choice of representative γ. The differential df of the function f is a section of the cotangent bundle T^*M. By definition, $df(v) := vf$.

A smooth n-manifold is orientable if and only if it admits an atlas where all the transition maps are orientation preserving, in the sense that they have differentials in $\mathrm{GL}_+(n)$.

Let us write $C_p^\infty(M)$ for the vector space of germs of smooth functions on M at p. This consists of equivalence classes of smooth functions on M defined in a neighbourhood of p such that $f \sim g$ if there exists an open set U in M containing p such that $f|_U = g|_U$. A *derivation* D at $p \in M$ is a linear functional on $C_p^\infty(M)$ that satisfies the Leibniz rule; i.e.,

$$D(fg) = D(f)g + fD(g).$$

Then there is a bijection between T_pM and the space of derivations at p. Hence, alternatively, we could have defined tangent vectors as derivations. Vector fields correspond to linear transformations on $C^\infty(M)$ that satisfy the Leibniz rule. For example, given a coordinate chart $\phi\colon U \to \mathbb{R}^n$ on M, we obtain the *coordinate vector fields* $\partial/\partial x_i$ on U for $i \in \{1, \dots, n\}$. For a function $f \in C^\infty(U, \mathbb{R})$, we let

$$(\partial/\partial x_i)(f) := \partial(f \circ \phi^{-1})/\partial x_i.$$

In Section 1.2, we defined the notion of Riemannian metric on a vector bundle. A Riemannian metric on a smooth manifold M is simply a Riemannian metric on its tangent bundle TM:

DEFINITION 1.3.12. A *Riemannian metric* g on a smooth manifold M is a positive definite, symmetric bilinear form g_p on T_pM for each $p \in M$ that varies smoothly in the sense that, for every coordinate chart $\phi\colon U \to \mathbb{R}^n$ and i, $j \in \{1, \dots, n\}$, the functions $g_p(\partial/\partial x_i, \partial/\partial x_j)$ are smooth in $p \in U$.

Differential topology is the study of the category of smooth manifolds and smooth maps between them. One of the main questions is the classification of smooth structures on topological manifolds.

THEOREM 1.3.13. *Every topological manifold of dimension at most* 3 *has a unique smooth structure, up to diffeomorphism.*

PROOF. Moise [**48, 47**] showed that every 2- and 3-manifold admits a unique PL structure; see also Hamilton [**20**]. For the uniqueness of smoothing of the PL structure, see the books of Hirsch and Mazur [**25**] and Thurston [**69**]. □

However, things suddenly change in dimension 4. There are topological manifolds that admit no smooth structure, and there are some that admit infinitely many. This will be the subject of Chapter 4.

Surprisingly, things get somewhat easier in dimensions 5 and higher, at least once one restricts the fundamental group, and considers only say simply-connected manifolds. This is because of the h-cobordism theorem that we will discuss in Section 1.9. In the heart of the proof of the h-cobordism theorem lies the *Whitney trick*, which is a geometric operation aimed at cancelling pairs of intersection points between two submanifolds of complementary dimensions, and hence realizing algebraic intersection numbers as geometric intersection numbers. In particular, Kervaire and Milnor [**32**] determined the set of smooth structures on all spheres of dimension at least 5. There is a unique smooth structure on S^5 and S^6, but there are 28 on S^7!

Given a smooth manifold M, a non-diffeomorphic smooth structure on M is called *exotic*. It is not known whether there is an exotic smooth structure on S^4. This is called the *smooth 4-dimensional Poincaré conjecture*.

We have already encountered the connected sum operation for topological manifolds. The corresponding operation in the smooth category is more subtle:

DEFINITION 1.3.14. We define the *connected sum* of two smooth, oriented, connected n-manifolds M_1 and M_2 as follows: Choose embeddings $e_i\colon D^n \hookrightarrow M_i$ for $i \in \{1, 2\}$ such that e_1 is orientation-preserving and e_2 is orientation-reversing. We obtain $M_1 \# M_2$ from the disjoint union

$$(M_1 \setminus \{e_1(0)\}) \sqcup (M_2 \setminus \{e_2(0)\})$$

by identifying $e_1(tv)$ with $e_2((1-t)v)$ for each unit vector $v \in S^{n-1}$ and $t \in (0, 1)$. We orient $M_1 \# M_2$ compatibly with M_1 and M_2.

This is independent of the choice of embeddings e_1 and e_2 up to diffeomorphism, by the work of Palais [**54**] and Cerf [**6**], who showed that, for any two embeddings e, $e'\colon D^n \to M$, there is a diffeomorphism $f\colon M \to M$ such that $e' = f \circ e$. Furthermore, the connected sum operation is commutative and associative up to diffeomorphism, and S^n is an identity element.

Boundary connected sum is a closely related operation for manifolds with boundary:

Definition 1.3.15. Let W_1 and W_2 be oriented $(n+1)$-manifolds with connected boundary. Let $H^{n+1} := D^{n+1} \cap \mathbb{R}^{n+1}_+$ be a half-disk. Choose embeddings $e_i \colon (H^{n+1}, D^n) \hookrightarrow (W_i, \partial W_i)$ for $i \in \{1,2\}$ such that $e_2 \circ e_1^{-1}$ is orientation-reversing. We obtain the *boundary connected sum* $W_1 \natural W_2$ from

$$(W_1 \setminus \{e_1(0)\}) \sqcup (W_2 \setminus \{e_2(0)\})$$

by identifying $e_1(tv)$ with $e_2((1-t)v)$ for every $v \in S^n \cap \mathbb{R}^{n+1}_+$ and $t \in (0,1)$.

Note that

$$\partial(W_1 \natural W_2) = W_1 \# W_2.$$

1.4. Embeddings, immersions, and submersions

Manifolds were first considered as subsets of some Euclidean space $\mathbb{R}^a$, and defined as common zero sets of a collection of smooth functions. In this section, we will show that every abstract smooth manifold defined using atlases indeed embeds into some $\mathbb{R}^a$.

Definition 1.4.1. A smooth map $f \colon M \to N$ is an *immersion* if its differential $df_p \colon T_pM \to T_{f(p)}N$ is injective for every $p \in M$.

We say that f is an *embedding* if it is an injective immersion that is a homeomorphism onto its image.

It is customary to use the notation $f \colon M \looparrowright N$ for immersions. An immersion f is locally an embedding, and its image is locally a submanifold of N. However, it might have self-intersections; i.e., distinct points p, $q \in M$ such that $f(p) = f(q)$. For example, consider a map from S^1 to $\mathbb{R}^2$ parametrising a figure eight curve in the plane.

Not every injective immersion is an embedding: The map $f \colon \mathbb{R} \to \mathbb{R}^2/\mathbb{Z}^2$ given by $f(t) = tv$ for $v = (1,s)$ and $s \in \mathbb{R} \setminus \mathbb{Q}$ has dense image in the torus $T^2 = \mathbb{R}^2/\mathbb{Z}^2$. So f is not a homeomorphism onto its image when that is endowed with the subspace topology. However, every injective immersion of a compact manifold is an embedding. If f is an embedding, then $f(M)$ is a submanifold of N.

Definition 1.4.2. We say that the embeddings f_0, $f_1 \colon M \to N$ are *isotopic* if there is a smooth map $F \colon M \times I \to N$ such that $F_t := F(-,t)$ is an embedding for every $t \in I$, and $F_i = f_i$ for $i \in \{0,1\}$. We call F an *isotopy.*

We can think of an isotopy as a path in the subspace $\mathrm{Emb}(M,N)$ of embeddings of M into N, considered as a subspace of $C^\infty(M,N)$. In Definition 1.3.10, we introduced the notion of ambient isotopy for two submanifolds. There is an analogous notion for embeddings. According to the isotopy extension theorem, every isotopy can be extended to an ambient isotopy:

Theorem 1.4.3 (Isotopy extension theorem). *Let M be a compact submaifold of N. Suppose that $F \colon M \times I \to N$ is an isotopy, and let us write $F_t := F(-,t)$. Then there is a smooth map $G \colon N \times I \to N$ such that $G_t := G(-,t)$ is a diffeomorphism and*

$$G_t \circ F_0 = F_t$$

for every $t \in I$.

Proof. We obtain G by integrating a vector field v on $N \times I$. This will be of the form $v_N + \partial/\partial t$, where v_N is tangent to the N-direction. We first define v_N along the submanifold

$$L := \{(f_t(x), t) : (x,t) \in M \times I\}$$

of $N \times I$. For $(x,t) \in M \times I$, we let $v_N(f_t(x), t)$ be the velocity vector of the curve $\gamma_x(t) := F_t(x)$ at t. Let $N(L)$ be a tubular neighbourhood of L in $N \times I$. We then extend v_N to $N(L)$ and multiply it with a fibrewise bump function on $N(L)$ that is 1 along L and 0 on a neighbourhood of $(N \times I) \setminus N(L)$. We finally let $v_N = 0$ on $(N \times I) \setminus N(L)$.

By construction, v is tangent to the curves $t \mapsto (\gamma_x(t), t)$ for $(x,t) \in M \times I$. Given $(x,t) \in N \times I$, we obtain $(G(x,t), t)$ by following the flow of v starting from $(x, 0)$ until we reach $N \times \{t\}$. $\square$

As the space of vector fields v tangent to $F(\{x\} \times I)$ for every $x \in M$ and of the form $v_N + \partial/\partial t$ in the proof of Theorem 1.4.3 is convex and hence contractible, the diffeomorphism $(G(x,t), t)$ of $N \times I$ is unique up to isotopy.

Proposition 1.4.4. *For every smooth n-manifold M, there is an embedding*

$$f\colon M \hookrightarrow \mathbb{R}^a$$

for some $a \in \mathbb{N}$.

Proof. We only give the proof when M is compact. Then there are charts $\varphi_i\colon U_i \to \mathbb{R}^n$ for $i \in \{1, \dots, k\}$ such that each U_i is diffeomorphic to the ball B^n, and there are concentric balls $U_i'' \subsetneq U_i' \subsetneq U_i$ with $M = \cup_{i=1}^k U_i''$. There are smooth functions $\mu_i\colon M \to I$ and $\lambda_i\colon M \to I$ such that

- $\mu_i|_{U_i'} \equiv 1$ and $\mu_i|_{M \setminus U_i} \equiv 0$, and
- $\lambda_i|_{U_i''} \equiv 1$ and $\lambda_i|_{M \setminus U_i'} \equiv 0$.

Then the maps

$$\psi_i := \mu_i \varphi_i\colon M \to \mathbb{R}^n$$

are smooth for $i \in \{1, \dots, k\}$. Furthermore, if we set $\psi = (\psi_1, \dots, \psi_k)$, then $\psi\colon M \to \mathbb{R}^{nk}$ is an immersion. Indeed, for $p \in M$, there is an $i \in \{1, \dots, k\}$ such that $p \in U_i''$, and $(d\psi_i)_p$ has rank n.

To lift ψ to an embedding, let $\Lambda := (\lambda_1, \dots, \lambda_k)$, and set

$$f := (\psi, \Lambda)\colon M \to \mathbb{R}^{nk} \times \mathbb{R}^k.$$

Suppose that $f(x) = f(y)$ for $x \neq y \in M$. Then $x \in U_i''$ for some $i \in \{1, \dots, k\}$. As $\psi_i(x) = \psi_i(y)$, and since $\psi_i = \mu_i \varphi_i = \varphi_i$ is injective in U_i', we have $y \notin U_i'$. But then $\lambda_i(y) = 0$ and $\lambda_i(x) = 1$, which is a contradiction. $\square$

Definition 1.4.5. Let $f\colon M \to N$ be a smooth map. We say that $p \in M$ is a *regular point* of f if $\mathrm{rk}(df_p) = \dim(N)$, and is a *critical point* otherwise. The point $q \in N$ is a *regular value* of f if every point of $f^{-1}(q)$ is regular, and is a *critical value* otherwise. The map f is a *submersion* if every $p \in M$ is regular.

If q is a regular value, then $f^{-1}(q)$ is a submanifold of M by the implicit function theorem.

EXERCISE 1.4.6 (Ehresmann's fibration lemma). Let $f\colon M \to N$ be a submersion. Show that if f is proper; i.e., $f^{-1}(K)$ is compact for every $K \subseteq N$ compact, then M is a fibre bundle over N with fibre $f^{-1}(\{q\})$ for $q \in N$. Give a counterexample when f is not proper.

A subset of a smooth manifold is said to be measure zero if its intersection with any coordinate chart has measure zero. This is well-defined since diffeomorphisms map measure zero sets to measure zero sets, and change of coordinate maps are diffeomorphisms. The following result of Sard plays an important role in differential topology.

THEOREM 1.4.7 (Sard's theorem). *Let $f\colon M \to N$ be a smooth map. Then the set of critical values of f forms a measure zero subset of N.*

For a proof, see the book of Milnor [**44**]. As a special case, when $\dim(M) < \dim(N)$, we obtain that $\mathrm{Im}(f)$ has measure zero in N.

It is an important question in differential topology what the smallest a is such that we can embed or immerse a given smooth n-manifold into $\mathbb{R}^a$.

THEOREM 1.4.8. *Every smooth n-manifold M can be embedded into $\mathbb{R}^{2n+1}$ and immersed into $\mathbb{R}^{2n}$.*

PROOF. Again, we only prove the case when M is compact. By Proposition 1.4.4, there is an embedding $f\colon M \hookrightarrow \mathbb{R}^a$ for some $a \in \mathbb{N}$. For $v \in S^{a-1}$, let p_v be orthogonal projection onto $v^\perp$. Our goal is to find a direction $v \in S^{a-1}$ such that $p_v \circ f$ is also an embedding or immersion.

For $p_v \circ f$ to be an immersion, $d(p_v \circ f)$ has to have rank n everywhere. To ensure this, v should not be tangent to $\mathrm{Im}(f)$. This means $v \neq df(w)/|df(w)|$ for w a unit tangent vector of M. The unit tangent bundle $STf(M)$ of $f(M)$ is $(2n-1)$-dimensional, as it has fibre S^{n-1}. Hence, by Sard's theorem, if $a > 2n$, for v outside a measure zero subset of S^{a-1}, the map $p_v \circ f$ is an immersion. Repeatedly projecting in this way, we obtain an immersion of M into $\mathbb{R}^{2n}$.

To make sure $p_v \circ f$ is also injective, consider the map $\varphi\colon (M \times M) \setminus \Delta \to S^{a-1}$ given by

$$\varphi(x, y) = \frac{f(x) - f(y)}{|f(x) - f(y)|}$$

for $x \neq y \in M$, where $\Delta = \{(x, x) : x \in M\}$ is the diagonal. By Sard's theorem, the image of φ has measure zero in S^{a-1} whenever $a - 1 > \dim(M \times M) = 2n$. If $v \in S^{a-1} \setminus \mathrm{Im}(\varphi)$, then $p_v \circ f$ is injective. Furthermore, if v also avoids the image of $STf(M)$, the map $p_v \circ f$ is also an immersion, and is hence an embedding. Repeating this process, we obtain an embedding of M into $\mathbb{R}^{2n+1}$. $\square$

The above proof can be modified to show that, in fact, embeddings of an n-manifold into $\mathbb{R}^{2n+1}$ are dense in the space of all smooth maps, which we endow with the weak C^∞ topology. Similarly, the set of immersions into $\mathbb{R}^{2n}$ is also dense. In other words, any smooth map of an n-manifold into $\mathbb{R}^{2n+1}$ can be perturbed into an embedding, and every smooth map into $\mathbb{R}^{2n}$ can be perturbed into an immersion.

Using the Whitney trick (Section 1.8), Theorem 1.4.8 can be improved, and every n-manifold embeds into $\mathbb{R}^{2n}$ and, when $n > 1$, immerses into $\mathbb{R}^{2n-1}$. However, the subsets of embeddings and immersions are no longer dense in these dimensions.

Given a manifold, it is a hard problem to determine the smallest a such that it embeds or immerses into $\mathbb{R}^a$. The immersion conjecture states that if $\alpha(n)$ is the number of ones in the binary expansion of n, then every n-manifold immerses into $\mathbb{R}^{2n-\alpha(n)}$. This has been shown by Cohen [**8**].

1.5. Morse theory

We can learn a lot about smooth manifolds by studying critical points of certain nice generic functions on them, which are called Morse. We write $C^\infty(M) := C^\infty(M, \mathbb{R})$, and let $f \in C^\infty(M)$. By Definition 1.4.5, $p \in M$ is a *critical point* of f if $df_p = 0$; i.e., if $vf = 0$ for every $v \in T_pM$. Furthermore, $c \in \mathbb{R}$ is a *critical value* of f if there is a critical point p of f with $f(p) = c$, and is a *regular value* otherwise.

Definition 1.5.1. Let $f \in C^\infty(M)$ be a smooth function on the n-manifold M. We say that the critical point p is *non-degenerate* if, in a local coordinate system $(x_1, \dots, x_n)$ about p, the *Hessian*

$$\left(\frac{\partial^2 f}{\partial x_i \partial x_j}(\underline{0}) \right)_{i,j=1,\dots,n}$$

is non-degenerate. The function f is called *Morse* if all of its critical points are non-degenerate.

We will denote the set of critical points of a smooth function f by $\mathrm{Crit}(f)$. The non-degeneracy condition is equivalent to requiring that the section $df \colon M \to T^*M$ is transverse to the 0-section at p. There exists a local coordinate system about each non-degenerate critical point in which the function has a particularly nice form. More specifically, if $(x_1, \dots, x_n)$ are local coordinates, then $\underline{0}$ is a critical point of f if and only if $\frac{\partial f}{\partial x_i}(\underline{0}) = 0$ for $i \in \{1, \dots, n\}$. Nondegeneracy of the critical point amounts to nondegeneracy of the Hessian, which is a real symmetric bilinear form, and hence is diagonalisable as $\sum_{i=1}^n a_i x_i^2$ for $a_i \in \{1, -1\}$ by Sylvester's law of inertia. The Morse lemma states that there is a local coordinate system about a nondegenerate critical point in which the Hessian is diagonal with entries ± 1, and all higher order terms of the Taylor series vanish.

Lemma 1.5.2 (Morse Lemma). *Let p be a non-degenerate critical point of $f \in C^\infty(M)$. Then there are local coordinates $(x_1, \dots, x_n)$ about p such that*

$$f(x_1, \dots, x_n) = f(\underline{0}) - x_1^2 - \dots - x_i^2 + x_{i+1}^2 + \dots + x_n^2$$

for some $i \in \{0, \dots, n\}$.

We call i the *index* of the critical point p and denote the set of index i critical points of f by $\mathrm{Crit}_i(f)$.

Proof (Non-examinable). By choosing a chart about p, we can assume that f is defined in a convex neighbourhood U of $\underline{0} \in \mathbb{R}^n$. Furthermore, by replacing f with $f - f(\underline{0})$, we can assume that $f(\underline{0}) = 0$. Then

$$f(\underline{x}) = \int_0^1 \frac{df}{dt}(t\underline{x})dt = \int_0^1 \sum_{i=1}^n \frac{\partial f}{\partial x_i}(t\underline{x}) x_i dt = \sum_{i=1}^n x_i \int_0^1 \frac{\partial f}{\partial x_i}(t\underline{x})dt,$$

where we used the fundamental theorem of calculus and $f(\underline{0}) = 0$ in the first step. Repeating the same computation with $g_i(\underline{x}) := \int_0^1 \frac{\partial f}{\partial x_i}(t\underline{x})dt$ in place of $f(\underline{x})$, and noting that $g_i(\underline{0}) = \frac{\partial f}{\partial x_i}(\underline{0}) = 0$, we obtain functions $H_{i,j}$ such that

$$f(\underline{x}) = \sum_{i,j=1}^{n} x_i x_j H_{i,j}(\underline{x}). \tag{1.5.1}$$

We can further assume that $H_{i,j} = H_{j,i}$ by replacing $H_{i,j}$ with $\frac{1}{2}(H_{i,j}+H_{j,i})$. Then $H_{i,j}(\underline{0}) = \frac{1}{2}\frac{\partial^2 f}{\partial x_i \partial x_j}(\underline{0})$.

We now adapt the proof of Sylvester's law of inertia to find the desired coordinate system. Suppose that there is a coordinate system $(u_1, \dots, u_n)$ about $\underline{0}$ in which

$$f(\underline{u}) = \sum_{i<r} \pm u_i^2 + \sum_{i,j=r}^{n} u_i u_j H_{i,j}(\underline{u}),$$

and $H_{i,j} = H_{j,i}$. This is true for $r = 1$ by equation (1.5.1), and the main result follows from this by induction on r, as follows.

By non-degeneracy of the Hessian of f, there are i, $j \geq r$ such that $H_{i,j}(\underline{0}) \neq 0$. If $i \neq j$, then let $y_i = u_i + u_j$, $y_j = u_i - u_j$, and $y_k = u_k$ for $k \notin \{i, j\}$, giving

$$2u_i u_j H_{i,j} = \frac{1}{2}(y_i^2 - y_j^2)H_{i,j}.$$

Hence, by possibly reindexing the coordinates, we can assume that $H_{r,r}(\underline{0}) \neq 0$.

Write $g(\underline{u}) := \sqrt{|H_{r,r}(\underline{u})|}$. This is a smooth, positive function in some neighbourhood U' of $\underline{0}$ in U. We now set

$$v_r(\underline{u}) := g(\underline{u}) \left(u_r + \sum_{i>r} u_i H_{i,r}(\underline{u}) / H_{r,r}(\underline{u}) \right),$$

and $v_i := u_i$ for $i \neq r$. Then the Jacobian $\left| \frac{\partial \underline{v}}{\partial \underline{u}}(\underline{0}) \right| \neq 0$, so $\underline{u} \mapsto \underline{v}$ is a change of coordinates in a neighbourhood $U'' \subset U'$ of $\underline{0}$ by the inverse function theorem. Furthermore,

$$f(\underline{v}) = \sum_{i \leq r} \pm v_i^2 + \sum_{i,j>r} v_i v_j H'_{i,j}(\underline{v})$$

for $\underline{v} \in U''$. □

Corollary 1.5.3. *A non-degenerate critical point is always isolated.*

A subset of a topological space is called *residual* if it is an intersection of countably many open dense subsets. We say that a property of elements of a topological space is *generic* if it holds for a residual subset.

Theorem 1.5.4. *Morse functions form a residual subset of $C^\infty(M)$.*

For a proof, see the book of Milnor [**43**]. Since very residual set is dense, every smooth manifold admits a Morse function, and every smooth function can be perturbed to a Morse function.

Now suppose that f is a Morse function such that different critical points have distinct values. This is a generic condition. If c is a regular value of f, then $f^{-1}(c)$ is a smooth submanifold of M by the implicit function theorem. This is the boundary of $M^c := f^{-1}((-\infty, c])$. As c increases, both $f^{-1}(c)$ and M^c change smoothly as long as c does not pass a critical value. The key observation of Morse theory is that if c passes a point of $f(\mathrm{Crit}_i(f))$, the manifold M^c changes by attaching a thickened

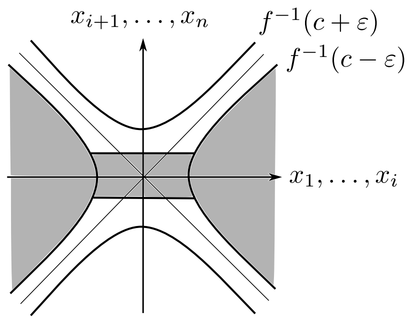

FIGURE 1. This figure shows the level sets $f^{-1}(c-\varepsilon)$, $f(c)$, and $f^{-1}(c+\varepsilon)$ of the function f in Lemma 1.5.2, where $c = f(\underline{0})$. The shaded region is $M^{c-\varepsilon}$ union the i-handle $D^i \times D^{n-i}$, which becomes diffeomorphic to $M^{c+\varepsilon}$ after smoothing the corners.

i-cell, called an n-dimensional i-handle. This is $D^i \times D^{n-i}$, which we attach along an embedding

$$h \colon \partial D^i \times D^{n-i} \hookrightarrow \partial M^c = f^{-1}(c)$$

and smooth the resulting corners. This follows from a simple local analysis of the normal form of f in Lemma 1.5.2. Indeed, if $c = f(\underline{0})$ and $\varepsilon > 0$ is small, then – up to diffeomorphism – we obtain $M^{c+\varepsilon}$ from $M^{c-\varepsilon}$ by gluing the i-handle

$$N(\{\, \underline{x} \in \mathbb{R}^n : x_{i+1} = \dots = x_n = 0 \,\}) \setminus M^{c-\varepsilon}$$

to $M^{c-\varepsilon}$; see Figure 1. We will study handle decompositions of manifolds in more detail in Section 1.6.

In particular, if we are only interested in the homotopy type of M^c, then it changes by attaching an i-cell. Successively attaching such cells, we obtain a cell complex homotopy equivalent to M. Indeed, by cellular approximation, we can homotope the attaching map of the i-cell into the $(i-1)$-skeleton of the complex corresponding to $M^{c-\varepsilon}$. Hence, using cellular homology, we obtain lower bounds on the Betti numbers of M: If c_i is the number of critical points of f of index i, then

$$b_i(M) \le c_i$$

for $i \in \{0, \dots, n\}$. These are called the *weak Morse inequalities.* Furthermore,

$$\chi(M) = \sum_{i=0}^{n} (-1)^n c_i.$$

Let v be the *gradient* of f with respect to the metric g, which is defined by the equation $g(v, w) = wf$ for all vector fields w on M. If c, $c' \in \mathbb{R}$ satisfy $c < c'$ and $f(\mathrm{Crit}(f)) \cap [c, c'] = \emptyset$, then the flow of v/vf gives a diffeomorphism between $f^{-1}([c, c'])$ and $f^{-1}(c) \times I$.

If p is an index i critical point of f, choose local coordinates $(x_1, \dots, x_n)$ about p as in Lemma 1.5.2. Assume that g is the usual Euclidean metric in these coordinates; i.e., $g(\partial/\partial x_i, \partial/\partial x_j) = \delta_{i,j}$. Then, in this chart,

$$v = 2(-x_1, \dots, -x_i, x_{i+1}, \dots, x_n).$$

We see that the *stable manifold* $D_s(v, p)$ of v at p – the set of points of M on flow lines of v converging to p – is an i-disk. Similarly, the *unstable manifold* $D_u(v, p)$

of v at p is a disk of dimension $n-i$. If $c=f(p)$, the manifold $M^{c+\varepsilon}$ is obtained from $M^{c-\varepsilon}$ by attaching a neighbourhood of the i-disk $D_s(v,p)\setminus M^{c-\varepsilon}$.

Remark 1.5.5. It is often more convenient to work with *gradient-like vector* fields instead of Riemannian metrics. Given a Morse function f on a smooth n-manifold M, a gradient-like vector field for f is a vector field v such that $vf(p)>0$ whenever $p\in M\setminus \mathrm{Crit}(f)$; furthermore, there are local coordinates $(x_1,\dots,x_n)$ about each critical point p in which

$$f(x_1,\dots,x_n)=f(p)-x_1^2-\dots-x_i^2+x_{i+1}^2+\dots+x_n^2,$$

and the vector field is the Euclidean gradient:

$$v=2\left(-x_1\frac{\partial}{\partial x_1}-\dots-x_i\frac{\partial}{\partial x_i}+x_{i+1}\frac{\partial}{\partial x_{i+1}}+\dots+x_n\frac{\partial}{\partial x_n}\right).$$

The advantage of working with gradient-like vector fields is that they are easier to manipulate and deform than Riemannian metrics.

We conclude this section with the following application of Morse theory, due to Reeb:

Theorem 1.5.6. *Let M be a smooth, closed n-manifold that admits a Morse function f with only two critical points. Then M is homeomorphic to S^n.*

Proof. Since f has only two critical points, one has to be the global minimum, which has index 0, and the other the global maximum, which has index n. Hence f can be obtained by gluing an n-handle $h^n=D^n$ to a 0-handle $h^0=D^n$ along a diffeomorphism $\phi\colon \partial h^n\to\partial h^0$. Consider the hemispheres $S^n_+ := S^n\cap\{x_{n+1}\geq 0\}$ and $S^n_- := S^n\cap\{x_{n+1}\leq 0\}$ and the points $p_\pm := S^n\cap\{x_{n+1}=\pm 1\}$. We define the homeomorphism $H\colon M\to S^n$ as follows: For $x\in h^0$, let

$$H(x)=\left(x,-\sqrt{1-|x|^2}\right).$$

We extend this to h^n radially via

$$H(x):=\left(|x|\phi(x/|x|),\sqrt{1-|x|^2}\right)\in S^n_+.$$ □

Remark 1.5.7. First, note that the above homeomorphism H is not smooth at the centre of the n-handle. The gluing map of the two handles is determined by the flow of a gradient vector field v on M between the minimum and the maximum. The smooth structure on an n-manifold obtained by gluing two n-disks depends on the isotopy class of the gluing map. This gluing map is called the *clutching function*. Kervaire and Milnor showed that there are 28 different smooth structures on S^7, all arising from this construction of gluing two copies of D^7.

For a detailed account of Morse theory, see the excellent book of Milnor [**43**]. Morse homology is developed in the book of Schwarz [**63**]. Above, we just gave an outline of the construction, and have omitted several deep analytical proofs.

1.6. Handle decompositions and surgery

When studying homotopy types of topological spaces, cell complexes play a fundamental role. These are built from cells of varying dimensions. In order to study n-manifolds, we instead study handle decompositions, which are obtained by thickening cells so that they all become n-dimensional.

Definition 1.6.1. An *n-dimensional k-handle* h^k is $D^k \times D^{n-k}$, which we can think of as a thickened k-cell. We call k the *index* of the handle. The *core* of h^k is $D^k \times \{0\}$, the *co-core* is $\{0\} \times D^{n-k}$, the *attaching sphere* is $A(h^k) := S^{k-1} \times \{0\}$, and the *belt sphere* is $B(h^k) := \{0\} \times S^{n-k-1}$.

Given an n-manifold M with boundary and an embedding

$$\varphi\colon S^{k-1} \times D^{n-k} \hookrightarrow \partial M,$$

called the *attaching map*, we can *attach* the k-handle $D^k \times D^{n-k}$ to M along φ as follows: We take the disjoint union $M \sqcup (D^k \times D^{n-k})$, identify $x \in S^{k-1} \times D^{n-k}$ with $\varphi(x) \in \partial M$, and round the corners. The result is unique up to diffeomorphism, and we denote it by $M(\varphi)$.

To specify φ, we need to define an embedding of the attaching sphere $S^{k-1}\times\{0\}$, together with a normal framing; i.e., an identification of a neighbourhood of the image with $S^{k-1} \times D^{n-k}$. So the normal bundle of the attaching sphere has to be trivial, and we also need to specify a trivialization of the normal bundle (the framing).

When attaching a handle, ∂M changes by removing the image of φ, and gluing in $D^k \times S^{n-k-1}$ using $\varphi|_{S^{k-1}\times S^{n-k-1}}$. This leads to the notion of surgery:

Definition 1.6.2. Let X be a smooth n-manifold, and $S \subset X$ an embedded k-sphere with trivial normal bundle and normal framing $\nu\colon S^k \times D^{n-k} \to N(S)$. Then the result of *surgery* on X along the framed sphere (S, ν) is

$$X(S,\nu) := (X \setminus N(S)) \cup_{\nu|_{S^k\times S^{n-k-1}}} (D^{k+1} \times S^{n-k-1}).$$

More generally, one could glue using any automorphism of $S^k \times S^{n-k-1}$, which will lead to the notion of Dehn surgery in dimension 3. Sometimes we will only write $X(\nu)$ instead of $X(S,\nu)$.

If $\Phi \in \mathrm{Diff}(X)$ is an automorphism of X, then it induces a diffeomorphism

$$\Phi(\nu)\colon X(\nu) \to X(\Phi \circ \nu).$$

We saw in Section 1.5 that, when we pass a critical point of index i of a Morse function, the sub-level set changes by attaching an i-handle. The attaching map is given by the negative gradient flow from the critical point. Consequently, the level set changes by a surgery.

Recall that a cell complex is obtained by taking a collection of 0-cells (a discrete topological space), then attaching 1-cells, followed by 2-cells, etc. The analogous construction for handles is called a handlebody:

Definition 1.6.3. An n-dimensional *handlebody* is obtained by taking finitely many n-dimensional 0-handles, and recursively attaching 1-handles, followed by 2-handles, etc.

A *handle decomposition* of the smooth n-manifold M consists of a handlebody H, together with a diffeomorphism $\phi\colon H \to M$.

Analogously, we can define relative handle decompositions, built on an n-manifold with boundary. Furthermore, the resulting handlebody can be a manifold with boundary.

Lemma 1.6.4. *If we attach a k-handle h^k followed by an l-handle h^l such that $k \geq l$, then we can isotope the attaching map of the l handle to be disjoint from h^k.*

PROOF. Note that $\dim(A(h^l)) = l-1$, $\dim(B(h^k)) = n-k-1$, and

$$(l-1)+(n-k-1) \leq n-2.$$

So we can perturb the attaching map of h^l such that its image is disjoint from $B(h^k)$. Then we can push off h^l from h^k by isotoping its attaching map radially along the rays of the core of h^k. □

PROPOSITION 1.6.5. *Every closed smooth manifold has a handle decomposition.*

PROOF. Let M be a closed smooth n-manifold. By Theorem 1.5.4, there is a Morse function $f\colon M \to \mathbb{R}$, and we can arrange that $f|_{\mathrm{Crit}(f)}$ is injective. Consider $M^c := f^{-1}((-\infty, c])$ for $c \in \mathbb{R}$. When $c < \min(f)$, we have $M^c = \emptyset$. We saw in Section 1.5 that, if $c \in f(\mathrm{Crit}_i(f))$, then $M^{c+\varepsilon}$ is obtained from $M^{c-\varepsilon}$ by attaching an n-dimensional i-handle. It follows that M can be constructed by recursively attaching handles. To arrange that the handles are attached such that their indices are nondecreasing, we apply Lemma 1.6.4. □

Any two handle decompositions of a smooth manifold are related by a set of elementary moves. However, we have to pass through handle decompositions where the handles are not necessarily attached with indices in increasing order during the intermediate steps.

The first move is an isotopy of the attaching map of one of the handles. We modify the attaching maps of subsequently attached handles suitably as well. A special case is a *handle slide*. In this case, one isotopes the attaching map of an i-handle h_1^i over the belt sphere of another i-handle h_2^i such that, along the isotopy, the image of $A(h_1^i)$ intersects $B(h_2^i)$ at a single point, and the manifold traced by $A(h_1^i)$ is transverse to $B(h_2^i)$ at the intersection point. Using the language of Morse theory, this corresponds to a flow line between two index i critical points. This does not happen for a generic gradient-like vector field, but occurs in 1-parameter families.

The second move is a *handle creation or cancellation*. Here, we add or remove an i-handle h^i and an $(i+1)$-handle h^{i+1}, where h^i and h^{i+1} are attached consecutively, and $|A(h^{i+1}) \cap B(h^i)| = 1$. This is possible since $h^i \cup h^{i+1}$ is diffeomorphic to a disk D^n, after smoothing corners. This is intuitively clear, but requires a careful analysis that we omit.

THEOREM 1.6.6. *Any two handle decompositions of a closed smooth manifold can be connected by isotopies of the attaching maps of the handles, and handle creations and cancellations.*

SKETCH OF PROOF. To prove that the above two moves are sufficient to connect any two handle decompositions of the same manifold, we again use Morse theory. We first construct Morse functions f and f' and gradient-like vector fields v and v', respectively, such that (f, v) and (f', v') induce the two handle decompositions. We obtain these by gluing model functions and vector fields on each handle. We then choose a generic 1-parameter family of smooth functions f_t for $t \in I$, such that $f_0 = f$ and $f_1 = f'$. The only singularity appearing in such a family is of the form

$$f_t(x_1, \dots, x_n) = c - x_1^2 - \cdots - x_i^2 + x_{i+1}^2 + \cdots + x_{n-1}^2 + x_n^3 \pm t x_n. \tag{1.6.1}$$

This is called a birth-death singularity, and is the suspension of the family $x^3 \pm tx$ for $t \in I$. The sign determines whether a pair of critical points are born or die.

The critical points have indices i and $i+1$, and there is a single gradient flow line connecting them. So, if we choose v_t to be the Euclidean gradient of f_t in this local coordinate system, the attaching sphere of the higher index handle will intersect the belt sphere of the lower index one in a single point.

As t increases, the attaching spheres of the handles corresponding to the critical points of f_t change by isotopies, except when f_t has a birth-death singularity, as in equation 1.6.1, which corresponds to a handle creation or cancellation. □

1.7. Cobordisms

Cobordism provides a coarser notion of equivalence between manifolds than homeomorphism or diffeomorphism. The idea stems from an attempt of Poincaré to define homology, was formalised by Pontryagin, and further studied by Thom.

Definition 1.7.1. Let M and N be closed n-manifolds. A *cobordism* from M to N is a compact $(n+1)$-manifold with boundary W such that

$$\partial W = M \sqcup N,$$

where $\sqcup$ denotes "disjoint union." If M and N are oriented, we say that W is an *oriented cobordism* if W is also oriented and

$$\partial W = -M \sqcup N.$$

Two manifolds are called *cobordant* if there is a cobordism between them.

Remark 1.7.2. If we did not require W to be compact, every manifold M would be cobordant to $\emptyset$ via $M \times [0, \infty)$.

We can define cobordisms in both the topological and the smooth category. Given cobordisms W from M_0 to M_1 and W' from M_1 to M_2, we can define their composition $W' \circ W$ by gluing them along M_1. In the smooth category, we need to smooth the corners along M_1; however, the result is unique up to diffeomorphism fixing M_0 and M_2. Hence cobordism is a transitive relation.

If W is an oriented cobordism from M to N, then

$$\partial W = -M \sqcup N = -(-N) \sqcup -M,$$

so we can also view W as a cobordism from $-N$ to $-M$. We denote this by $\overline{W}$, and call it the *reverse* of W. Note that this is not the same as $-W$, which is a cobordism from N to M. Clearly, $W = I \times M$ provides a cobordism from M to M. We have obtained the following:

Proposition 1.7.3. *Cobordism is an equivalence relation.*

In fact, cobordism classes of manifolds with the operation of disjoint union form an abelian group, called the *cobordism group* of n-manifolds, where the identity element is the class of the empty n-manifold. (This class consists of the n-manfolds that are boundaries of compact $(n+1)$-manifolds.) Furthermore, in the oriented cobordism group, the inverse of $[M]$ is $[-M]$, since we can view $I \times M$ as a cobordism from $M \sqcup -M$ to $\emptyset$. The unoriented cobordism groups are denoted by $\mathfrak{N}_n$, and the oriented ones by Ω_n^{SO}. These groups have been determined using the pioneering work of René Thom [**68**]. For the purposes of low-dimensional topology, the following groups will be important:

$$\Omega_n^{\mathrm{SO}} \cong \begin{cases} \mathbb{Z} & \text{if } n = 0, \\ 0 & \text{if } n = 1, \\ 0 & \text{if } n = 2, \\ 0 & \text{if } n = 3, \\ \mathbb{Z} & \text{if } n = 4. \end{cases}$$

EXERCISE 1.7.4. Prove that $\Omega_0^{\mathrm{SO}} \cong \mathbb{Z}$.

More generally, cartesian product of manifolds endows $\Omega_*^{\mathrm{SO}} := \bigoplus_{i \geq 0} \Omega_i^{\mathrm{SO}}$ with a ring structure. Thom proved that $\Omega_*^{\mathrm{SO}} \otimes \mathbb{Q}$ is a polynomial ring generated by the cobordism classes of the complex projective spaces $\mathbb{CP}^{2i}$ for $i > 0$. More generally, two oriented manifolds are oriented cobordant if and only if they have the same Stiefel–Whitney and Pontryagin numbers (see the book of Milnor and Stasheff [**46**] for an overview of characteristic classes).

Two 4-manifolds are cobordant if and only if they have the same signature, which is a numerical invariant that we now define in full generality. Let M be a closed oriented $4k$-manifold, with fundamental class $[M] \in H_{4k}(M)$. The cup product defines a symmetric bilinear form

$$Q_M \colon H^{2k}(M) \otimes H^{2k}(M) \to \mathbb{Z}$$

by the formula $Q_M(x \otimes y) = \langle\, x \cup y, [M]\, \rangle$, which is unimodular by Poincaré duality. Then $\sigma(M) \in \mathbb{Z}$, the *signature* of M, is the signature of Q_M; i.e., the dimension of a maximal positive definite subspace minus the dimension of a maximal negative definite subspace. René Thom showed that the signature is a cobordism invariant.

We say that two cobordisms from M_0 to M_1 are *equivalent* if they are diffeomorphic relative to $M_0 \sqcup M_1$. The *cobordism category* Cob_n has objects closed (smooth and/or orientable) n-manifolds and morphisms equivalence classes of cobordisms between them. The identity cobordism of M is the class of $I \times M$. A *topological quantum field theory*, or TQFT in short, is a certain functor from Cob_n to the category of vector spaces and linear maps that take disjoint unions to tensor products. To be completely rigorous, we should define cobordisms as follows:

DEFINITION 1.7.5. A cobordism from M_0 to M_1 as a 5-tuple $(W, N_0, N_1, \phi_0, \phi_1)$, where $\partial W = N_0 \sqcup N_1$ (or $-N_0 \sqcup N_1$ if we are working with oriented manifolds), and $\phi_i \colon N_i \to M_i$ are diffeomorphisms or homeomorphisms (depending on the category) for $i \in \{0, 1\}$.

Why this is necessary becomes clear when trying to construct the identity cobordism from M to M. Using the less rigorous definition, this would be a manifold W with $\partial W = M \sqcup M$, which is impossible unless $M = \emptyset$.

DEFINITION 1.7.6. The identity cobordism from M to M is given by the tuple $(I \times M, \{0\} \times M, \{1\} \times M, e_0, e_1)$, where $e_i(i, x) = x$ for $x \in M$ and $i \in \{0, 1\}$.

More generally, Definition 1.7.5 makes it very easy to associate a cylindrical cobordism to a diffeomorphism:

DEFINITION 1.7.7. The *cylindrical cobordism* W_ψ associated to a diffeomorphism $\psi \colon M \to N$ is the tuple $(I \times M, \{0\} \times M, \{1\} \times M, \phi_0, \phi_1)$, where $\phi_0(0, x) = x$ and $\phi_1(1, x) = \psi(x)$ for $x \in M$.

Definition 1.7.8. The cobordisms $(W, N_0, N_1, \phi_0, \phi_1)$ and $(W', N_0', N_1', \phi_0', \phi_1')$ from M_0 to M_1 are equivalent if there is a diffeomorphism (or homeomorphism) $\Phi\colon W \to W'$ such that $\Phi|_{N_i} = (\phi_i')^{-1} \circ \phi_i$ for $i \in \{0,1\}$.

Morphisms in the cobordism category are equivalence classes of cobordisms, since the composition of two cobordisms is only well-defined up to equivalence due to the smoothing involved. Throughout this work, we usually use the less rigorous definition of cobordism, as it is usually straightforward to make the arguments precise.

Exercise 1.7.9. Let ψ, $\psi'\colon M \to N$ be diffeomorphism of n-manifolds. Show that the cobordisms W_ψ and $W_{\psi'}$ are equivalent if and only if ψ and ψ' are pseudo-isotopic.

If W is a cobordism from M_0 to M_1, then we say that $f\colon W \to [a,b]$ is a *Morse function* if it has only non-degenerate critical points, and $f^{-1}(a) = M_0$ and $f^{-1}(b) = M_1$. Using such a Morse function, we can obtain a relative handle decomposition of W by successively attaching handles to $I \times M_0$. Furthermore, we can arrange that the handles are attached with nondecreasing indices by Lemma 1.6.4.

Definition 1.7.10. If S is an embedded k-sphere with normal framing ν in the n-manifold M, we define the *trace of the surgery* on M along (S,ν) to be the cobordism $W(S,\nu)$ from M to $M(S,\nu)$ obtained by attaching an $(n+1)$-dimensional $(k+1)$-handle to $I \times M$ along $\{1\} \times S$ using the framing $1 \times \nu$.

Traces of surgeries admit Morse functions with a single critical point. They are known as *elementary cobordisms*. Since every cobordism admits a Morse function, every cobordism is a product of elementary cobordisms.

1.8. The Whitney trick

The Whitney trick plays a fundamental role in manifold topology. Its failure in lower dimensions is the main reason why the classification of manifolds in dimensions 3 and 4 is more difficult and has a different flavour than in higher dimensions. While this course focuses on low-dimensional topology, it is important to understand the reason behind this distinction. The Whitney trick is the key component of the proof of the h-cobordism theorem, which is the subject of the following section. However, it was originally used by Whitney [**74**] to prove that every n-manifold embeds into $\mathbb{R}^{2n}$.

Let A and B be smooth submanifolds of the n-manifold M, and write $a = \dim(A)$ and $b = \dim(B)$. If A and B are transverse and $a + b = n$, then the intersection is a discrete set of points. When A, B, and M are oriented, each point $p \in A \cap B$ has a positive or negative sign, which we denote by $\operatorname{sgn}(p) \in \{\pm 1\}$. We have $\operatorname{sgn}(p) = +1$ if and only if an oriented basis of T_pA followed by an oriented basis of T_pB gives an oriented basis of $T_pM = T_pA \oplus T_pB$. Note that this also depends on the order of the submanifolds A and B. In $B \cap A$, the intersection signs are $(-1)^{\dim(A)\dim(B)}$ times those in $A \cap B$, which is the sign of the permutation swapping the bases of T_pA and T_pB.

If A and B are closed and oriented, they represent homology classes $[A]$, $[B] \in H_*(M)$. Their algebraic intersection is

$$\#(A \cap B) := \sum_{p \in A \cap B} \operatorname{sgn}(p) \in \mathbb{Z}.$$

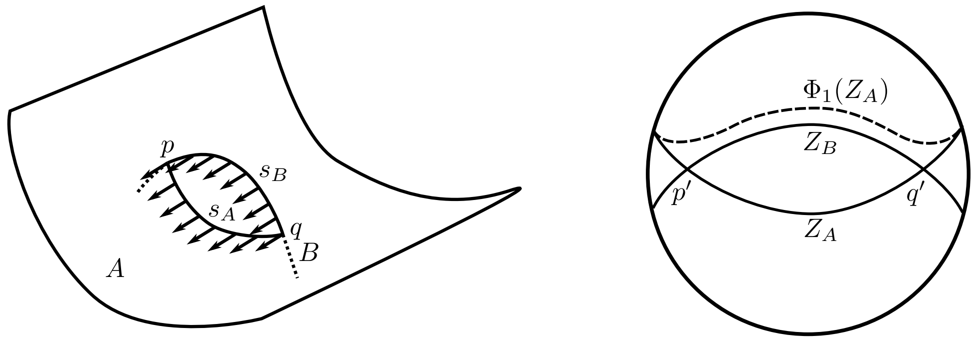

FIGURE 2. An illustration of the proof of the Whitney trick. The left shows the trivialisation of the normal bundle of the Whitney disk. The right shows the standard model in the disk.

If α, $\beta \in H^*(M)$ are the Poincaré duals of $[A]$ and $[B]$, respectively, then this agrees with $\langle \alpha \cup \beta, [M] \rangle$. More generally, $\alpha \cup \beta$ is dual to $A \cap B$, even if we do not assume that $\dim(A) + \dim(B) = \dim(M)$. In particular, $\#(A \cap B)$ only depends on the homology classes that A and B represent.

PROPOSITION 1.8.1 (Whitney trick). *Let A and B be oriented, smooth submanifolds of the oriented n-manifold M that intersect transversely, and write $a = \dim(A)$ and $b = \dim(B)$. Suppose the following hold:*

(1) $a + b = n$,
(2) $n \geq 5$, $a \geq 1$, and $b \geq 3$,
(3) p, $q \in A \cap B$ have opposite signs,
(4) when a is 1 or 2, then the map $\pi_1(M \setminus B) \to \pi_1(M)$ is injective,
(5) there are embedded paths s_A and s_B from p to q in A and B, respectively, with interiors disjoint from $A \cap B$, such that $s_A \bar{s}_B$ is contractible in M.

Then there is an isotopy $h_t \colon M \to M$ for $t \in I$ such that $h_0 = Id_M$ and

$$h_1(A) \cap B = A \cap B \setminus \{p, q\}.$$

In other words, we can eliminate intersection points of opposite signs using an ambient isotopy, under the assumptions above.

PROOF (NON-EXAMINABLE). Since the curve $C := s_A \bar{s}_B$ is 0-homotopic in M, there is a map $h \colon D^2 \to M$ such that $h|_{\partial D^2} = C$. As $n \geq 5$, we can perturb the map h to be an embedding by Theorem 1.4.8 and the remark following it, and to be transverse to A and B along $\mathrm{Int}(D^2)$. When a, $b \geq 3$, the latter implies that $h(\mathrm{Int}(D^2)) \cap (A \cup B) = \emptyset$, since both A and B have codimension at least 3, while D^2 is 2-dimensional. When a is 1 or 2, condition (4) implies that C is also contractible in $M \setminus B$, and hence we can choose h such that $h(\mathrm{Int}(D^2)) \cap B = \emptyset$. As the codimension of A is still $b \geq 3$, we have $h(\mathrm{Int}(D^2)) \cap A = \emptyset$ by transversality. We can also assume that $T_x h(D^2) \cap T_x A = T_x s_A$ for every $x \in s_A$ and $T_x h(D^2) \cap T_x B = T_x s_B$ for every $x \in s_B$.

Let $v_1, \ldots, v_{a-1}$ be linearly independent normal vector fields along s_A tangent to A; see the left-hand side of Figure 2. Then $v_i(p)$ and $v_i(q)$ are normal to both $h(D^2)$ and B for $i \in \{1, \ldots, a-1\}$. We extend these vector fields to s_B such that they are normal to both $h(D^2)$ and B. This is possible since the normal to

$Th(D^2) + TB$ over s_B is a trivial rank $a-1$ vector bundle. Furthermore, the frames $v_1(p), \dots, v_{a-1}(p)$ and $v_1(q), \dots, v_{a-1}(q)$ lie in the same path component of $\mathrm{GL}(a-1)$ as $\mathrm{sgn}(p) = -\mathrm{sgn}(q)$.

Since D^2 is contractible, the normal bundle of $h(D^2)$ in M is a trivial $(n-2)$-bundle ε^{n-2}. The vector fields $v_1, \dots, v_{a-1}$ give a map φ from the curve C to the Stiefel manifold $V_{a-1}(\mathbb{R}^{n-2})$, which is $(b-2)$-connected. (The Stiefel manifold $V_k(\mathbb{R}^l)$ consists of k-frames in $\mathbb{R}^l$, and $\pi_i(V_k(\mathbb{R}^l)) = 0$ for $i < l-k$.) Hence, the frame $v_1, \dots, v_{a-1}$ extends to the normal bundle of $h(D^2)$ in M, giving a splitting $\varepsilon^{n-2} \cong \varepsilon^{a-1} \oplus \varepsilon^{b-1}$ (the complement of ε^{a-1} is trivial since every bundle is trivial over D^2).

Now we consider the standard model $D^2 \times \mathbb{R}^{a-1} \times \mathbb{R}^{b-1}$, together with properly embedded arcs Z_A, $Z_B \subset D^2$ that intersect transversely in a pair of points p' and q'. See the right-hand side of Figure 2. Let D' be the closure of the component of $D^2 \setminus (Z_A \cup Z_B)$ disjoint from ∂D^2. The above splitting of the normal bundle of $h(D^2)$ allows us to define a diffeomorphism

$$\psi\colon D^2 \times \mathbb{R}^{a-1} \times \mathbb{R}^{b-1} \to N(h(D^2)),$$

such that

- $\psi(D' \times \{0\} \times \{0\}) = h(D^2)$,
- $\psi(p', 0, 0) = p$ and $\psi(q', 0, 0) = q$,
- $s_A \subset \psi(Z_A \times \{0\} \times \{0\})$ and $s_B \subset \psi(Z_B \times \{0\} \times \{0\})$,
- $\psi(Z_A \times \mathbb{R}^{a-1} \times \{0\}) \subset A$ and $\psi(Z_B \times \{0\} \times \mathbb{R}^{b-1}) \subset B$.

Let Φ_t for $t \in I$ be an isotopy of D^2 such that $\Phi_0 = \mathrm{Id}_{D^2}$, $\Phi_1(Z_A) \cap Z_B = \emptyset$, and $\Phi_t|_{\partial D^2} = \mathrm{Id}_{\partial D^2}$ for every $t \in I$ (first define Φ_t on Z_A, then extend it to D^2 using Theorem 1.4.3). Furthermore, let $\lambda\colon \mathbb{R} \to I$ be a smooth function such that $\lambda(0) = 1$ and $\lambda(t) = 0$ whenever $|t| \geq 1$. Then we define the desired isotopy h_t of M on $\psi(D^2 \times \{x\} \times \{y\})$ to be $\psi \circ \Phi_{t\lambda(|(x,y)|)} \circ \psi^{-1}$ for $x \in \mathbb{R}^{a-1}$ and $y \in \mathbb{R}^{b-1}$, and the identity outside $N(h(D^2))$. $\square$

1.9. The h-cobordism theorem

The h-cobordism theorem is the key technical tool that allows one to reduce the classification of simply-connected smooth and topological manifolds in dimension at least 5 to algebraic topology, proven by Smale [**65**]. It was extended to simply-connected topological manifolds in dimension 4 by Freedman [**11**]. However, Donaldson [**10**] showed it fails in the smooth category in dimension 4. In dimension 3, it is still open whether it holds, and is equivalent to the smooth 4-dimensional Poincaré conjecture. In dimension 2, it follows from the classical 3-dimensional Poincaré conjecture, proven by Perelman.

Definition 1.9.1. Let W be a cobordism from M_0 to M_1. Then we say that W is an *h-cobordism* (where "h" stands for "homotopy") if the embeddings $e_i\colon M_i \hookrightarrow W$ are homotopy equivalences for $i \in \{0, 1\}$.

Remark 1.9.2. By the relative Hurewicz theorem, when W, M_0, and M_1 are simply-connected, the following are equivalent:

(1) the embedding e_0 is a homotopy equivalence,
(2) $H_*(W, M_0) = 0$,
(3) $H^*(W, M_1) = 0$,
(4) $H_*(W, M_1) = 0$,

(5) e_1 is a homotopy equivalence.

THEOREM 1.9.3 (h-cobordism theorem). *If $n \geq 6$ and W is a simply-connected h-cobordism between the $(n-1)$-manifolds M_0 and M_1, then W is diffeomorphic to $I \times M_0$. In particular, M_0 and M_1 are diffeomorphic.*

Using the terminology of Definition 1.7.5, the conclusion of the h-cobordism theorem is that W is equivalent to W_ψ for some equivalence $\psi\colon M_0 \to M_1$.

The simple-connectivity assumption is necessary: Milnor showed that the 7-manifolds $L(7,1) \times S^4$ and $L(7,2) \times S^4$ with fundamental group $\mathbb{Z}_7$ are h-cobordant but not diffeomorphic. Here $L(7,1)$ and $L(7,2)$ are 3-dimensional lens spaces, which we will study in detail in Section 2.2.

Donaldson gave an example of a 5-dimensional smooth h-cobordism W from M_0 to M_1 that is not diffeomorphic to $I \times M_0$. On the other hand, Freedman proved the Whitney trick in the topologically locally flat category in dimension 4, which implies the h-cobordism theorem for topological 5-dimensional h-cobordisms as well, and gave his classification of simply-connected topological 4-manifolds that we will review in Section 4.2.

As an application of the h-cobordism theorem, we first give a characterisation of the smooth n-disc for $n \geq 6$.

PROPOSITION 1.9.4. *Let W be a compact simply-connected smooth n-manifold for $n \geq 6$ with simply-connected boundary. Then the following are equivalent:*

(1) W^n is diffeomorphic to D^n.
(2) W^n is homeomorphic to D^n.
(3) W^n is contractible.
(4) W^n has the same integral homology as a point.

PROOF. As the four statements are respectively weaker, it suffices to prove that (4) implies (1). Let $D_0 \subseteq \mathrm{Int}(W)$ be a smooth n-disk. By excision,

$$H_*(W \setminus \mathrm{Int}(D_0), \partial D_0) \cong H_*(W, D_0) = 0.$$

Since $\pi_1(W \setminus \mathrm{Int}(D_0)) = 1$ and $\pi_1(\partial W) = 1$, the manifold $W \setminus \mathrm{Int}(D_0)$ is a simply-connected h-cobordism from $\partial D_0 \approx S^{n-1}$ to ∂W by Remark 1.9.2. Hence, by the h-cobordism theorem, $W \setminus \mathrm{Int}(D_0)$ is equivalent to the product $I \times \partial D_0$, and so W is diffeomorphic to D^n. □

We now state and prove the Generalised Poincaré Conjecture in dimensions at least six.

THEOREM 1.9.5 (Generalised Poincaré Conjecture). *Let M be a closed, simply-connected, smooth n-manifold that has the same homology as S^n, and suppose that $n \geq 6$. Then M is* homeomorphic *to S^n.*

PROOF. Choose a handle decomposition of M. Analogously to Step 1 of the proof of the h-cobordism theorem, we can arrange that there is a single 0-handle h^0. Then

$$W := M \setminus \mathrm{Int}(h^0)$$

is simply-connected and has the same homology as a point by excision and the long exact sequence of the pair (M, W). Hence W is diffeomorphic to D^n by Proposition 1.9.4. Reattaching h^0 gives a twisted sphere, which is homeomorphic to S^n by Theorem 1.5.6, as it admits a Morse function with only two critical points. □

Note that the Generalised Poincaré Conjecture holds in every dimension in the topological category, but it is beyond the scope of this book to prove it in full generality. The 1- and 2-dimensional cases simply follow from the classification of manifolds in these dimensions. The 3-dimensional case is the classical Poincaré conjecture, proven by Perelman using the Ricci flow. The 4-dimensional case is due to Freedman. In dimension 5, the result follows from work of Kervaire and Milnor, which in fact implies the result in the smooth category in dimensions 5 and 6. However, as mentioned earlier, there are 28 non-diffeomorphic smooth structures on S^7 by the work of Milnor. It is still unknown if there is an exotic smooth structure on S^4.

CHAPTER 2

3-manifolds

Every topological 3-manifold admits a unique smooth structure up to diffeomorphism by the work of Moise [**48**], hence the categories of topological and smooth manifolds are equivalent in dimension three.

In these notes, we only discuss classical results from 3-manifold topology, preceding Thurston's revolutionary work. Most of the proofs use cut-and-paste techniques. Some of these results will then be applied in the following section on knots and links.

2.1. The Schönflies theorem

The Jordan–Schönflies theorem states that every Jordan curve in $\mathbb{R}^2$ bounds a disk. One might wonder whether this generalises to higher dimensions: Does an embedding of an n-sphere into $\mathbb{R}^{n+1}$ bound a disk? The answer depends on the category we are working in. The Alexander horned sphere is an example of a continuous embedding of D^3 into $\mathbb{R}^3$ whose complement has non-finitely generated fundamental group; see Rolfsen [**61**] for the construction. Its boundary is hence a continuously embedded S^2 in S^3 that does not bound a disk on one side.

However, if we restrict to the smooth or PL category, then the generalised Schönflies theorem is known to hold for $n \neq 3$: A smoothly or PL embedded S^n bounds a D^{n+1} in $\mathbb{R}^{n+1}$. In the smooth category, for $n > 3$ this follows from the h-cobordism theorem; see Proposition 1.9.4. The case $n = 3$ is still open. The case $n = 2$ is due to Alexander; we prove it in the smooth category:

THEOREM 2.1.1 (Schönflies theorem). *Any smoothly embedded 2-sphere S in $\mathbb{R}^3$ bounds a 3-disk.*

PROOF. We follow the exposition of Casson [**5**]. We use the following observation: If $M = M_1 \cup M_2$, where M_1 is a 3-manifold, $M_2 \approx D^3$, and

$$M_1 \cap M_2 \approx \partial M_1 \cap \partial M_2 \approx D^2,$$

then $M \approx M_1$.

We can perturb S such that the height function $x_3|_S$ is Morse for $x_3 \colon \mathbb{R}^3 \to \mathbb{R}$. Let M be the closure of the bounded component of $\mathbb{R}^3 \setminus S$. The proof proceeds by induction on the number n of index 1 critical points of $x_3|_S$.

If $n = 0$, then S has one local minimum and one local maximum. Let $t_{\min} := \min(x_3|_S)$ and $t_{\max} := \max(x_3|_S)$. Then, for every $t \in (t_{\min}, t_{\max})$, the level set $(x_3|_S)^{-1}(t)$ is a circle in $\mathbb{R}^2 \times \{t\}$ that bounds a 2-disk in $\mathbb{R}^2 \times \{t\}$ by the smooth Jordan curve theorem, and hence $M \approx D^3$.

If $n = 1$, then there are either two local minima and one local maximum, or one local minimum and two local maxima. Without loss of generality, assume it is the former. Consider the level sets of $x_3|_S$. As we pass the two local minima, two

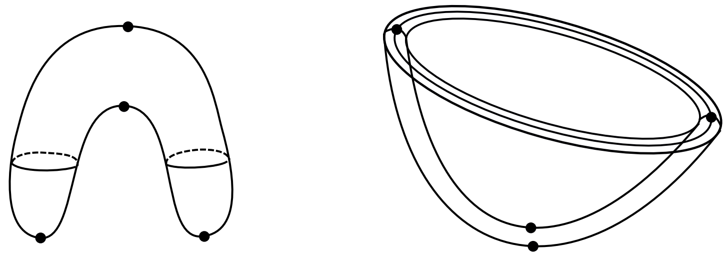

FIGURE 3. Two embeddings of S^2 in $\mathbb{R}^3$ with one saddle point and two local minima.

circles appear. If these have disjoint interiors, then M is a regular neighbourhood of a $\cap$-shaped curve; see the left of Figure 3. If one circle is contained in the interior of the other, M is a "tilted bowl;" see the right of Figure 3. In both cases, M bounds a disk.

Now suppose that $n \geq 2$. Let t be a regular value of $x_3|_S$ such that there is at least one saddle point of $x_3|_S$ on each side of $H := \mathbb{R}^2 \times \{t\}$. Then $S \cap H$ is a compact 1-manifold. Let C be an innermost component of $H \cap S$. This bounds a disk $D \subset H$ such that $D \cap S = \partial D = C$. Then C separates S into disks D_1 and D_2, so $S_1 = D \cup D_1$ and $S_2 = D \cup D_2$ are embedded 2-spheres in $\mathbb{R}^3$.

If both $x_3|_{S_1}$ and $x_3|_{S_2}$ have saddles, then they bound disks M_1 and M_2, respectively, by the inductive hypothesis. There are three cases: If $M_1 \cap M_2 = D$, then $M = M_1 \cup M_2 \approx D^3$ by the observation. If $M_1 \subset M_2$, then we can apply the observation to $M = M_1 \cup (\overline{M_2 \setminus M_1})$ as $M_1 \cap (\overline{M_2 \setminus M_1}) = D_1$ is a disk. The case $M_2 \subset M_1$ is analogous.

If S_1 has no saddle, it bounds a disk M_1 by the $n = 0$ case. As before, using the observation, S bounds a disk if and only if S_2 bounds a disk. We push S_2 a bit to eliminate C from $S_2 \cap H$, reducing the number of its components. We repeat the above procedure with a new innermost circle, which has to terminate since there is a saddle on each side of H. It follows that S also bounds a disk. $\square$

EXERCISE 2.1.2. Using the same method as in the proof of the Schönflies theorem (and assuming the Schönflies theorem), show that if T is a smooth torus in S^3, then one of the components of $S^3 \setminus T$ has closure $S^1 \times D^2$. Give an example of a torus T in $\mathbb{R}^3$ such that the closure of the bounded component of $\mathbb{R}^3 \setminus T$ is not a solid torus.

2.2. Heegaard decompositions and diagrams

A *genus g handlebody* is a 3-ball with g oriented 1-handles attached. Alternatively, it can be described as a regular neighbourhood of a wedge of g unknotted circles in $\mathbb{R}^3$. Every closed, oriented, and connected 3-manifold can be obtained by gluing together two handlebodies of the same genus along their boundaries.

DEFINITION 2.2.1. Let M be a closed, connected, and oriented 3-manifold. A *Heegaard decomposition* of M consists of a closed, oriented surface $\Sigma \subset M$ that separates M into the union of two handlebodies H_α and H_β, where Σ is oriented as the boundary of H_α. We call Σ a Heegaard surface. The *Heegaard genus* of M is the minimal genus of a Heegaard surface Σ.

PROPOSITION 2.2.2. *Every closed, connected, and oriented 3-manifold M admits a Heegaard decomposition.*

PROOF. Choose a handle decomposition of M. Since M is connected, we can assume that it has a single 0-handle and a single 3-handle. Indeed, if there is more than one 0-handle, we can cancel one of them against a 1-handle until we are left with just one. The union of the 0-handle and the 1-handles is one of the handlebodies, and the union of the 3-handle and the 2-handles is another handlebody. The two handlebodies have the same genus since they share the same boundary. □

REMARK 2.2.3. An alternative proof often found in the literature uses the fact that every 3-manifold admits a triangulation. Then a Heegaard decomposition can be obtained by taking a regular neighbourhood of the 1-skeleton and its complement. The complement is a handlebody because it is a regular neighbourhood of the 1-skeleton of the dual cell decomposition, whose 0-cells are the centres of the tetrahedra of the original triangulation, and two 0-cells are connected by a 1-cell if and only if the corresponding tetrahedra share a face. It is a rather technical result that every smooth manifold admits a triangulation; see Munkres [**51**].

Given two genus g handlebodies, the diffeomorphism class of the 3-manifold obtained by gluing them together only depends on the isotopy class of the gluing map. The group of isotopy classes of orientation-preserving automorphisms of a genus g surface Σ_g is called the *mapping class group*, and is denoted by $\mathrm{MCG}(\Sigma_g)$. Due to Proposition 2.2.2, and since every handlebody admits an orientation-reversing symmetry, one can study 3-manifolds via the mapping class group.

EXERCISE 2.2.4. Show that if M is a 3-manifold, then $\pi_1(M)$ has a presentation with the same number of generators and relations. Show that $\mathbb{Z}^4$ is not the fundamental group of a closed 3-manifold.

Since every orientation-reversing automorphism of S^2 is isotopic to a reflection, the only 3-manifold of Heegaard genus zero is S^3. Heegaard genus one manifolds form an important class:

DEFINITION 2.2.5. A *lens space* is a closed, oriented 3-manifold of Heegaard genus one that is not $S^1 \times S^2$.

In other words, a lens space is a manifold that can be obtained by gluing two solid tori along their boundaries, with the exception of S^3 and $S^1 \times S^2$. The gluing map is described by an automorphism of T^2. Since $\pi_0(\mathrm{Diff}(T^2)) \cong \mathrm{GL}(2, \mathbb{Z})$, an orientation-reversing automorphism of T^2 is isotopic to a transformation given by a matrix

$$\begin{pmatrix} q & r \\ p & s \end{pmatrix}$$

with determinant -1, so p and q are relatively prime. If m is the meridian and l is the longitude, then m is mapped to $qm + pl$. This already determines the resulting 3-manifold $L(p, q)$, since every automorphism of T^2 that preserves m extends to $S^1 \times D^2$.

Alternatively, we can also describe $L(p, q)$ as the quotient of $S^3 \subset \mathbb{C}^2$ by the action of the cyclic group C_p given by

$$(z, w) \mapsto (\zeta z, \zeta^q w)$$

for a p-th root of unity $\zeta \in C_p$. This is a free action, hence the universal cover of $L(p,q)$ is S^3, and $\pi_1(L(p,q)) \cong C_p$. We orient $L(p,q)$ such that this covering map is orientation-preserving. Alternatively, one can compute the fundamental group by applying the Seifert–van Kampen theorem to the first description of lens spaces.

Exercise 2.2.6. Show that the two descriptions of a lens space are equivalent. (Hint: Construct a cell decomposition of $L(p,q)$ using a C_p-equivariant cell decomposition of S^3.)

Theorem 2.2.7. *Let (p,q) and (p',q') be relatively prime pairs of integers. Then $L(p,q)$ and $L(p',q')$ are homotopy equivalent if and only if $p = p'$ and $qq' \equiv \pm n^2 \mod p$ for some $n \in \mathbb{Z}$. They are homeomorphic if and only if $p = p'$ and $q' \equiv \pm q^{\pm 1} \mod p$.*

This was shown by Reidemeister [**59**] using Reidemeister torsion. It follows from Theorem 2.2.7 that the lens spaces $L(7,1)$ and $L(7,2)$ are homotopy equivalent but not homeomorphic.

A Heegaard decomposition is not unique. For example, S^3 has a genus g Heegaard decomposition for every g: just consider the standard genus g surface in S^3. By a result of Waldhausen [**71**], every genus g Heegaard splitting of S^3 is isotopic to this.

Given 3-manifolds M and M', together with Heegaard surfaces Σ and Σ', respectively, we can take their *connected sum*: Choose 3-balls $B \subset M$ and $B' \subset M'$ such that both $B \cap \Sigma$ and $B' \cap \Sigma'$ are 2-disks. Then, if we take the connected sum of M and M' along B and B', then $\Sigma \# \Sigma'$ is a Heegaard surface of $M \# M'$.

Definition 2.2.8. Let Σ be a Heegaard surface in M. Then a stabilisation of Σ is obtained by taking the connected sum with (S^3, T^2).

The following theorem is due to Reidemeister and Singer:

Theorem 2.2.9. *Let Σ_0 and Σ_1 be Heegaard surfaces in the closed, oriented 3-manifold M. Then there is a Heegaard surface Σ that is isotopic to a stabilisation of both.*

If one would also like to encode how the two handlebodies are glued together, one can record the belt circles of the 1-handles of the two handlebodies:

Definition 2.2.10. An *abstract Heegaard diagram* is a triple $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta})$, where Σ is an oriented, genus g surface, and $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_g)$ and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_g)$ are two g-tuples of pairwise disjoint simple closed curves in Σ that are linearly independent in $H_1(\Sigma)$.

An *embedded Heegaard diagram* of the closed, connected, and oriented 3-manifold M is an abstract Heegaard diagram $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta})$ such that $\Sigma \subset M$ is a Heegaard surface, each α_i bounds a disk in H_α, and each β_j bounds a disk in H_β.

For example, (T^2, μ, λ) is a genus 1 embedded Heegaard diagram of S^3, where μ is a meridian and λ is a longitude of T^2; see Remark 1.1.11. A stabilisation of a Heegaard diagram is obtained by taking the connected sum with this.

An abstract Heegaard diagram $\mathcal{H} = (\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta})$ defines a 3-manifold as follows: Take $\Sigma \times I$, and attach 3-dimensional 2-handles along $\alpha \times \{0\}$ and $\beta \times \{1\}$ for every $\alpha \in \boldsymbol{\alpha}$ and $\beta \in \boldsymbol{\beta}$. Since the α-curves are linearly independent in $H_1(\Sigma)$ and there is g of them, the lower boundary component, which is obtained by surgering

Σ along $\boldsymbol{\alpha}$, is a 2-sphere. Similarly, the upper boundary component is also S^2. After attaching two 3-balls to these two spheres, we obtain a closed, connected, and oriented 3-manifold, which we call the 3-manifold defined by $\mathcal{H}$. If $\mathcal{H}$ is an embedded Heegaard diagram of M, then the 3-manifold defined by $\mathcal{H}$ is diffeomorphic to M relative to Σ.

If $\mathcal{H} = (\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta})$ is a Heegaard diagram of M, then we say that $\boldsymbol{\alpha}' = \boldsymbol{\alpha} \setminus \{\alpha_i\} \cup \{\alpha_i'\}$ is obtained from $\boldsymbol{\alpha}$ by sliding α_i over α_j if α_i, α_i', and α_j is the boundary of a pair-of-pants (i.e., $S^2 \setminus P$ for $|P| = 3$) in Σ disjoint from the other curves in $\boldsymbol{\alpha}$. (This corresponds to sliding the 3-dimensional 2-handle corresponding to α_i over the handle corresponding to α_j in the 3-manifold defined by the diagram.) We can define a handle slide among $\boldsymbol{\beta}$ analogously. Then we have the following strengthening of the Reidemeister–Singer theorem:

Theorem 2.2.11. *Suppose that $\mathcal{H}$ and $\mathcal{H}'$ are embedded Heegaard diagrams of the closed, connected, oriented 3-manifold M. Then they become isotopic after a sequence of handle slides, stabilisations, and destabilisations.*

2.3. Incompressible surfaces and the loop theorem

Definition 2.3.1. Let S be a surface in a 3-manifold M. Then we say that S is *compressible* either if there is a disk $D \subset M$ such that $D \cap S = \partial D$ and ∂D does not bound a disk in S, or if S has an S^2 component that is the boundary of a 3-ball in M, or if S has a boundary-parallel D^2 component. We say that S is *incompressible* otherwise. We call such a D a *compressing disk*.

Given a compressing disk D for a surface S in the 3-manifold M, we can compress it along D as follows: Let $N(D) \approx D \times [-1, 1]$ be a regular neighbourhood of D such that $N(D) \cap S \approx \partial D \times [-1, 1]$. Then the compressed surface is

$$S(D) := S \setminus (\partial D \times (-1, 1)) \cup (D \times \{-1, 1\}).$$

One can think of this as a surgery operation for embedded surfaces.

Definition 2.3.2. A surface S in a 3-manifold M is called *2-sided* if its normal bundle is trivial. We say that S is *π_1-injective* if the map $\pi_1(S) \to \pi_1(M)$ induced by the inclusion is injective.

For example, if both M and S are orientable, then S is 2-sided. The following fundamental result was first stated by Dehn in 1910, but gaps in his proof were found by Kneser. The first complete proof was given by Papakyriakopoulos [**55**] in 1957.

Theorem 2.3.3 (Dehn's lemma). *Let M be a 3-manifold with boundary, and $f\colon D^2 \to M$ a continuous map that is an embedding near ∂D^2 and $f(\partial D^2) \subset \partial M$. Then there exists an embedding $g\colon D^2 \to M$ such that f and g agree in a neighbourhood of ∂D^2.*

The following generalisation of Dehn's lemma is known as the loop theorem, also due to Papakyriakopoulos:

Theorem 2.3.4 (Loop theorem). *Let M be a 3-manifold with boundary (not necessarily compact). If ∂M is not π_1-injective, then it is compressible.*

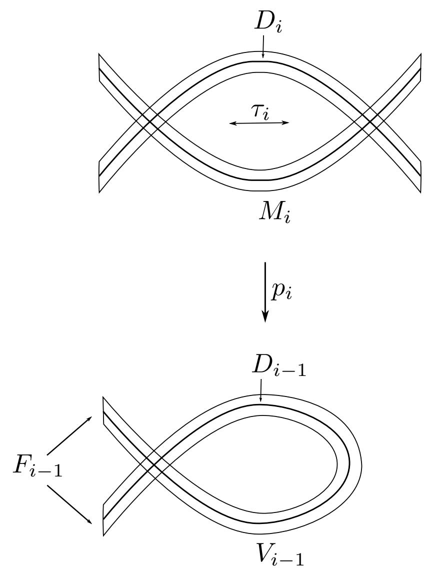

FIGURE 4. The double cover $p_i\colon M_i \to V_{i-1}$ in the proof of the loop theorem. Here, V_{i-1} is a regular neighbourhood of the disk D_{i-1} and τ_i is the covering involution.

PROOF (NON-EXAMINABLE). We follow the exposition of Hatcher [**21**]. Let $f\colon (D^2, \partial D^2) \to (M, \partial M)$ be a continuous map such that $[f|_{\partial D^2}] \neq 1 \in \pi_1(\partial M)$. Choose a triangulation of M. By the relative version of the simplicial approximation theorem, there is a triangulation of D^2 such that f is homotopic to a simplicial map $f_0\colon (D^2, \partial D^2) \to (M, \partial M)$. Let $D_0 := f_0(D^2)$, and V_0 a regular neighbourhood of D_0 obtained by taking the union of the simpices of the second barycentric subdivision of M incident to D_0.

Let $p_1\colon M_1 \to V_0$ be a connected double cover of V_0, if one exists. See Figure 4. As $\pi_1(D^2) = 1$, we can lift f_0 to a map $f_1\colon D^2 \to M_1$. We write $D_1 := f_1(D^2)$ and V_1 for a regular neighbourhood of D_1, as above. We repeat this procedure to obtain a sequence of maps $f_i\colon D^2 \to M_i$ whose image is D_i with regular neighbourhood V_i, where $p_i\colon M_i \to V_{i-1}$ is a connected double cover. We can triangulate M_i by lifting the triangulation of V_{i-1}, and D_i and V_i are subcomplexes.

We claim that this process terminates in finitely many steps, and write n for the largest index for which V_n admits no connected double cover. Indeed, let $E_i := p_i^{-1}(D_{i-1})$, and note that the double cover $E_i \to D_{i-1}$ is connected as M_i is connected and deformation retracts onto E_i by lifting the retraction of V_{i-1} onto D_{i-1}. If τ_i is the covering automorphism of $E_i \to D_{i-1}$, then $E_i = D_i \cup \tau_i(D_i)$. As E_i is connected, $D_i \cap \tau_i(D_i) \neq \emptyset$. So there is a simplex σ of D_i such that $\tau_i(\sigma) \subset D_i$. As τ_i has no fixed points, $\sigma \neq \tau_i(\sigma)$. This means that D_{i-1}, which is a quotient of D_i, has fewer simplices than D_i. The number of simplices in each D_i is bounded from above by the number of simplices in D^2, which hence gives an upper bound on the number of steps n.

We now show that each component of ∂V_n is a 2-sphere. It suffices to prove that $H_1(\partial V_n; \mathbb{Z}_2) = 0$. As V_n has no connected double cover, $\pi_1(V_n)$ has no index two subgroup, and hence $H^1(V_n; \mathbb{Z}_2) \cong \mathrm{Hom}(\pi_1(V_n), \mathbb{Z}_2) = 0$. So, by the universal

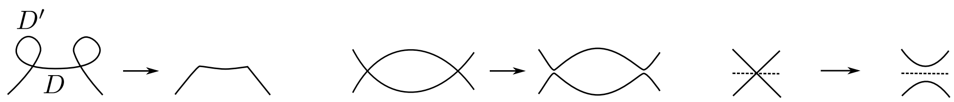

FIGURE 5. Removing double point curves of an immersed disk. Left: the preimages of the double point curve are nested circles. Middle: the preimages of the double point curve are non-nested circles. Right: the preimage of the double point of curve is connected.

coefficient theorem, $H_1(V_n;\mathbb{Z}_2) = 0$, and by Poincaré duality, $H_2(V_n, \partial V_n;\mathbb{Z}_2) = 0$. Combined with the exact sequence

$$H_2(V_n, \partial V_n;\mathbb{Z}_2) \to H_1(\partial V_n;\mathbb{Z}_2) \to H_1(V_n;\mathbb{Z}_2)$$

of the pair $(V_n, \partial V_n)$, we see that $H_1(\partial V_n;\mathbb{Z}_2) = 0$, as claimed.

Let $\partial_0 V_i$ be the component of ∂V_i containing $f_i(\partial D^2)$, and let

$$F_i := (p_1 \circ \cdots \circ p_i)^{-1}(\partial M) \cap \partial_0 V_i.$$

We denote the kernel of the homomorphism $(p_1 \circ \cdots \circ p_i)_* \colon \pi_1(F_i) \to \pi_1(\partial M)$ by N_i. As $[f|_{\partial D^2}] \neq 0 \in \pi_1(\partial M)$, we have $[f_i|_{\partial D^2}] \notin N_i$.

Since each component of ∂V_n is a sphere, the surface F_n is planar. Hence $\pi_1(F_n)$ is normally generated by the components of ∂F_n. As $N_n \neq \pi_1(F_n)$, there is a component C of ∂F_n that represents an element of $\pi_1(F_n) \setminus N_n$. Each component of $\partial_0 V_n \setminus F_n$ is a disk. We can push the interior of the disk bounding C into V_n to obtain a smooth embedding $g_n \colon D^2 \hookrightarrow V_n$ such that $[g_n|_{\partial D^2}] \notin N_n$.

Starting from g_n, we recursively construct a sequence of embeddings $g_i \colon D^2 \hookrightarrow V_i$ with $[g_i|_{\partial D^2}] \notin N_i$. Then g_0 is the desired embedding of D^2 into M whose boundary is homotopically non-trivial.

Suppose we have already constructed g_i. To obtain g_{i-1}, we consider the immersion $p_i \circ g_i$. After a small perturbation of g_i, this is an immersion with transverse double point curves and arcs, since p_i is a 2-fold cover. We use cut-and-paste techniques to get rid of these, resulting in the embedding g_{i-1}.

Let C be a double point curve of $p_i \circ g_i(D^2)$. Let $N(C)$ be a regular neighbourhood of C in $p_i \circ g_i(D^2)$. Then $N(C)$ is an X-bundle over S^1. This can be obtained from $X \times [0,1]$ by idenfitying $X \times \{0\}$ and $X \times \{1\}$ via an automorphism φ of the figure X. This automorphism cannot be a 90° rotation or a reflection in one of the two lines in X as otherwise D^2 would contain a Möbius band. So φ is either the identity of the figure X, or a reflection in a vertical or horizontal line.

First, suppose that φ is the identity. Then the preimage of C in D^2 consists of two circles that bound disks D and D'. If these disks are nested, say $D \subset D'$, then we replace $p_i \circ g_i|_{D'}$ with $p_i \circ g_i|_D$ and smooth the corners; see the left of Figure 5. If $D \cap D' = \emptyset$, then we swap $p_i \circ g_i|_D$ and $p_i \circ g_i|_{D'}$ and again smooth the corners; see the middle of Figure 5. If φ is a reflection in the horizontal axis, we smooth each X-fibre as on the right of Figure 5, which replaces the annulus mapping to $N(C)$ with another annulus. The case of reflection in a vertical axis is analogous. In each case, we remove C from the double point set, without introducing any new double point curves or changing $p_i \circ g_i|_{\partial D^2}$.

Now suppose that A is a double point arc of $p_i \circ g_i(D^2)$. Its preimage in D^2 consists of a pair of arcs a, $a' \subset D^2$ with boundary on ∂D^2; see the left of Figure 6. We label the components of $\partial D^2 \setminus (a \cup a')$ by α, β, γ, and δ counterclockwise such

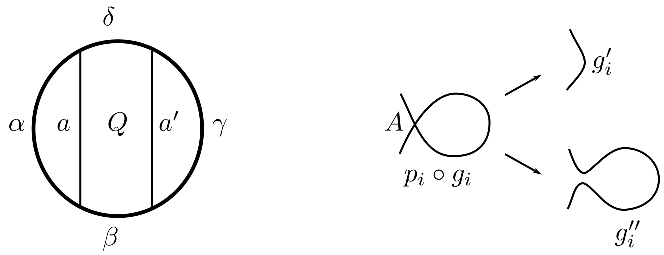

FIGURE 6. Left: The preimage of a double point arc A. Right: The two resolutions g_i' and g_i''.

that a and α and a' and γ form bigons B and B' (and hence a, β, a', δ are sides of a quadrilateral Q).

There are two ways of smoothing the double point arc A; see the right of Figure 6. In the first case, we glue together the maps $p_i \circ g_i|_B$ and $p_i \circ g_i|_{B'}$, smooth the corners, and remove the quadrilateral Q, to obtain the map g_i'. The other smoothing of A gives rise to a map we denote g_i''. This is obtained by removing $p_i \circ g_i|_Q$ and vertically reversing it, gluing the side a of B to the side a' of Q, and the side a' of B' to the side a of Q.

We claim that either $[g_i'|_{\partial D^2}] \notin N_{i-1}$ or $[g_i''|_{\partial D^2}] \notin N_{i-1}$. Choose an orientation of A; this induces orientations on a and a'. If exactly one of them is oriented coherently with ∂Q, then

$$\begin{aligned}\alpha\beta\gamma\delta &= (\alpha\gamma)\delta^{-1}(\alpha\beta^{-1}\gamma\delta^{-1})^{-1}(\alpha\gamma)\delta \\ &= (g_i'|_{\partial D^2})\delta^{-1}(g_i''|_{\partial D^2})^{-1}(g_i'|_{\partial D^2})\delta.\end{aligned}$$

Otherwise, we have

$$\begin{aligned}\alpha\beta\gamma\delta &= (\alpha\gamma^{-1})(\gamma\delta)^{-1}(\alpha\gamma^{-1})^{-1}(\alpha\delta\gamma\beta)(\gamma\delta) \\ &= (g_i'|_{\partial D^2})(\gamma\delta)^{-1}(g_i'|_{\partial D^2})^{-1}(g_i''|_{\partial D^2})(\gamma\delta).\end{aligned}$$

As $\alpha\beta\gamma\delta \notin N_{i-1}$, at least one of $[g_i'|_{\partial D^2}]$ and $[g_i''|_{\partial D^2}]$ is not in N_{i-1}, as claimed, and we continue with this map until $p_i \circ g_i$ becomes an embedding. $\square$

It is clear that every π_1-injective surface without S^2 components that bound 3-balls and without boundary-parallel D^2 components is incompressible. The converse also holds for 2-sided surfaces:

THEOREM 2.3.5. *If S is a 2-sided, incompressible surface in a 3-manifold M, then S is π_1-injective.*

PROOF. By contradiction, suppose that there is a homotopically non-trivial curve $\gamma\colon S^1 \to S$ that is null-homotopic in M. Then there is a smooth map $u\colon D^2 \to M$ transverse to S such that $u|_{S^1} = \gamma$. The preimage $u^{-1}(S)$ is a smooth 1-manifold C in D^2. Let C_0 be an innermost component of C. Let $D_0 \subset D^2$ be the disk bounded by C_0. If $u|_{C_0}$ is null-homotopic in S, then we can replace u on D_0 with the null-homotopy and push it off S such that we obtain a map v with $v^{-1}(S) = C \setminus C_0$. Repeating this procedure, we can assume that $u|_{C_0}$ is not null-homotopic. Note that $u(D_0) \cap S = u(C_0)$. Hence, we can assume the curve γ on S is null-homotopic in the complement of S.

Since S is 2-sided, $N(S) \approx S \times [-1,1]$. The curve $\gamma \times \{1\}$ is null-homotopic in $M' := M \setminus (S \times (-1,1))$. Hence, by Theorem 2.3.4 applied to M', the surface $S \times \{1\}$ is compressible in M', so S is compressible in M, which is a contradiction. $\square$

EXERCISE 2.3.6. Let K denote the Klein bottle, and let $K \widetilde{\times} I$ be an orientable I-bundle over K. Then $\partial K \widetilde{\times} I$ is a 2-torus T. Show that we can attach $S^1 \times D^2$ to $K \widetilde{\times} I$ along T such that in the resulting 3-manifold M, the Klein bottle K is incompressible but not π_1-injective.

2.4. Haken manifolds

In this section, we review some results on a class of particularly nice 3-manifolds called Haken manifolds that includes knot exteriors. They were used by Haken in his unknot detection algorithm.

DEFINITION 2.4.1. A 3-manifold M is *irreducible* if every embedded 2-sphere in M bounds a 3-ball. We say that M is *boundary-irreducible* if each component of ∂M is incompressible in M.

An oriented 3-manifold M is *Haken* or *sufficiently large* if it is irreducible, and contains a properly embedded, orientable, incompressible surface.

We will need a result that is known as the "half lives, half dies lemma."

LEMMA 2.4.2. *Let M be a compact, oriented 3-manifold. Then*

$$rk(\ker(H_1(\partial M) \to H_1(M))) = \frac{rk(H_1(\partial M))}{2}.$$

PROOF. From the homological and cohomological long exact sequences of the pair $(M, \partial M)$ with real coefficients, we obtain the commutative diagram

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

where the vertical arrows are Poincaré duality isomorphisms. Hence

$$\begin{aligned}
\dim \operatorname{coker}(i^*) = \dim \operatorname{coker}(\partial_*) &= \dim H_1(\partial M;\mathbb{R}) - \dim \operatorname{Im}(\partial_*) \\
&= \dim H_1(\partial M;\mathbb{R}) - \dim \ker(i_*).
\end{aligned} \tag{2.4.1}$$

Since we are working over the filed $\mathbb{R}$, we have $i^* = \operatorname{Hom}(i_*, \mathbb{R})$, and hence

$$\dim \ker(i_*) = \dim \operatorname{coker}(i^*).$$

Together with equation (2.4.1), we obtain that $2 \dim \ker(i_*) = \dim H_1(\partial M;\mathbb{R})$, as claimed. $\square$

PROPOSITION 2.4.3. *Let M be a compact, oriented, irreducible 3-manifold with boundary with $b_1(M) > 0$. Then M is Haken.*

PROOF. Since $b_1(M) > 0$, we have

$$H^1(M) = \operatorname{Hom}(H_1(M), \mathbb{Z}) \neq 0.$$

If $a \in H^1(M) \setminus \{0\}$, then there is a map $f \colon M \to S^1$ such that $f^*(1) = a$, where 1 is the generator of $H^1(S^1)$ dual to the fundamental class of S^1, using the identification between $H^1(S^1)$ and the set of homotopy classes $[M, S^1]$. Let x be a regular value of f. Then $S := f^{-1}(x)$ is Poincaré dual to a. It is orientable since we can pull back $\partial/\partial\theta \in T_1S^1$ to give a trivialisation of the normal bundle of S, and M is orientable. After compressing S, we obtain a properly embedded, orientable, incompressible surface S' homologous to S. Since M is irreducible, every S^2 component of S' bounds a ball. So we can remove these and all boundary-parallel D^2 components to obtain a surface S'' without changing the homology class. As $[S''] \neq 0$, we have $S'' \neq \emptyset$. So S'' is incompressible, and hence M is Haken. $\square$

A Haken manifold M admits a *Haken hierarchy*, where we successively cut it along incompressible surfaces until we obtain 3-balls. This makes them amenable to inductive proofs. Let S be a properly embedded, orientable, incompressible surface in M. Then we cut M along S; i.e., consider $M_1 := M \setminus N(S)$. If a component M_1' of M_1 has a boundary component that is not a sphere, it has $b_1(M_1') > 0$ by the half lives, half dies lemma (Lemma 2.4.2), and is hence also Haken by Proposition 2.4.3. So we can cut M_1 along an incompressible surface. We can continue this procedure until all boundary components are spheres. As M was irreducible, each such sphere bounds a ball in M, and so all the resulting components are balls. The fact that this procedure terminates in finitely many steps relies on Theorem 2.5.8 below; see Hempel [**24**, Theorem 13.3].

According to the following result of Waldhausen [**72**], Haken manifolds are determined by their fundamental groups:

THEOREM 2.4.4. *Let M and N be compact, orientable 3-manifolds that are irreducible and boundary-irreducible. Suppose M is Haken. Let $\psi\colon \pi_1(N) \to \pi_1(M)$ be an isomorphism that "respects the peripheral structure;" i.e., for each component F of ∂N there exists a component G of ∂M such that $\psi(i_*(\pi_1(F))) \subset A$ and A is conjugate in $\pi_1(M)$ to $i_*(\pi_1(G))$, where the i_* denote inclusion homomorphisms. Then there exists a homeomorphism $f\colon N \to M$ which induces ψ.*

The following deep result is due to Agol [**1**], building on work of Kahn and Markovic [**29**][**30**], which was known as the virtually Haken conjecture:

THEOREM 2.4.5. *Every compact, orientable, irreducible 3-manifold with infinite fundamental group is virtually Haken; i.e., has a finite cover that is Haken.*

2.5. Normal surfaces and prime decomposition

This section is based on unpublished notes of Casson [**5**].

DEFINITION 2.5.1. We say that the closed 3-manifold M is *prime* if, whenever $M = A\#B$, we have $A \approx M$ and $B \approx S^3$, or $A \approx S^3$ and $B \approx M$.

Every irreducible 3-manifold is clearly prime. Conversely, we have the following:

EXERCISE 2.5.2. Show that if a closed 3-manifold is prime, then it is either irreducible, or $S^1 \times S^2$, or the non-orientable S^2-bundle over S^1.

Normal surfaces were introduced by Kneser [**35**] in order to prove that every closed 3-manifold can be written as a connected sum of prime 3-manifolds, and further developed by Haken [**18**] into what is now known as normal surface theory. It is a fundamental tool for many 3-manifold algorithms, including the following result due to Haken [**19**], Waldhausen [**72**], Jaco–Shalen [**26**], Johannson [**27**], and Hemion [**23**]:

FIGURE 7. A normal surface intersects each tetrahedron in triangles (left) or quads (right).

THEOREM 2.5.3. *There is an algorithm to decide whether two Haken 3-manifolds are homeomorphic.*

PROOF. A knot exterior is a Haken 3-manifold, so we can use Theorem 2.5.3 to decide whether it is homeomorphic to $S^1 \times D^2$, the complement of the unknot. □

While most of this book focuses on differential topological methods, normal surface theory is specific to the PL category of triangulated 3-manifolds. In dimension 3, every topological 3-manifold has a unique smooth structure and a unique PL structure, so the two approaches are in some sense equivalent.

Let M be a 3-manifold with triangulation T. We write T^i for the i-skeleton of T. Let F be a closed surface in M; not necessarily connected. We can make F transverse to T^i for every i.

DEFINITION 2.5.4. We say that the surface F in the 3-manifold M with triangulation T is *normal* if, for every 3-simplex τ of T, each component of $F \cap \tau$ is either

(1) a *triangle* with vertices on three edges meeting at a vertex of τ, one vertex on each edge, as on the left of Figure 7, or
(2) a quadrilateral (*quad*) with vertices on four edges that are adjacent to an edge of τ, one on each edge, as on the right of Figure 7.

There are four types of triangles, corresponding to the four vertices of τ, and three types of quads, corresponding to disjoint pairs of edges of τ. For each τ, there might be several components of $F \cap \tau$ of each triangle type, but at most one type of quad as otherwise F would have double points.

DEFINITION 2.5.5. Let $S \subset M$ be a union of 2-spheres. We say that these spheres are *independent* if no component of $M \setminus S$ is homeomorphic to $S^3 \setminus P$ for $P \subset S^3$ finite.

LEMMA 2.5.6. *Let S be a subsurface of the 3-manifold M with triangulation T.*

(1) If S is incompressible and M is irreducible, then S is isotopic to a normal surface.
(2) If S is a collection of k independent 2-spheres, then M also contains an independent set of k 2-spheres that is normal with respect to T.

PROOF. We first prove statement (1). We perturb S such that it becomes transvere to all simplices of T. We then define the *weight* of S to be $w(S) := |S \cap T^1|$. Suppose that S minimises $w(S)$ in its isotopy class. Then we claim that, for each 3-simplex Δ of T, every component C of $S \cap \partial\Delta$ is a simple closed curve that either

(i) lies in a 2-simplex of T,

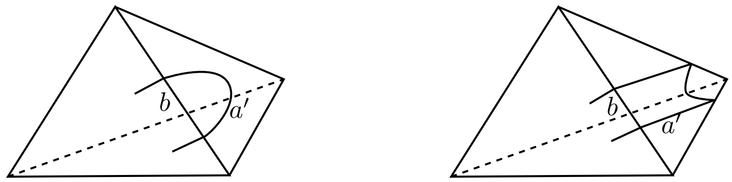

FIGURE 8. A component C of $S \cap \partial\Delta$ that is not of the required form admits one of these configurations.

(ii) intersects three edges of T exactly once that meet at a vertex of Δ (see the left of Figure 7), or
(iii) intersects four edges of T exactly once that are adjacent to an edge of Δ (see the right of Figure 7).

If C is not of this form, then there is a sub-arc b of an edge of Δ and a sub-arc a' of C such that $|\mathrm{Int}(a') \cap T^1| \in \{0, 3\}$; see Figure 8. The curve $a' \cup b$ bounds a bigon D' in $\partial\Delta$ that is either disjoint from the 1-skeleton, or contains a single vertex of Δ.

In either case, let a be an arc in $S \cap \Delta$ such that $a \cap \partial\Delta = \partial a'$, and such that a and a' are parallel. Then there is a disk D in Δ parallel to D' such that $\partial D \cap S = a$ and $\partial D \cap \partial\Delta = b$. (Note that the interior of D might intersect S.) Then a Whitney move (see Section 1.8) supported in $N(D)$ eliminates the intersection points ∂b from $S \cap T^1$, contradicting the minimality of $w(S)$.

Let $m(S)$ be the sum of $|S \cap \delta|$ over 2-simplices δ of T, and isotope S such that $m(S)$ is minimal. Suppose that C is an innermost component of $S \cap \partial\Delta$ for a 3-simplex Δ that lies in a 2-simplex of Δ; i.e., C is of type (i). Then C bounds a disk D in the complement of S. Since S is incompressible, this means that C also bounds a disk D' in S. Then $D \cup D'$ is a 2-sphere that bounds a 3-ball B in M, since M is irreducible. Hence, we can eliminate C from the intersection of S with 2-simplices of T by isotoping S across B, decreasing $m(S)$. Hence we can assume that no component of $S \cap \partial\Delta$ is of the form (i).

Now suppose that there is a 3-simplex Δ of T such that a component F of $S \cap \Delta$ is not a disk. Let C be a component of ∂F; this bounds a disk D' in $\partial\Delta$. We choose F, C, and D' such that D' does not contain in its interior the boundary of a non-disk component of $S \cap \Delta$. Then there is a disk $D \subset \Delta$ with $D \cap S = \partial D = C$. Since S is incompressible, we can again eliminate C by an isotopy of S, decreasing $m(S)$. Hence S is a normal suface when $m(S)$ is minimal. This concludes the proof of statement (1).

The proof of statement (2) proceeds similarly, with the exception of the last two paragraphs. Suppose that C is an innermost component of $S \cap \partial\Delta$ of type (i). Let the components of S be $S_1, \dots, S_k$, and suppose that C lies in S_1. If we compress S_1 along D, we obtain the 2-spheres S_1' and S_1''. Then at least one of $S' = S_1' \cup S_2 \cup \dots \cup S_k$ and $S'' = S_1'' \cup S_2 \cup \dots \cup S_k$ is independent. Indeed, suppose there were components B' and B'' of $M \setminus S'$ and $M \setminus S''$, respectively, that were punctured spheres. Since $S_1, \dots, S_k$ are independent, we must have $S_1' \subset \partial B'$ and $S_1'' \subset \partial B''$. If $S_1 \subset B'$ or $S_1 \subset B''$, then S would not be independent by the Schönflies theorem (Theorem 2.1.1). Otherwise, $B' \cap B'' = D$, so $B' \cup B''$ would be a punctured sphere component of $M \setminus S$, which is again impossible. Note that $m(S') < m(S)$ and $m(S'') < m(S)$. Hence, when $m(S)$ is minimal, there is no component C of $S \cap \partial\Delta$ that lies in a 2-simplex. Similarly, we can arrange that $S \cap \Delta$ has only disk components for every 3-simplex Δ of T. □

DEFINITION 2.5.7. Let S and S' be disjoint embedded surfaces in the 3-manifold M. Then we say that S and S' are *parallel* if $M \setminus (S \cup S')$ has a component C with $\partial C = S \cup S'$ and $\overline{C} \approx S \times I$ (in particular, $S \approx S'$).

THEOREM 2.5.8. *For any compact, irreducible 3-manifold M, there exists a constant $h(M) \in \mathbb{N}$, such that, for any incompressible surface S in M with $|S| \geq h(M)$, there are two components of S that are parallel.*

PROOF. Let T be a triangulation of M with t tetrahedra. Then we let

$$h(M) := 6t + 2b_2(M),$$

where $b_2(M) = \dim H_2(M; \mathbb{Z}_2)$. By statement (1) of Lemma 2.5.6, we can isotope S such that it becomes normal with respect to T.

If Δ is a 3-simplex of T, a component of $\partial\Delta \setminus S$ is called *good* if it is an annulus, and is *bad* otherwise. At most six components of $\partial\Delta \setminus S$ are bad. (The worst-case scenario is when we have all four types of triangles and one type of quad in Δ, in which case the four components containing the vertices and two components between triangles and quads are bad.) A component X of $M \setminus S$ is *good* if every component of $X \cap \partial\Delta$ is good for every 3-simplex Δ of T, and is *bad* otherwise. At most $6t$ components of $M \setminus S$ are bad. If $|S| \geq 6t + 2b_2(M)$, then $|M \setminus S| \geq 6t + 1 + b_2(M)$, so there are at least $1 + b_2(M)$ good components.

A good component X is obtained by gluing regions homeomorphic to triangle times I and square times I along their edges times I. So X is an I-bundle over a surface. Every non-trivial I-bundle contributes one $\mathbb{Z}_2$ direct summand to $H_2(M; \mathbb{Z}_2)$, so there is at least one trivial I-bundle X. The boundary components of X are parallel components of S. □

This finiteness result allows one to enumerate all normal surfaces by writing down so called matching equations. The following result and corollary are due to Kneser [**35**]:

THEOREM 2.5.9. *Let M be a triangulated 3-manifold with t 3-simplices. If M contains an independent set of k 2-spheres, then $k \leq 6t + 2b_2(M)$.*

PROOF. The proof is analogous to the proof of Theorem 2.5.8, except now we invoke statement (2) of Lemma 2.5.6. □

COROLLARY 2.5.10. *Every closed 3-manifold can be expressed as a connected sum of finitely many prime 3-manifolds.*

Uniqueness was shown 30 years later by Milnor [**42**].

THEOREM 2.5.11. *Let M be a closed oriented 3-manifold. If $M \approx M_1 \# \dots \# M_k$ and $M \approx N_1 \# \dots \# N_l$ with M_i, N_j prime and not S^3, then $k = l$, and after reindexing, M_i and N_i are orientation-preserving homeomorphic.*

PROOF (NON-EXAMINABLE). First, suppose that every two-sphere in M is separating. Then every M_i and N_j is irreducible. Let S be the union of the connected sum spheres in $M_1 \# \dots \# M_k$. If Σ is a two-sphere separating N_1 and $N_2 \# \dots \# N_l$, then we can isotope it such that it is transverse to S. Furthermore, we choose S

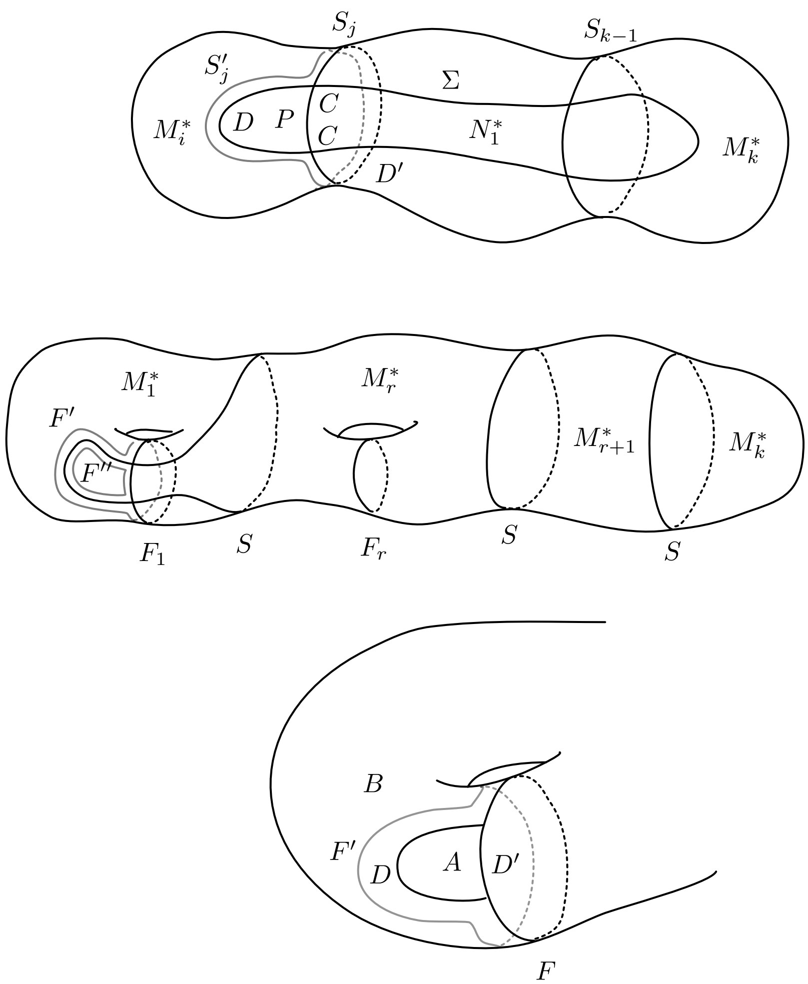

FIGURE 9. Illustrations for the proof of uniqueness of prime decompositions of three-manifolds.

to minimise $|S \cap \Sigma|$. We write $M_1^*, \ldots, M_k^*$ for the closures of the components of $M \setminus S$, where M_i^* is a punctured M_i. We define $N_1^*, \ldots, N_l^*$ analogously.

If $S \cap \Sigma \neq \emptyset$, then let C be a component of $S \cap \Sigma$ innermost in Σ, bounding a disk $D \subset \Sigma$ with $D \cap S = C = \partial D$; see the top of Figure 9. Suppose $D \subset M_i^*$. Since M_i is irreducible, one of the components of $M_i^* \setminus D$ has closure a punctured ball P. Suppose that C lies in a component S_j of ∂M_i^*, and let $D' = \overline{S_j \setminus P}$. Replace S_j by $D \cup D'$, pushed slightly to eliminate the intersection component C. This contradicts the choice of S to minimise the number of components of $S \cap \Sigma$.

So $S \cap \Sigma = \emptyset$. Suppose that $S \cap N_1^* \neq \emptyset$. Since N_1 is irreducible, some component of S is parallel to Σ and $M_i^* \subset N_1^*$ with $N_1^* \setminus M_i^* \approx S^2 \times I$ for some $i \in \{1, \ldots, k\}$. If $S \cap N_1^* = \emptyset$, then $N_1^* \subset M_i^*$ for some i. Since M_i is irreducible, $\overline{M_i^* \setminus N_1^*}$ is a punctured ball. In either case, $M_i \approx N_1$ and

$$M_1 \# \ldots \# M_{i-1} \# M_{i+1} \# \ldots \# M_k \approx N_2 \# \ldots \# N_l.$$

Hence, the result follows by induction.

Now suppose that F is a maximal collection of two-spheres in M such that $M \setminus F$ is connected, and let $r = |F|$. Furthermore, let S be as above. If $F \cap S \neq \emptyset$, then we choose a component C of $F \cap S$ innermost in S, and compress F along the disk that C bounds in S. This gives rise to collections of spheres F' and F'', one of which – say F' – does not separate; see the middle of Figure 9. Then we can isotope F' to eliminate C. Repeating this procedure, we obtain a sequence of nonseparating collections of spheres $F = F^0, F^1, \dots, F^n$ such that $F^n \cap S = \emptyset$. If $F_j^n \subset M_i^*$ for a component F_j^n of F^n, then $M_i \approx S^1 \times S^2$ by Exercise 2.5.2, since M_i is prime but not irreducible as it contains the nonseparating sphere F_j^n. Hence, after reindexing, $M_1 \approx \dots \approx M_r \approx S^1 \times S^2$. Furthermore, $M \setminus N(F^n)$ is a $2r$-punctured $M_{r+1} \# \dots \# M_k$. The same applies to the factorisation $N_1 \# \dots \# N_l$.

We claim that $\overline{M \setminus N(F)} \approx \overline{M \setminus N(F^n)}$. It suffices to show that, if D is a disk in M with $D \cap F = \partial D$ such that ∂D bounds a disk $D' \subset F$ and $F' := \overline{(F \setminus D')} \cup D$ does not separate M, then $\overline{M \setminus N(F)} \approx \overline{M \setminus N(F')}$; see the bottom of Figure 9. Indeed, since F was maximal, $D \times I$ separates $\overline{M \setminus N(F)}$ into two components, whose closures we denote by A and B. Then $\overline{M \setminus N(F)} \approx A \cup_D B$ and $\overline{M \setminus N(F')} \approx A \cup_{D'} B$. These are homeomorphic since both are obtained from $A \sqcup B$ by gluing disks in their boundaries.

We conclude that a $2r$-punctured $M_{r+1} \# \dots \# M_k$ is homeomorphic to a $2r$-punctured $N_{r+1} \# \dots \# N_l$, as they are both homeomorphic to $M \setminus N(F)$. Hence,

$$M_{r+1} \# \dots \# M_k \approx N_{r+1} \# \dots \# N_l,$$

and it does not contain a separating two-sphere, so the result now follows from the first part of the proof. $\square$

CHAPTER 3

Knots and links

Knots and links play a fundamental role in low-dimensional topology. Their exteriors provide an important class of 3-manifolds. One can obtain all 3-manifolds using surgery on links in S^3. Furthermore, all 4-manifolds can be obtained from traces of surgeries on links in connected sums of copies of $S^1 \times S^2$ (after adding 3-handles and a 4-handle in a unique way); see Section 4.1 on Kirby calculus for more information. Invariants of 3- and 4-manifolds are often easier to understand for links, and some of them are in fact defined as extensions of link invariants.

3.1. Knots and links

Definition 3.1.1. A *k-component link* in a 3-manifold Y is a smooth embedding

$$L\colon \bigcup_{i=1}^{k} S^1 \times \{i\} \hookrightarrow Y.$$

The links L_0 and L_1 are *equivalent* if they are ambient isotopic; i.e., there is an isotopy $\Phi\colon Y \times I \to Y$ such that $\Phi_0 = \mathrm{Id}_Y$ and $\Phi_1 \circ L_0 = L_1$. A *knot* is a 1-component link.

One could also define piecewise linear (PL) knots and links, whose study is amenable to combinatorial techniques. However, we will take the differential topological point of view. The two are equivalent for links.

The image $\mathrm{Im}(L)$ is a smooth 1-dimensional submanifold of M that is oriented such that the map L is orientation-preserving. Given an orientation of $\mathrm{Im}(L)$, the parametrisation L is uniquely determined up to isotopy as $\mathrm{Diff}^+(S^1)$ is connected. Hence, we will often think of a link as an oriented smooth 1-dimensional submanifold of M. We will denote by $\overline{L}$ the link L with its orientation reversed, called the *reverse* of L.

Since $\pi_0(\mathrm{Diff}_+(S^3)) = 0$ by Cerf [**7**], any orientation-preserving automorphism of S^3 is isotopic to the identity. So two links in S^3 are ambient isotopic if and only if they are orientation preserving diffeomorphic. We denote by $-L$ the mirror of the link L, obtained by reflecting L in a plane.

We can represent a link in $\mathbb{R}^3$ by looking at its projection onto $\mathbb{R}^2$. If this projection is an immersion with only transverse double point singularities, and at each double point we record which strand is higher with respect to the coordinate z, we obtain a *link diagram*. This allows one to reconstruct the link up to equivalence.

Definition 3.1.2. A *link diagram* is an immersed, oriented 1-manifold in $\mathbb{R}^2$ or S^2 that has only transverse double point singularities, and at each double point, a designation of one of the two strands as the over-strand.

Every link has a diagram:

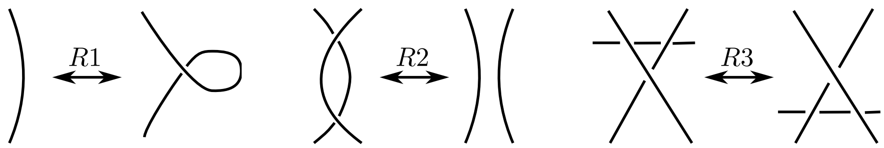

FIGURE 10. The three Reidemeister moves.

PROPOSITION 3.1.3. *Let L be a link in $\mathbb{R}^3$. Then there is an isotopy $\{\psi_t : t \in I\}$ of $\mathbb{R}^3$ such that $\psi_0 = Id_{\mathbb{R}^3}$ and orthogonal projection of $\psi_1(L)$ onto $\mathbb{R}^2$ is an immersion with only transverse double point singularities.*

The proof is similar to the proof of Theorem 1.4.8, which we hence omit.

3.2. Reidemeister moves

Given two link diagrams, they represent equivalent links if and only if they can be connected by a sequence of Reidemeister moves and planar isotopies. We describe these next. For an illustration, see Figure 10.

We say that two diagrams are related by a *planar isotopy* if there is an isotopy of $\mathbb{R}^2$ taking one diagram to the other.

Let $\mathcal{D}$ be a diagram of the link L. We can perform the *first Reidemeister move R1* if there is a component C of $\mathbb{R}^2 \setminus \mathcal{D}$ that contains a single crossing in its boundary. The resulting diagram is obtained by smoothing $\overline{\mathcal{D} \setminus \partial C}$, and has one fewer crossings than $\mathcal{D}$. We will also refer to the reverse of this move as an R1 move.

Now suppose that $\mathbb{R}^2 \setminus \mathcal{D}$ contains a component C that has two crossings along its boundary, and such that $\overline{C} \approx D^2$. We further assume that, as we traverse one of the two strands along ∂R, both crossings are either overcrossings, or they are both undercrossings. Then the *second Reidemeister move R2* applied to the component C removes a neighbourhood of C from $\mathcal{D}$, and reconnects the resulting two pairs of endpoints in $N(C)$ without intersection points, as in the middle of Figure 10. Again, we also call the reverse of this move an R2 move.

Finally, suppose that $\mathbb{R}^2 \setminus \mathcal{D}$ has a component C with $\overline{C} \approx D^2$ whose boundary contains three crossings, and one of the three strands along ∂C – call it s – has two undercrossings; see the right-hand side of Figure 10. Let c be the crossing along ∂C opposite s, and let C' be the component of $\mathbb{R}^2 \setminus \mathcal{D}$ that meets C at c. Then we can remove a small extension of s from $\mathcal{D}$, and reconnect its endpoints across C'. This is called the *third Reidemeister move R3*.

The following result is due to Reidemeister [**58**]:

THEOREM 3.2.1. *Let $\mathcal{D}_0$ and $\mathcal{D}_1$ be link diagrams. Then they represent equivalent links if and only if they can be connected by a sequence of planar isotopies and Reidemeister moves R1–R3.*

SKETCH OF PROOF. We denote by $\pi \colon \mathbb{R}^3 \to \mathbb{R}^2$ the projection. Let L_i be a link with diagram $\mathcal{D}_i$ for $i \in \{0,1\}$. Suppose there is an isotopy ϕ_t for $t \in I$ of $\mathbb{R}^3$ such that $\phi_0 = \mathrm{Id}_{S^3}$ and $\phi_1(L_0) = L_1$. We write $L_t := \phi_t(L)$. If the isotopy ϕ_t is generic, then $\pi(L_t)$ is an immersion with only transverse double points for every $t \in I$ outside a finite set $S \subset (0,1)$. Furthermore, for $s \in S$, the projection $\pi(L_s)$ has exactly one of the following singularities:

(1) a cusp, corresponding to a point $p \in L_t$ where T_pL_t is parallel to $\partial/\partial z$,
(2) a self-tangency modelled on the graphs of x^2 and $-x^2$,
(3) a triple point where any two branches of the curve are transverse.

We write $\mathcal{D}_t$ for the diagram of L_t for $t \in I \setminus S$. If t and t' lie in the same component of $I \setminus S$, then the diagrams $\mathcal{D}_t$ and $\mathcal{D}_{t'}$ are related by a planar isotopy. Furthermore, for $s \in S$ and ε sufficiently small, $\mathcal{D}_{s-\varepsilon}$ and $\mathcal{D}_{s+\varepsilon}$ are related by a Reidemeister move R1 in case (1), move R2 in case (2), and move R3 in case (3).

Conversely, every planar isotopy and Reidemeister move can be lifted to an isotopy of the knot. Furthermore, if two knots have the same diagram, they are isotopic via linearly interpolating between the z-coordinate functions. □

A corollary of the above theorem is that any quantity that is defined for link diagrams and is invariant under Reidemeister moves gives rise to a link invariant. The Jones polynomial, which we will encounter later, is a particularly interesting example of this construction.

Definition 3.2.2. Let $\mathcal{D}$ be a link diagram. We say that a crossing of $\mathcal{D}$ is *positive* if, as we traverse the over-strand following the orientation, the under-strand crosses from right to left, and is *negative* otherwise.

The *writhe* $w(\mathcal{D})$ of the diagram $\mathcal{D}$ is the number of positive crossings minus the number of negative crossings.

The *winding number* of the knot diagram $\mathcal{D}$ is the number of rotations the unit tangent vector of $\mathcal{D}$ makes as we traverse the knot.

Writhe and winding number are *not* knot invariants, but are only changed by move R1. Trace [**70**] showed that two diagrams of the same knot are related only using moves R2 and R3 if and only if they have the same writhe and winding number.

The writhe is closely related to the notion of linking. By Alexander duality, $H_1(S^3 \setminus K) \cong \mathbb{Z}$ for any knot K. Given disjoint knots K and K', their *linking number* $\mathrm{lk}(K, K')$ is the integer corresponding to $[K'] \in H_1(S^3 \setminus K)$. Given a diagram for the link $K \cup K'$, this can be computed by taking the number of positive crossings minus the number of negative crossings between K and K', divided by two.

Definition 3.2.3. The *crossing number* $c(\mathcal{D})$ of a diagram $\mathcal{D}$ of the link L is the number of its crossings. We define the crossing number $c(L)$ of the link L to be the minimum of $c(\mathcal{D})$ over all diagrams of L.

Knots and links are usually tabulated according to their crossing number. For example, the Rolfsen table of knots [**61**] tabulates all knots up to ten crossings. The *unknot* is represented by the standard circle S^1 in S^3, and is the only knot with crossing number zero. There is no knot of crossing number two. Up to mirroring, there is just one knot of crossing number three: the *trefoil* (see Figure 12), and one crossing number four knot: the *figure eight*. Even though the crossing number is easy to define, it is unknown whether it is additive under connected sums.

3.3. Seifert surfaces

Surfaces bounding knots in the 3-sphere play an important role in defining various knot invariants.

Definition 3.3.1. Let L be a link in the 3-manifold Y. A *Seifert surface* for L is a compact, oriented surface S embedded in Y without closed components such that $\partial S = L$.

Proposition 3.3.2. *Every link in S^3 admits a Seifert surface.*

Proof. Let $\mathcal{D}$ be a diagram of L. We resolve each crossing according to the orientation of $\mathcal{D}$, obtaining an embedded, oriented 1-manifold C in $\mathbb{R}^2$ whose components are called *Seifert circles*. We choose a collection S_0 of disjoint disks in $\mathbb{R}^3$ with boundary C. For each crossing in $\mathcal{D}$, we add a half-twisted band to S_0, obtaining the desired Seifert surface S. □

We say that the diagram $\mathcal{D}$ is *simple* if it has no nested Seifert circles. The above proof only works for links in S^3. More generally, we have the following result:

Proposition 3.3.3. *Let L be a closed 0-homologous oriented $(n-2)$-manifold in the n-manifold Y, and $a \in H_{n-1}(Y, L)$ a homology class such that $\partial a = [L] \in H_{n-2}(L)$, where $\partial\colon H_{n-1}(Y, L) \to H_{n-2}(L)$ is the boundary map in the long exact sequence of the pair (Y, L). Then there is a compact, connected, oriented $(n-1)$-manifold embedded in Y with boundary L representing a.*

Proof. Let $E := Y \setminus N(L)$ be the exterior of L. Since L is 0-homologous in Y, there is a trivialisation $\partial E \approx L \times S^1$ such that $L \times \{1\}$ is 0-homologous in E. Furthermore, there is a cohomology class $\alpha \in H^1(E)$ dual to $a \in H_{n-1}(E, \partial E) \cong H_{n-1}(Y, L)$ such that $\alpha(\mu) = 1$ for any meridian μ of L. Since the Eilenberg–MacLane space $K(\mathbb{Z}, 1) \cong S^1$, the map $i\colon [E, S^1] \to H^1(E)$ given by $i(f) := f^*(1)$ for $[f] \in [E, S^1]$ and $1 \in H^1(S^1)$ is an isomorphism. Let $f_\alpha\colon E \to S^1$ be a map such that $i(f_\alpha) = \alpha$ and $f_\alpha|_{\partial E}\colon L \times S^1 \to S^1$ is projection onto the second factor. We can perturb f_α such that it is smooth, and such that $1 \in S^1$ is a regular value. After taking the connected sum of the components of $f^{-1}(\{1\})$, it extends in $N(L)$ to a compact, connected, oriented manifold with boundary L representing a. □

Seifert surfaces are not unique. For example, if S is a Seifert surface of a link L in the 3-manifold Y, then one can always choose a 3-disk $D \subset Y$ such that $D \cap S \approx D^2 \sqcup D^2$, and replace $D \cap S$ with an unknotted annulus. This operation is called a *stabilisation*, and its reverse a *destabilisation*.

Proposition 3.3.4. *Let L be a 0-homologous link in the 3-manifold Y and $a \in H_2(Y, L)$ a homology class. Then any two Seifert surfaces of L representing the class a can be connected by a sequence of isotopies, stabilisations, and destabilisations.*

Proof. Let S_0 and S_1 be Seifert surfaces of L representing the class a, and write E for the link exterior. Choose maps f_0, $f_1\colon E \to S^1$ such that 1 is a common regular value of both, $f_0^{-1}(\{1\}) = S_0$ and $f_1^{-1}(\{1\}) = S_1$, and which are projections onto the S^1-factor along $\partial N(L) \approx L \times S^1$.

Let $\{f_t : t \in I\}$ be a generic 1-parameter family of smooth functions connecting f_0 and f_1 relative to ∂E. Then f_t is a circle-valued Morse function, except for finitely many values of t, when f_t has a single birth-death critical point p, and generically $f(p) \neq 1$; see equation (1.6.1).

Hence, the preimage $f_t^{-1}(\{1\})$ only changes when there is a nondegenerate critical point p of f_t with $f_t(p) = 1$. If 1 is a regular value of f_t, we write S_t for the connected sum of the components of $f_t^{-1}(\{1\})$. When the value of an index 0 or 3 critical point passes through 1, the preimage $f_t^{-1}(\{1\})$ changes by the birth or death of a small 2-sphere that bounds a 3-ball, hence $S_{t-\varepsilon}$ and $S_{t+\varepsilon}$ are isotopic. When the value of an index 1 or 2 critical point passes through 1, the level set $f_t^{-1}(\{1\})$ changes by adding or removing a tube. Hence, the surfaces $S_{t-\varepsilon}$ and $S_{t+\varepsilon}$ are related by an isotopy if the tube connects distinct components, and a stabilisation or destabilisation otherwise. □

REMARK 3.3.5. While a knot in S^3 has no unique Seifert surface, the normal framing given by a Seifert surface is well-defined, and is called the *Seifert framing*. This follows by applying the half lives half dies lemma to the knot exterior. The *blackboard framing* of a knot diagram is given by the normal in the plane of the diagram. The difference between the Seifert framing and the blackboard framing is the writhe of the diagram.

DEFINITION 3.3.6. Let K be a knot in S^3. Then its *Seifert genus* $g(K)$ is the minimum of $g(S)$, where S is a Seifert surface of K and $g(S)$ is the genus of S.

The Seifert genus detects the unknot, in the following sense:

LEMMA 3.3.7. *Let K be a knot in S^3. Then $g(K) = 0$ if and only if K is the unknot U.*

PROOF. Clearly, $g(U) = 0$. Conversely, suppose that $g(K) = 0$. Then there is an embedding $e\colon D^2 \hookrightarrow S^3$ such that $e|_{S^1} = K$. Then $e_t := e|_{(1-t)\cdot S^1}$ for $t \in [0, 1-\varepsilon]$ provides an isotopy of K to the curve $e_{1-\varepsilon}$. If ε is small enough, then we can linearly isotope $e_{1-\varepsilon}$ to the tangent plane $e_*(T_0 D^2)$ and obtain the unknot. □

The following celebrated result is due to Haken [**18**]:

THEOREM 3.3.8. *There is an algorithm to decide whether a knot in S^3 is trivial.*

PROOF. The only knot K in S^3 with exterior $S^1 \times D^2$ is the unknot, since $\{1\} \times \partial D^2$ has to be a longitude of K for homological reasons. So $g(K) = 0$ and $K = U$ by Lemma 3.3.7. A knot exterior is a Haken three-manifold, so we can use Theorem 2.5.3 to decide whether it is homeomorphic to $S^1 \times D^2$. □

Given knots K_1 and K_2, we can form their connected sum $K_1 \# K_2$, as follows: Choose diagrams D_1 and D_2, respectively, such that $D_1 \subset \mathbb{R}^2_+$ and $D_2 \subset \mathbb{R}^2_-$. Let $r\colon I \times I \hookrightarrow \mathbb{R}^2$ be an embedding such that $r(I \times I) \cap D_1 = r(\{0\} \times I)$ and $r(I \times I) \cap D_2 = r(\{1\} \times I)$, and the orientation along $\partial r(I \times I)$ is coherent with the orientations of D_1 and D_2. Then $K_1 \# K_2$ has diagram the symmetric difference $(D_1 \cup D_2) \triangle (\partial r(I \times I))$.

PROPOSITION 3.3.9. *Let K_1 and K_2 be knots in S^3. Then*

$$g(K_1 \# K_2) = g(K_1) + g(K_2).$$

PROOF. By considering the boundary connected sum of Seifert surfaces of K_1 and K_2, we obtain that $g(K_1 \# K_2) \leq g(K_1) + g(K_2)$.

We now show that $g(K_1 \# K_2) \geq g(K_1) + g(K_2)$. Let S be a Seifert surface for $K_1 \# K_2$. Make S transverse to the connected sum sphere C. Then $C \cap S$ is a 1-manifold with a single arc component, and a collection of circle components. Choose an innermost circle component γ of $C \cap S$, which bounds a disk $D \subset C$ such that $D \cap S = \gamma$. If we compress S along D, we obtain a surface S' such that $\chi(S') = \chi(S) + 2$. If S' has an S^2 component, we remove it and obtain a surface with Euler characteristic $\chi(S)$.

Repeating this process, we obtain a surface S'' with boundary K such that $S'' \cap C$ is a single arc and $\chi(S'') \geq \chi(S)$. Furthermore, S'' has no S^2 component. Hence, removing the closed components of S'', we obtain a Seifert surface S''' of K such that $\chi(S''') \geq \chi(S)$, and so $g(S''') \leq g(S)$. As $S''' \cap C$ consists of a single component, S''' is a boundary connected sum of Seifert surfaces of K_1 and K_2. Hence $g(S) \geq g(S''') \geq g(K_1) + g(K_2)$, as claimed. □

Definition 3.3.10. We say that a knot $K \neq U$ is *prime* if $K = K_1 \# K_2$ implies that $K_1 = U$ or $K_2 = U$.

Corollary 3.3.11. *Every nontrivial knot can be written as a connected sum of prime knots.*

Proof. We prove the claim by induction on $g(K)$. If K is not prime, then it can be written as $K = K_1 \# K_2$, where $K_1 \neq U$ and $K_2 \neq U$. Furthermore, $g(K_1) < g(K)$ and $g(K_2) < g(K)$ by Proposition 3.3.9 and Lemma 3.3.7. We can now write K_1 and K_2 as a connected sum of prime knots by the inductive hypothesis. □

Schubert [**62**] proved the uniqueness of prime decompositions of knots:

Theorem 3.3.12. *Any two prime decompositions of a knot are related by a permutation of the summands.*

3.4. The Seifert form

We now introduce the Seifert form, which is the source of a number of powerful knot invariants.

Definition 3.4.1. Let S be a Seifert surface of a knot K in S^3. We define the *Seifert form*

$$\langle \, , \, \rangle_S \colon H_1(S) \times H_1(S) \to \mathbb{Z}$$

as follows. For $a, b \in H_1(S)$, choose 1-cycles $\alpha, \beta \subset S$ representing them. We write β^+ for the 1-cycle in $S^3 \setminus S$ obtained by pushing β off S along the positive normal of S. Then

$$\langle a, b \rangle_S := \mathrm{lk}(\alpha, \beta^+).$$

If $a_1, \dots, a_{2g}$ is a basis of the free $\mathbb{Z}$-module $H_1(S)$, then the corresponding *Seifert matrix* V has (i,j)-th entry $\langle a_i, a_j \rangle_S$ for $i, j \in \{1, \dots, 2g\}$.

The Seifert form is bilinear, but not necessarily symmetric. Given a Seifert matrix V of K, we call $V + V^T$ the corresponding *symmetrised Seifert matrix*.

Definition 3.4.2. The *Alexander polynomial* of a knot K in S^3 is defined as

$$\Delta_K(t) := \det(V - tV^T),$$

where V is a Seifert matrix for K.

Proposition 3.4.3. *The Alexander polynomial is well-defined up to multiplication by t^k for $k \in \mathbb{Z}$.*

Proof. We first show that $\det(V - tV^T)$ is independent of the choice of basis $a_1, \dots, a_{2g}$. Let $b_1, \dots, b_{2g}$ be another basis of $H_1(S)$. Then the Seifert matrix V_1

obtained from $b_1, \dots, b_{2g}$ is of the form $M^T V M$ for an invertible integral matrix M. Then $\det(M) = \pm 1$ as it is a unit in $\mathbb{Z}$. Hence

$$\begin{aligned}\det(V_1 - tV_1^T) = &\det(M^T V M - tM^T V^T M) = \\ &\det(M^T)\det(V - tV^T)\det(M) = \\ &\det(V - tV^T).\end{aligned}$$

By Proposition 3.3.4, it suffices to prove that $\det(V - tV^T)$ is unchanged by a stabilisation of S. Suppose that S' is a stabilisation of S. If V is the Seifert matrix of S, then

$$V' = \begin{pmatrix} V & * & 0 \\ * & * & 1 \\ 0 & 0 & 0 \end{pmatrix} \text{ or } \begin{pmatrix} V & * & 0 \\ * & * & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

is a Seifert matrix of S', where we extend a basis of $H_1(S)$ by adding a longitude and a meridian of the stabilisation tube. After a change of basis, this becomes

$$\begin{pmatrix} V & * & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \text{ or } \begin{pmatrix} V & 0 & 0 \\ * & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Hence, we have $\det(V' - t(V')^T) = t\det(V - tV^T)$, and the result follows. $\square$

The Alexander polynomial satisfies the symmetry relation $\Delta_K(t) \doteq \Delta_K(t^{-1})$, where "$\doteq$" denotes equality up to multiplication by t^k for some $k \in \mathbb{Z}$. The Conway normalization of the Alexander polynomial is the unique Laurent polynomial satisfying $\Delta_K(t) = \Delta_K(t^{-1})$. It is immediate from the definition that

$$\deg \Delta_K(t) \le g(K).$$

Since $V - V^T$ agrees with the matrix of the intersection form of S in the basis $a_1, \dots, a_{2g}$, which is unimodular, we have

$$\Delta_K(1) = \det(V - V^T) = \pm 1. \tag{3.4.1}$$

We now introduce the signature of a knot, which, as we shall see, gives lower bounds on the genera of orientable surfaces the knot bounds in the 4-ball.

Definition 3.4.4. The *signature* $\sigma(K)$ of the knot K is defined as the signature of the symmetrised Seifert matrix $V + V^T$ (i.e., the number of positive eigenvalues minus the number of negative eigenvalues).

Proposition 3.4.5. *The signature $\sigma(K)$ of a knot K is well-defined.*

Proof. For a given basis of $H_1(S)$, the symmetrised Seifert matrix is the matrix of the symmetric bilinear form

$$\langle a, b\rangle := \langle a, b\rangle_S + \langle b, a\rangle_S.$$

The signature is then the dimension of a maximal positive definite subspace of $H_1(S)$ minus the dimension of a maximal negative definite subspace of $H_1(S)$. This shows that the signature is independent of the choice of basis.

Invariance under stabilisation of the Seifert surface follows from the observation that $V+V^T$, up to change of basis, changes by taking the direct sum with the matrix

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

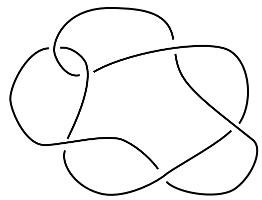

FIGURE 11. The twist knot 6_1 is the first non-trivial slice knot.

which has signature 0. □

It immediately follows from the definition that the signature is additive under connected sum. We have the following analogues of Seifert surfaces and the Seifert genus in dimension four:

DEFINITION 3.4.6. Let K be a knot in S^3. A *slice surface* for K is a connected, compact, oriented surface F embedded in D^4 such that $\partial F = K$. The *4-ball genus* $g_4(K)$ of K is the minimal genus of a slice surface for K.

Clearly, $g_4(K) \le g(K)$. A knot is called *slice* if $g_4(K) = 0$. There are nontrivial slice knots. The smallest one is the twist knot 6_1 in the Rolfsen table; see Figure 3.4.

EXERCISE 3.4.7. Show that the knot 6_1 is slice by attaching a single band, resulting in a 2-component unlink.

LEMMA 3.4.8. *Let K and K' be disjoint knots in S^3 with Seifert surfaces S and S' and slice surfaces F and F', respectively. Then*

$$lk(K, K') = S \cdot K' = S' \cdot K = F \cdot F',$$

where $\cdot$ denotes the algebraic intersection number.

PROOF. Since S is Poincaré–Lefschetz dual to the meridian of K in the exterior of K, we have $\mathrm{lk}(K, K') = S \cdot K'$. Similarly, $\mathrm{lk}(K, K') = S' \cdot K$.

View D^4 as S^4_+, and push S' into S^4_-. Then $F \cup -S$ and $F' \cup -S'$ are closed oriented surfaces in S^4 that are null-homologous, hence

$$0 = (F \cup -S) \cdot (F' \cup -S') = F \cdot F' + S \cdot S'.$$

As $S \cap S' = S \cap K'$, and since the intersection signs are opposite, the result follows from the previous paragraph. □

PROPOSITION 3.4.9. *Let K be a knot in S^3. Then*

$$\frac{|\sigma(K)|}{2} \le g_4(K).$$

PROOF. Let $g = g_4(K)$, and let F be a genus g slice surface for K. Choose an arbitrary Seifert surface S for K. Then $S \cup F$ is a closed, oriented, null-homologous surface in D^4, and hence bounds a Seifert 3-manifold M according to Proposition 3.3.3. Let

$$U = \ker(H_1(\partial M) \to H_1(M)).$$

By Lemma 2.4.2,

$$\mathrm{rk}(U) = \frac{\mathrm{rk}(H_1(\partial M))}{2} = g(S) + g(F).$$

Note that $H_1(\partial M) \cong H_1(S) \oplus H_1(F)$, and write $U_S := U \cap H_1(S)$. Then

$$\operatorname{rk}(U_S) \geq \operatorname{rk}(U) - \operatorname{rk}(H_1(F)) = g(S) - g(F). \tag{3.4.2}$$

We claim that $\langle a, b\rangle_S = 0$ for a, $b \in U_S$. Indeed, a and b are 0-homologous in M, and let A and B be 2-chains in M with $\partial A = a$ and $\partial B = b$. We write B^+ for the 2-chain in D^4 obtained by pushing B off M in the direction of the positive normal of M. Then $\partial B^+ = b^+$ and $A \cap B^+ = \emptyset$, hence

$$\langle a, b\rangle_S = \operatorname{lk}(a, b^+) = A \cdot B^+ = 0,$$

where the second equality follows from Lemma 3.4.8.

On the other hand, if V is a Seifert matrix associated to S, then

$$\det(V + V^T) \equiv \det(V - V^T) \equiv 1 \mod 2$$

by equation (3.4.1), so $V + V^T$ is nondegenerate. Hence, if $P \leq H_1(S)$ and $N \leq H_1(S)$ are maximal positive and negative definite subspaces of the Seifert form, respectively, then $P \oplus N = H_1(S)$. So, if we write $p = \operatorname{rk}(P)$ and $n = \operatorname{rk}(N)$, then $p + n = 2g(S)$.

As $P \cap U_S = \{0\}$ and $N \cap U_S = \{0\}$, we have

$$\max(p, n) \leq 2g(S) - \operatorname{rk}(U_S) \leq g(S) + g(F),$$

where the last inequality follows from equation (3.4.2). Hence

$$\sigma(K) = p - n = (p + n) - 2n \geq 2g(S) - 2(g(S) + g(F)) = -2g(F).$$

Similarly, $-\sigma(K) \geq -2g(F)$, and the result follows. □

The signature of a knot can also be computed from an unoriented spanning surface for the knot due to the work of Gordon and Litherland [**16**].

Definition 3.4.10. Let S be a compact, connected, embedded surface in S^3. The *Gordon–Litherland pairing*

$$\langle\, ,\, \rangle_S \colon H_1(S) \times H_1(S) \to \mathbb{Z}$$

is defined as follows: Let a, $b \in H_1(S)$, represented by oriented multi-curves α, $\beta \subset S$. Consider the unit normal bundle $p_S \colon UN(S) \to S$ of S in S^3. Then

$$\langle \alpha, \beta\rangle_S := \operatorname{lk}(\alpha, p_S^{-1}(\beta)).$$

The Gordon–Litherland pairing is symmetric and bilinear. It agrees with the symmetrised Seifert form when S is orientable. Let $\{b_1, \dots, b_n\}$ be a basis of $H_1(S)$. Then the *Goeritz matrix* G_S is an $n \times n$ symmetric matrix with (i, j)-th entry $\langle b_i, b_j\rangle_S$ for i, $j \in \{1, \dots, n\}$. Furthermore, the *normal Euler number* $e(S)$ of S is defined to be $-\operatorname{lk}(K, K')$, where K' is the framing of K given by S. Gordon and Litherland proved the following:

Theorem 3.4.11. *Let S be an unoriented surface bounding the knot K in S^3. Then*

$$\sigma(K) = \sigma(G_S) + \frac{e(S)}{2},$$

where $\sigma(G_S)$ is the signature of the Goeritz matrix.

3.5. Important classes of knots

In this section, we present several important classes of knots.

3.5.1. Alternating knots.

DEFINITION 3.5.1. The knot K is *alternating* if it admits a diagram $\mathcal{D}$ such that, as we travel along $\mathcal{D}$, we alternatingly encounter undercrossings and overcrossings.

Alternating knots have a number of nice properties. A large proportion of small crossing knots are alternating, but their proportion tends to zero as the crossing number goes to infinity.

DEFINITION 3.5.2. We say that a crossing c of a knot diagram $\mathcal{D}$ is *reducible* or *nugatory* if there is a circle $C \subset \mathbb{R}^2$ such that $C \cap \mathcal{D} = \{c\}$. The diagram $\mathcal{D}$ is called *reduced* if it does not have a reducible crossing.

We can always remove such a crossing by applying $\pm\pi$-rotation about a line in $\mathbb{R}^2$ through the crossing to the part of $\mathcal{D}$ in the circle C. Consequently, if a diagram realises the crossing number of a knot, it has to be reduced. A related operation on knot diagrams is called a flype:

DEFINITION 3.5.3. Let $\mathcal{D}$ be a knot diagram, and suppose that $C \subset \mathbb{R}^2$ is a circle that intersects $\mathcal{D}$ in a crossing c and two additional points that are not crossings. A *flype* on $\mathcal{D}$ is the operation of rotating the portion of $\mathcal{D}$ inside C by angle $\pm\pi$ about a line in $\mathbb{R}^2$ passing through c such that we remove the crossing c, but create a new crossing between the other two strands intersecting C.

The following results were conjectured by Tait. The first two statements were shown by Kauffman [**31**], Murasugi [**53**], and Thistlethwaite [**67**] using the Jones polynomial, and the third by Menasco and Thistlethwaite [**40**].

THEOREM 3.5.4. *Let $\mathcal{D}$ and $\mathcal{D}'$ be reduced alternating diagrams of a knot K. Then the following hold:*

(1) $c(\mathcal{D}) = c(\mathcal{D}')$,
(2) $w(\mathcal{D}) = w(\mathcal{D}')$, and,
(3) if K is prime, we can obtain $\mathcal{D}'$ from $\mathcal{D}$ using flypes.

In fact, any reduced alternating diagram of a knot has minimal crossing number. Conversely, if K is a prime alternating knot, then any minimal crossing number diagram of K is alternating. (This does not hold for nonprime knots; see for example the square knot $3_1\# - 3_1$, which admits a nonalternating six-crossing diagram.) The reduced alternating diagram of a prime alternating knot is unique up to flypes according to statement (3), known as the Tait flyping conjecture. Since a flype preserves both the crossing number and the writhe, statement (3) implies statements (1) and (2).

THEOREM 3.5.5. *If we apply Seifert's algorithm to a reduced alternating diagram of a knot, we obtain a minimal genus Seifert surface.*

3.5.2. Torus knots.

DEFINITION 3.5.6. Let p and q be coprime integers. The (p, q)-torus knot $T_{p,q}$ is given by the curve on the standard torus T^2 in $\mathbb{R}^3$ that winds around the longitude p times and the meridian q times.

In other words, $T_{p,q}$ is the projection of the line $\langle(p, q)\rangle \subset \mathbb{R}^2$ under the covering map $\mathbb{R}^2 \to \mathbb{R}^2/\mathbb{Z}^2 \cong T^2$ that sends $(1, 0)$ to the longitude and $(0, 1)$ to the meridian. When p and q are not coprime and n is their highest common factor, we obtain

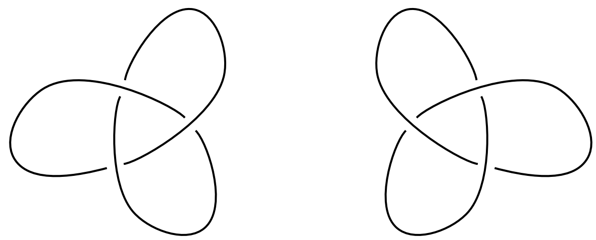

FIGURE 12. The left shows the right-handed trefoil $T_{2,3}$, and the right its mirror, the left-handed trefoil $T_{2,-3}$.

an n-component link, where each component is $T_{(p/n,q/n)}$. For example, $T_{2,2}$ is the Hopf link.

When $p = 1$ or $q = 1$, the torus knot $T_{p,q}$ is the unknot. The first nontrivial torus knot is $T_{2,3}$, which is the trefoil knot, denoted 3_1 in the Rolfsen table; see Figure 12.

Torus knots are typically far from being alternating. They are particularly important as they arise from singularities of complex plane curves:

EXERCISE 3.5.7. The complex polynomial $f(w, z) = w^p + z^q$ has a critical point at the origin of $\mathbb{C}^2$. Let the plane curve $V_f \subset \mathbb{C}^2$ be the zero-set of f. Show that, for $\varepsilon > 0$ sufficiently small, the link of the singular point of V_f at the origin, defined as $V_f \cap S^3_\varepsilon$, is the torus knot $T_{p,q}$.

The torus knots $T_{p,q}$ and $T_{q,p}$ are equivalent. Indeed, there is an orientation-preserving automorphism of S^3 that swaps the solid tori on the two sides of T^2, and interchanges the longitude and meridian. The torus knots $T_{p,q}$ and $T_{p,-q}$ are mirror images.

PROPOSITION 3.5.8. *Let p, $q > 0$. Then*

$$c(T_{p,q}) = \min\{(p-1)q, (q-1)p\} \text{ and } g(T_{p,q}) = g_4(T_{p,q}) = \frac{(p-1)(q-1)}{2}.$$

EXERCISE 3.5.9. Compute the signature of the right-handed trefoil using Seifert matrices, and also using the formula of Gordon and Litherland from the checkerboard surfaces.

3.5.3. Satellite knots.

DEFINITION 3.5.10. Let K' be a knot in the solid torus $S^1 \times D^2$ that does not lie in a 3-ball, and is not isotopic to $S^1 \times \{0\}$. Furthermore, let K be a knot in S^3, called the *companion*, and $n \in \mathbb{Z}$ a framing coefficient. Choose a diffeomorphism $f\colon S^1 \times D^2 \to N(K)$ representing the framing n. Then the *n-twisted satellite of K with pattern K'* is $f(K')$, which we denote by $K(K', n)$. We say that J is a *satellite knot* if $J = K(K', n)$ for $K \neq U$.

EXAMPLE 3.5.11. If there is a point $p \in S^1$ such that $|(\{p\} \times D^2) \cap K'| = 1$, then $K(K', n) = K \# K'$ for any n, where we view K' as a knot in S^3 by identifying $S^1 \times D^2$ with the standard solid torus in S^3.

EXAMPLE 3.5.12. If $K' = T_{p,q}$, then $K(K', 0)$ is called the (p, q)-*cable* of K. The $(n, 0)$-cable of K is also known as the n-cable of K (note that $T_{n,0}$ is an n-component unlink).

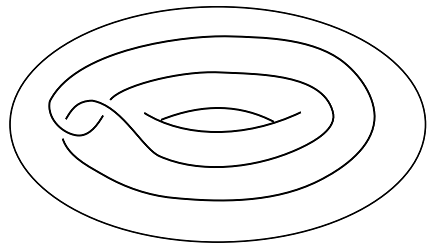

FIGURE 13. The Whitehead knot in the solid torus.

EXAMPLE 3.5.13. If W is the Whitehead knot in $S^1 \times D^2$ shown in Figure 13, then $K(W, n)$ is called the *n-twisted Whitehead double* of K.

The *algebraic winding number* of an oriented knot in $S^1 \times D^2$ is its signed intersection number with $\{1\} \times D^2$ (or, equivalently, its homology class in $H_1(S^1 \times D^2) \cong \mathbb{Z}$), while its *geometric winding number* is the minimal number of intersection points with $\{1\} \times D^2$ counted without signs. The algebraic winding number of W is zero, while its geometric winding number is two.

EXERCISE 3.5.14. If K is a Whitehead double, it has a genus one Seifert surface. Use this to show that $\Delta_K(t) = 1$.

Satellite knots are characterised by the property that their exterior E contains a torus T not parallel to ∂E that is incompressible; i.e., there is no homotopically nontrivial simple closed curve on ∂E that bounds an embedded disk in E.

3.5.4. Hyperbolic links.

DEFINITION 3.5.15. A link L in S^3 is *hyperbolic* if $S^3 \setminus L$ admits a complete hyperbolic metric; i.e., a Riemannian metric of constant sectional curvature -1.

The following result is a special case of Thurston's hyperbolisation theorem:

THEOREM 3.5.16. *Every knot in the 3-sphere is either hyperbolic, a torus knot, or a satellite, and these classes are mutually exclusive.*

In other words, we can build all knots from hyperbolic knots and torus knots using satellite operations. By the Mostow–Prasad rigidity theorem [**50**][**56**], which we now state, the hyperbolic structure is unique, and so any geometric quantity one can assign to the hyperbolic structure on the knot complement is a knot invariant:

THEOREM 3.5.17. *Let M and N be complete finite volume hyperbolic manifolds of dimensions at least* 3. *Then any isomorphism $\pi_1(M) \to \pi_1(N)$ is induced by a unique isometry from M to N.*

The most fundamental hyperbolic knot invariant is the volume. Other invariants describe the shape of the cusp, such as the cusp volume, the lengths of the longitude and meridian, and the longitudinal and meridional translations.

The hyperbolic structure can often be found using the computer software SnapPy [**9**], which can also compute many of the associated invariants.

Every complete, connected hyperbolic n-manifold has universal cover $\mathbb{H}^n$, hence can be written as $\mathbb{H}^n/\Gamma$, where Γ is a torsion-free discrete group of isometries of $\mathbb{H}^n$. When $n = 3$, this isometry group is $\mathrm{PSL}(2, \mathbb{C})$. Hence hyperbolic 3-manifolds can also be studied using group theory.

3.6. The knot group

Given a knot K in S^3, its *complement* is $S^3 \setminus K$, and its *exterior* is $S^3 \setminus N(K)$. The complement and the exterior uniquely determine each other. Knot complements are considered more often when studying hyperbolic structures, while the exterior has the advantage of being compact.

By a deep result of Gordon and Luecke [**17**], knots in S^3 are determined by their exteriors up to equivalence:

THEOREM 3.6.1. *The isotopy class of a knot in S^3 is determined by the orientation-preserving homeomorphism type of its exterior.*

While the longitude of the knot is homologically determined by the half lives, half dies lemma (Lemma 2.4.2), there are infinitely many possible choices for the meridian. What Gordon and Luecke actually showed is that there is exactly one gluing of a solid torus to the knot exterior that gives S^3. The fact that the knots are in S^3 is important, since there are different knots in lens spaces with homeomorphic complements; see Bleiler–Hodgson–Weeks [**3**].

The fundamental group of the knot complement is a powerful knot invariant. A presentation, called the *Wirtinger presentation*, can be computed as follows:

THEOREM 3.6.2. *Let $\mathcal{D}$ be a diagram of the knot K in S^3. Label the components of $\mathcal{D}$ by $a_1, \dots, a_c$ as we follow the orientation of K, where c is the number of crossings of $\mathcal{D}$, and we call crossing i the endpoint of the arc a_i. If a_j is the over-strand at crossing i, let*

$$w_i = \begin{cases} a_i a_j^{-1} a_{i+1}^{-1} a_j & \text{if crossing } i \text{ is positive,} \\ a_i a_j a_{i+1}^{-1} a_j^{-1} & \text{otherwise.} \end{cases}$$

Then

$$\langle\, a_1, \dots, a_c \,|\, w_1, \dots, w_c \,\rangle$$

is a presentation of $\pi_1(S^3 \setminus K)$. In fact, $S^3 \setminus K$ is homotopy equivalent to a cell complex with a single 0-cell, 1-cells $a_1, \dots, a_c$, 2-cells attached along $w_1, \dots, w_c$, and a single 3-cell.

PROOF. It clearly suffices to construct the claimed cell decomposition. Let D^3 be the upper hemisphere of S^3, which we view as a zero-handle, with $\mathcal{D}$ lying on its boundary.

Let $N(a_i) \approx I \times [-1, 1]$ be a regular neighbourhood of a_i in S^2, where $a_i \approx I \times \{0\}$ and $a_i \times \{-1\}$ lies to the right of a_i. We attach a two-dimensional one-handle to D^3 along $I \times \{-1, 1\}$, which we view as a tunnel over a_i; see Figure 14. We orient its core from $I \times \{-1\}$ to $I \times \{1\}$. We repeat this for each arc a_i, after which we have a space homotopy equivalent to a wedge of c circles.

Over crossing i, we then attach a two-cell of the form $[-1, 1] \times [-1, 1]$. Its boundary goes over a_i in the positive direction, passes over the tunnel over a_j in the positive or negative direction depending on the sign of crossing i, goes over a_{i+1} in the negative direction, and finally goes over a_j in the opposite direction. Hence, the attaching map of the two-cell is precisely w_i.

Finally, we attach the three-dimensional three-handle corresponding to the lower hemisphere of S^3, and obtain a CW decomposition of $S^3 \setminus K$ of the required form. We remark that we can thicken the one-cells and two-cells to obtain a handle decomposition of the knot exterior. $\square$

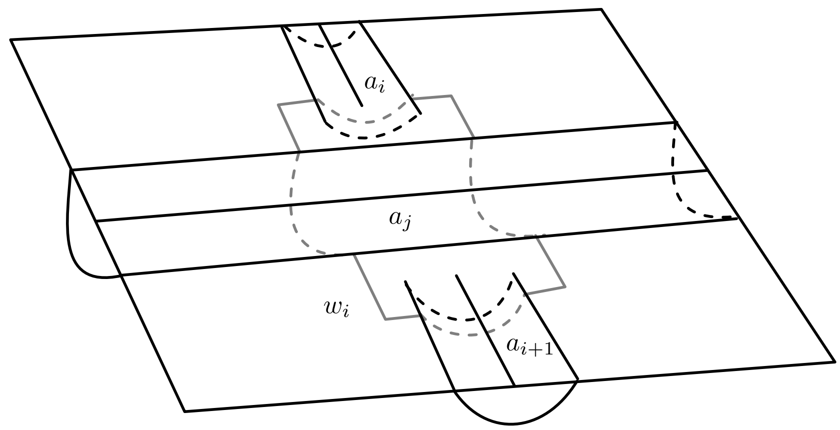

FIGURE 14. Illustration of the proof of the Wirtinger presentation. The zero-handle is the upper half-space, and the tunnels are one-handles attached below the plane. The two-cell over crossing i is attached from underneath along the curve w_i shown in grey.

EXERCISE 3.6.3. Show using the Wirtinger presentation that $H_1(S^3 \setminus K) \cong \mathbb{Z}$. Use Alexander duality to compute $H_1(S^3 \setminus L)$ for a link L in S^3.

The knot group is not a complete knot invariant. For example, the square knot $3_1\# - 3_1$ and the granny knot $3_1\#3_1$ have isomorphic groups, where 3_1 is the right-handed and -3_1 is the left-handed trefoil. However, the knot group together with the subgroup $\pi_1(\partial M)$, called a *peripheral system*, is a complete knot invariant:

THEOREM 3.6.4. *Let K and K' be knots in S^3 with exteriors M and M', respectively. If there is an isomorphism $\pi_1(M) \to \pi_1(M')$ that induces an isomorphism $\pi_1(\partial M) \to \pi_1(\partial M')$, then K and K' are equivalent. If K and K' are prime, then they are equivalent if and only if they have isomorphic groups.*

PROOF. As M and M' are 3-manifolds with boundary other than D^3, they are Haken, and we can apply Waldhausen's theorem (Theorem 2.4.4) to see that M and M' are homeomorphic. To conclude that K and K' are equivalent knots, we now invoke the theorem of Gordon and Luecke (Theorem 3.6.1). □

We can compute the Alexander polynomial of a knot from a presentation of its fundamental group via Fox calculus: Given a free group F generated by $g_1, \dots, g_n$, the *Fox derivatives* $\partial/\partial g_i \colon \mathbb{Z}[F] \to \mathbb{Z}[F]$ is a group homomorphism characterised by the following axioms:

(1) $\partial g_i/\partial g_j = \delta_{ij}$,
(2) $\partial e/\partial g_i = 0$, and
(3) $\partial(uv)/\partial g_i = \partial u/\partial g_i + u(\partial v/\partial g_i)$ for any u, $v \in F$.

EXERCISE 3.6.5. Deduce from the axioms that

$$\partial u^{-1}/\partial g_i = -u^{-1}(\partial u/\partial g_i)$$

for any $u \in F$.

Now suppose that $\langle g_1, \dots, g_n \,|\, r_1, \dots, r_m \rangle$ is a presentation of $G := \pi_1(S^3 \setminus K)$. Let F be the free group on $g_1, \dots, g_n$, and

$$H := G/[G,G] = H_1(S^3 \setminus K) \cong \mathbb{Z}.$$

Consider the homomorphism $\phi\colon \mathbb{Z}[F] \to \mathbb{Z}[H]$, and write t for the generator of H. We form the matrix J whose (i,j)-th entry is $\phi(\partial r_i/\partial g_j)$ for $i \in \{1,\dots,m\}$ and $j \in \{1,\dots,n\}$. Then the greatest common divisor of all $(n-1)\times(n-1)$ minors of J agrees with the Alexander polynomial of K, up to multiplication by $\pm t^k$ for $k \in \mathbb{Z}$. (Recall that the Alexander polynomial is only well-defined up to multiplication by t^k for $k \in \mathbb{Z}$.)

Exercise 3.6.6. Using Fox calculus, compute the Alexander polynomial of the trefoil knot.

3.7. Fibred knots

Definition 3.7.1. A knot K in a 3-manifold Y is said to be *fibred* if its complement $Y \setminus K$ is a fibre bundle over S^1 such that the closure of each fibre is a Seifert surface of K.

Example 3.7.2. The simplest example of a fibred knot is the unknot U, since $S^3 \setminus U \approx S^1 \times B^2$. The first non-trivial example is given by the trefoil knot. The figure eight knot is also fibred; we will shortly introduce a method that will allow us to easily prove this.

Example 3.7.3. The torus knot $T_{p,q}$ is fibred for any pair of relatively prime integers p and q. Indeed, if we use the description of $T_{p,q}$ as the link of the singularity $w^p + z^q$ at the origin in $\mathbb{C}^2$ given in Exercise 3.5.7, then

$$\frac{w^p + z^q}{|w^p + z^q|}\colon S^3 \setminus K \to S^1$$

is a fibration. Indeed, the restriction of this map to the knot exterior is a proper submersion, which is hence a fibre bundle by Ehresmann's fibration lemma (Exercise 1.4.6).

More generally, suppose that $f\colon \mathbb{C}^2 \to \mathbb{C}$ is a polynomial such that $f(\underline{0}) = 0$, $\frac{df}{\partial z}(\underline{0}) = 0$, $\frac{\partial f}{\partial w}(\underline{0}) = 0$, and $\underline{0} \in \mathbb{C}^2$ is an isolated common zero of these functions. Then $V := f^{-1}(0)$ is an algebraic variety with an isolated singularity at $\underline{0}$. The *link of the singularity* is the knot $K := V \cap S^3_\varepsilon$ for some $\varepsilon > 0$ small. Then K is a fibred knot with fibration $g/|g|\colon S^3 \setminus K \to S^1$, where $g := f|_{S^3\setminus K}$. This is called the *Milnor fibration* of the singularity. For more detail, see the book of Milnor [**41**].

The following result of Stallings [**4**] gives a complete characterisation of fibred knots:

Theorem 3.7.4. *Let K be a knot in S^3. Then K is fibred if and only if the commutator subgroup of $\pi_1(S^3 \setminus K)$ is finitely generated.*

Proposition 3.7.5. *Let K be a fibred knot in S^3. Then the Alexander polynomial $\Delta_K(t)$ is monic.*

Proof. Let S be a fibre surface, and fix a basis $\mathcal{B} = \{b_1, \dots, b_{2g}\}$ of $H_1(S)$. Then we can write the exterior of K as a mapping torus S_φ, where φ is called the *monodromy* of the fibration. Let V be the Seifert matrix and M the matrix of $\varphi_*\colon H_1(S) \to H_1(S)$ with respect to $\mathcal{B}$. Note that

$$\mathrm{lk}(b_i, b_j^+) = \mathrm{lk}(b_i^+, \varphi_* b_j) = \mathrm{lk}(\varphi_* b_j, b_i^+),$$

by translating b_i and b_j^+ along the mapping cylinder $(S \times I)/_{(x,1)\sim(\varphi(x),0)}$ until b_j^+ reaches $S \times \{1\}$. This implies that $V = M^T V^T$ and V is invertible, so $M = V^{-1}V^T$. Hence, the Alexander polynomial

$$\Delta_K(t) = \det(V - tV^T) = \det(V)\det(I - tV^{-1}V^T) = \det(I - tM).$$

As φ is an automorphism of S, the matrix M is invertible over $\mathbb{Z}$, so the leading coefficient of $\Delta_K(t)$ is $\det(M) = \pm 1$. $\square$

If K is fibred, the surface obtained by taking the closure of a fibre of the fibration $S^3 \setminus K \to S^1$ is a minimal genus Seifert surface of K, and is unique up to isotopy relative to K.

Gabai introduced sutured manifolds to give a simple method to check whether a knot is fibred.

Definition 3.7.6. A sutured manifold is a pair (M, γ), where M is a compact, oriented 3-manifold with boundary, and γ is a collection of thickened, oriented simple closed curves in ∂M that divide ∂M into two subsurfaces $R_+(\gamma)$ and $R_-(\gamma)$ that meet along γ. Furthermore, $R_+(\gamma)$ is oriented as ∂M, while $R_-(\gamma)$ is oriented as $-\partial M$, and γ is oriented as the boundary of both $R_+(\gamma)$ and $R_-(\gamma)$.

Given a knot K in an oriented 3-manifold Y and a Seifert surface S for K, the *sutured manifold complementary to* S is given by the pair (M, γ), where $M = Y \setminus (S \times [-1, 1])$ where $S \times [-1, 1]$ is a product neighbourhood of S in Y such that $S \times \{1\}$ lies on the positive side of S. Furthermore, $\gamma = \partial S \times [-1, 1]$ and $R_\pm(\gamma) = S \times \{\pm 1\}$.

We say that (M, γ) is a *product sutured manifold* if $M = R \times [-1, 1]$, $\gamma = \partial R \times [-1, 1]$, and $R_\pm(\gamma) = R \times \{\pm 1\}$ for some compact, oriented surface R.

Proposition 3.7.7. *The knot K in S^3 is fibred with fibre S if and only if the sutured manifold complementary to S is a product.*

Proof. This is a straightforward corollary of the definitions. $\square$

Given a sutured manifold (M, γ), a *product disk* in (M, γ) is a properly embedded disk $D \subset M$ such that $|D \cap \gamma| = 2$. We can decompose (M, γ) along D by taking $M' := M \setminus N(D)$, and γ' is obtained by reconnecting the ends of $\gamma \setminus N(D)$ along $D \times \{-1\}$ and $D \times \{1\}$, where we identify $N(D)$ with $D \times [-1, 1]$. Then $(M, \gamma) \overset{D}{\rightsquigarrow} (M', \gamma')$ is called a *product decomposition*. The following result is straightforward:

Proposition 3.7.8. *The connected sutured manifold (M, γ) is a product if and only if it admits a sequence of product decompositions terminating at a 3-ball with a single suture on it.*

This gives us a practical method for showing that a knot K in S^3 is fibred: Choose a Seifert surface S for K, usually using Seifert's algorithm. Then consider the product sutured manifold $(S \times [-1, 1], \partial S \times [-1, 1])$ by thickening S, and considering K as the suture. We then look for product disks in the *complement* of $(S \times [-1, 1], \partial S \times [-1, 1])$. Decomposing the complement along such a disk D

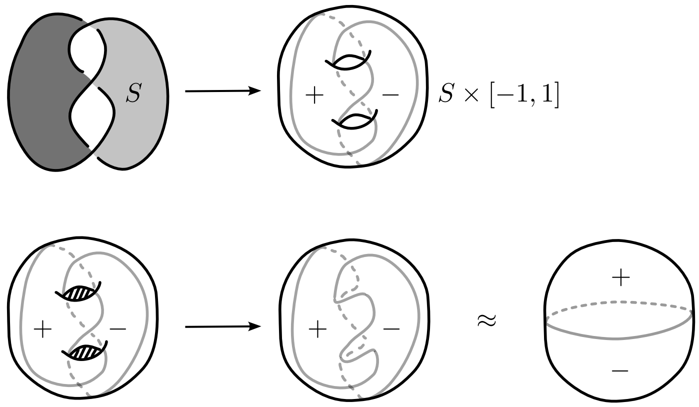

FIGURE 15. A proof that the trefoil knot is fibred using sutured manifolds.

amounts to adding $D \times [-1,1]$ to $(S \times [-1,1], \partial S \times [-1,1])$, and reconnecting the sutures along the boundary. If, at the end of this process, we end up with D^3 with a single suture, then K is fibred, and S is a minimal genus Seifert surface that is a fibre.

EXAMPLE 3.7.9. Figure 15 illustrates the above procedure in the case of the right-handed trefoil knot. Consider the Seifert surface S shown in the upper left. Its positive side is on the left. Thickening S results in the genus two handlebody $H := S \times [-1,1]$ on the right. The original knot is our suture γ on ∂H. This divides ∂H into $R_+(\gamma)$ and $R_-(\gamma)$. By construction, this is a product sutured manifold.

We now focus on the complementary sutured manifol (M,γ), where $M := S^3 \setminus \mathrm{Int}(H)$. We have shaded two product disks in (M,γ) on the bottom left. If we decompose (M,γ) along these, then we obtain the sutured manifold in the middle of the bottom row, which is diffeomorphic to D^3 with a single suture.

EXERCISE 3.7.10. Use product decompositions to show that the figure eight knot is fibred.

3.8. The Jones polynomial

Vaughan Jones [**28**] introduced a novel invariant of links in 1984 using von Neumann algebras, which was the first new knot polynomial after the Alexander polynomial. It admits a much simpler definition due to Kauffman [**31**]. Unlike the Alexander polynomial, its relationship with the geometry of the link is much less clear. It is conjectured to detect the unknot. It was the main tool for proving Tait's conjectures (Theorem 3.5.4). It admits a categorification called Khovanov homology – a homology theory whose graded Euler characteristic is the Jones polynomial – that has recently been shown to detect the unknot by Kronheimer and Mrowka [**36**].

Let L be a link in S^3. The Jones polynomial $V_L(t)$ is a Laurent polynomial in $\mathbb{Z}[t^{1/2}, t^{-1/2}]$ characterised by $V_U(t) = 1$ and the oriented skein relation

$$(t^{1/2} - t^{-1/2})V_{L_0}(t) = t^{-1}V_{L_+}(t) - tV_{L_-}(t),$$

where L_+, L_-, and L_0 are link diagrams that agree outside a single crossing, L_- is obtained from L_+ by changing a positive crossing to a negative one, and L_0 is the

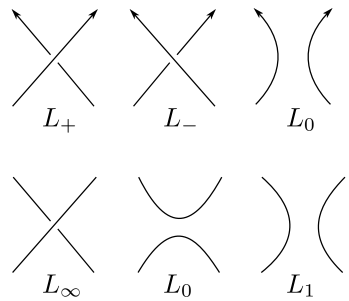

FIGURE 16. The top row shows the three links featuring in the oriented skein relation, and the bottom row the links in the unoriented skein relation.

oriented resolution of the crossing; see the top row of Figure 16. When the number of component of L is odd, then $V_L(t) \in \mathbb{Z}[t, t^{-1}]$.

It is not clear from this description that there exists such a polynomial, and that it is unique. We now give the definition of the Jones polynomial using the Kauffman bracket, which satisfies a different type of skein relation.

DEFINITION 3.8.1. Let L be an unoriented link diagram in S^2. Then its *Kauffman bracket* $\langle L \rangle \in \mathbb{Z}[A, A^{-1}]$ is characterised by the following relations:

(1) $\langle \bigcirc \rangle = 1$,
(2) $\langle L \sqcup \bigcirc \rangle = -(A^2 + A^{-2})\langle L \rangle$, and
(3) if the diagrams L_∞, L_0, and L_1 are as in the bottom row of Figure 16, then $\langle L_\infty \rangle = A^{-1}\langle L_0 \rangle + A\langle L_1 \rangle$ (unoriented skein relation).

If a diagram L has n crossings, there are two ways of resolving each, giving rise to 2^n unlink diagrams. The Kauffman bracket $\langle L \rangle$ is then recursively obtained as a combination of the bracket polynomials of these unlinks. A c-component unlink has bracket $(-A^2 - A^{-2})^{c-1}$. To see that this is independent of the order of resolutions, one checks that the result is unchanged if we swap the order of two adjacent crossings. The Kauffman bracket is not quite a link invariant:

LEMMA 3.8.2. *Let L be a link diagram in S^2.*

(1) Suppose that L' is obtained from L via an R1 move. Then either the 0-resolution or the 1-resolution of L' at the new crossing splits off an unknot component. In the former case, $\langle L' \rangle = -A^{-3}\langle L \rangle$, and, in the latter, $\langle L' \rangle = -A^3\langle L \rangle$.
(2) If L' is obtained from L by an R2 or R3 move, then $\langle L \rangle = \langle L' \rangle$.

PROOF. Consider part (1), and assume that the 1-resolution splits off an unknot. If we set $L_\infty := L'$ and apply the unoriented skein relation to it, then $L_0 = L$ and $L_1 = L \sqcup \bigcirc$, so we obtain that

$$\langle L_\infty \rangle = A^{-1}\langle L \rangle + A\langle L \sqcup \bigcirc \rangle = A^{-1}\langle L \rangle - A(A^2 + A^{-2})\langle L \rangle = -A^3\langle L \rangle,$$

as claimed. The cased when the 0-resolution splits off an unknot is analogous.

Invariance under Reidemeister move R2 is obtained by applying the skein relation to one of the two new crossings, followed by part (1).

Finally, R3-invariance can be seen by applying the skein relation to the topmost crossing before and after the R3 move, and applying R2-invariance twice. We leave the details as an exercise. □

It is immediate that the following is well-defined:

DEFINITION 3.8.3. Let L be a link in S^3 and choose a diagram in S^2. Then the Jones polynomial $V_L(t)$ is obtained from $(-A)^{-3w(L)}\langle L\rangle$ by substituting $t^{1/2} = A^{-2}$.

Unlike the Alexander polynomial, the Jones polynomial is hard to compute, and current algorithms are exponential time in the crossing number n of the diagram (corresponding to the 2^n resolutions).

It is a simple consequence of the above characterisation of the Jones polynomial that $V_{-K}(t) = V_K(t^{-1})$. Hence, if $V_K(t) \neq V_K(t^{-1})$, then the knot is *chiral*; i.e., $K \neq -K$. For example, $V_{3_1}(t) = t + t^3 - t^4$ is not a symmetric polynomial, hence the left- and right-handed trefoil are inequivalent knots. As for the Alexander polynomial, the Jones polynomial is also alternating for alternating links, as shown by Thistlethwaite.

It is clear from the definition that the span of the Jones polynomial gives a lower bound on the crossing number of the link diagram (and hence on the crossing number of the link). For alternating, reducible link diagram, Lickorish and Millett [**37**] showed that the span is equal to the crossing number of the diagram. This implies the first Tait conjecture (Theorem 3.5.4).

3.9. Constructing 3-manifolds using links

In this section, we introduce two methods for constructing 3-manifolds from links in the 3-sphere: Dehn surgery and cyclic branched covers.

We have already encountered surgery in an arbitrary n-manifold M along a framed, embedded k-sphere S. This described the outgoing boundary of the handle attachment to $I \times M$ along $\{1\} \times S$. Concretely, one removes a regular neighbourhood $N(S) \approx S^k \times D^{n-k}$ of S, and glues in $D^{k+1} \times S^{n-k-1}$ along $\partial N(S) = S^k \times S^{n-k-1}$ using the framing.

When we have a framed knot K in a 3-manifold Y, we are in the special situation that $\partial N(K) = S^1 \times S^1$ has a large automorphism group, namely $\mathrm{SL}(2, \mathbb{Z})$, and we are not restricted to gluing via a framing. This leads to the notion of *Dehn surgery*. As in the case of lens spaces, any automorphism of $S^1 \times S^1$ that preserves $S^1 \times \{x\}$ for some $x \in S^1$ extends to $D^2 \times S^1$. Hence, the result of gluing $D^2 \times S^1$ to $Y \setminus N(K)$ is determined, up to diffeomorphism, by the homology class of the image of $S^1 \times \{x\}$ in $\partial N(K)$. Given a slope $\alpha \in H_1(\partial N(K))$, we write $Y_\alpha(K)$ for the result of Dehn surgery along K with slope α.

If Y is a homology 3-sphere, then K is null-homologous in Y, and we have a well-defined longitude $l \in H_1(\partial N(K))$, which is the class null-homologous in $Y \setminus N(K)$. This exists by the half lives half dies lemma (Lemma 2.4.2). If we write $m \in H_1(\partial N(K))$ for the class of the meridian of K, the result of the surgery is determined by an extended rational number $p/q \in \mathbb{Q} \cup \{\infty\}$, where p and q are coprime, and the image of $S^1 \times \{x\}$ is $pm + ql \in H_1(\partial N(K))$. Gluing along a framing of K corresponds to $q = 1$; i.e., to integer surgery. When $p/q \notin \mathbb{Z}$, the surgery does not correspond to a handle attachment.

More generally, one can consider surgery along a framed link whose components are labelled by extended rational numbers. When the link is in a homology 3-sphere, there is a canonical framing. We have the following result of Lickorish [**38**] and Wallace [**73**]:

THEOREM 3.9.1. *Every closed, connected, and oriented 3-manifold Y can be obtained by Dehn surgery along a link in S^3. Furthermore, the surgery coefficients can be chosen to be integers.*

PROOF. We have seen that the oriented cobordism group Ω_3^{SO} of 3-manifolds vanishes. Hence Y bounds an oriented 4-manifold X. Choose a handle decomposition of X. By cancelling 0- and 1-handles, we can assume that X has a single 0-handle. Similarly, we can assume that X has no 4-handles. Given a 1-handle, we can cap off its core in the 0-handle to obtain an embedded S^1. If we perform surgery along it, we replace $S^1 \times D^3$ with $D^2 \times S^2$, which corresponds to attaching a 2-handle. (In the language of Kirby calculus, this corresponds to replacing a dotted circle representing a 1-handle with a 0-framed 2-handle.) If we turn the handle decomposition upside down, we can similarly replace 3-handles with 2-handles. Hence, X admits a handle decomposition with a single 0-handle and some 2-handles. This means that M can be obtained by performing integer surgery along a link in S^3. $\square$

Given a 3-manifold M with a torus boundary component T, we can glue $S^1 \times D^2$ to M along T. This operation is called *Dehn filling.* So one can think of Dehn surgery as removing a neighbourhood of a knot, and Dehn filling the resulting manifold with torus boundary.

Branched covers along links provide another useful operation for constructing and studying 3-manifolds.

DEFINITION 3.9.2. Let L be a k-component link in S^3. Then the *n-fold cyclic branched cover* $\Sigma_n(S^3, L)$ of S^3 along L is obtained by taking the n-fold cover of $S^3 \setminus N(L)$ corresponding to the kernel of the homomorphism

$$\pi_1(S^3 \setminus N(L)) \to H_1(S^3 \setminus N(L)) \cong \mathbb{Z}^k \to \mathbb{Z}_n,$$

where the first map is abelianisation, and the second map is

$$(a_1, \dots, a_k) \mapsto a_1 + \dots + a_k \mod n.$$

We then fill the boundary with k solid tori, mapping their meridians to the lifts of n times the meridians of each link component.

By a classical result of Alexander, every 3-manifold can be obtained as a branched cover of the 3-sphere.

DEFINITION 3.9.3. Let p, q, $r \geq 2$ be integers. Then we define the *Brieskorn manifold*

$$M(p, q, r) := \{\, (z_1, z_2, z_3) \in \mathbb{C}^3 : z_1^p + z_2^q + z_3^r = 0 \,\} \cap S^5.$$

EXERCISE 3.9.4. Show that $M(p, q, r)$ is a manifold of dimension 3.

The Brieskorn manifold $M(p, q, r)$ is a *homology sphere*; i.e.,

$$H_*(M(p, q, r)) \cong H_*(S^3),$$

if and only if p, q, r are pairwise relatively prime. The manifold $M(2, 3, 5)$ is known as the *Poincaré* homology sphere, sometimes also denoted by Σ_P. Poincaré originally conjectured that every homology sphere is homeomorphic to S^3, but found this counterexample and reformulated his conjecture to homotopy spheres. It can also be described by gluing the opposite faces of a dodecahedron with a twist.

The manifolds $M(p, q, r)$ can also be described as branched coverings, due to work of Milnor [**45**]:

THEOREM 3.9.5. *The Brieksorn manifold $M(p, q, r)$ is homeomorphic to the r-fold cyclic branched cover of S^3, branched along the torus link $T_{p,q}$.*

CHAPTER 4

4-manifolds

4-manifold theory has a completely different flavour than 3-manifold topology. This is the first dimension where the topological and the smooth classification differ. Furthermore, since the Whitney trick, and hence the h-cobordism theorem and surgery theory fail in dimension 4, there are numerous exotic phenomena not present in higher dimensions.

By Theorem 1.1.13, every finitely presented group appears as the fundamental group of a closed, smooth 4-manifold. Furthermore, there is no algorithm to decide whether a finitely presented group is trivial. Hence, unless one restricts the fundamental group, there is no hope of obtaining a classification. Much of 4-manifold topology focuses on simply-connected manifolds.

By the work of Freedman, the classification of simply-connected, topological 4-manifolds is governed by algebraic topology; namely, the intersection form on the second cohomology group.

The classification of smooth 4-manifolds relies on invariants that can distinguish homeomorphic but non-diffeomorphic manifolds. One of the main tools for constructing and manipulating 4-manifolds is Kirby calculus, which allows us to visualise a smooth 4-manifold as a framed link in the 3-sphere. Furthermore, it gives a set of moves such that two 4-manifolds are diffeomorphic if and only if they are related by a sequence of such moves. There are further constructions originating from algebraic and symplectic geometry, such as blow-ups and fibre sums.

The two main invariants are the Donaldson and the Seiberg–Witten invariants. These come from gauge theory. Roughly, one chooses a Riemannian metric on the manifold, writes down some non-linear PDEs, and counts the solutions. Then one shows this count is independent of the chosen metric. Heegaard Floer theory is a more combinatorial version of Seiberg–Witten theory.

4.1. Kirby calculus

Kirby calculus gives us a way to represent and manipulate 4-manifolds using framed links in the 3-sphere. If the 4-manifold has boundary, it can also be used to study the boundary 3-manifold. More generally, one can even use it to study 4-dimensional cobordisms.

Let X be a smooth, connected 4-manifold (not necessarily closed). Then it admits a handle decompositon with a single 0-handle. The 0-handle can be identified with D^4, which has boundary S^3. Each 1-handle is attached along a framed 0-sphere, which one can represent by a pair of 2-spheres, with their boundaries identified by an orientation-reversing diffeomorphism – usually a reflection.

Alternatively, one can visualise a 1-handle h^1 as an unknotted circle with a dot on it: If we attach a 2-handle h^2 cancelling h^1, we can identify the resulting manifold with D^4. The belt circle of h^2 is an unknotted circle in S^3, which we

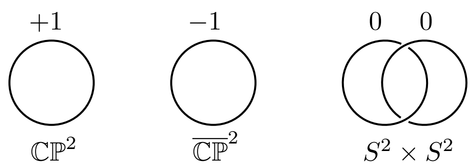

FIGURE 17. Kirby diagrams for $\mathbb{CP}^2$, $\overline{\mathbb{CP}}^2$, and $S^2 \times S^2$.

mark with a dot. If D is a 2-disk with boundary the dotted circle, we can push it into D^4 to obtain the co-core of h^2, and removing a neighbourhood of it amounts to removing h^2, leaving us with h^1.

After attaching all the 1-handles, we obtain the boundary connected sum of a number of copies of $S^1 \times D^3$, whose boundary is $\#_n S^1 \times S^2$. The 2-handles are attached along a framed link L in $\#_n S^1 \times S^2$. If we denote the 1-handles with pairs of 2-spheres, a component of L is a collection of arcs in S^3 with endpoints on the 2-spheres, which are paired up using the reflections. In dotted circle notation, we can visualise L as a usual link in S^3. The framing can be encoded by a parallel copy F of L. When there are no 1-handles, we can encode the framing by an integer on each component of L, which gives the linking between the corresponding components of L and F. It is unknown whether every closed, simply-connected 4-manifold admits a handle decomposition without 1-handles.

After the 2-handles are attached, we have a 4-manifold X_2 with boundary a 3-manifold Y. Then the 1-handles with the framed link can be viewed as a diagram of X_2, or as a diagram of the boundary 3-manifold Y. If X is closed, we further have to attach some 3-handles and a single 4-handle. For this to be possible, Y has to be a connected sum of copies of $S^1 \times S^2$. If this is the case, by the following result of Laudenbach and Poénaru, this uniquely determines X, up to diffeomorphism:

PROPOSITION 4.1.1. *Every automorphism of $\#_k S^1 \times S^2$ extends to $\natural_k S^1 \times D^3$.*

In other words, there is an essentially unique way to attach the 3- and 4-handles, which we hence do not need to encode in the diagram.

EXAMPLE 4.1.2. We now give examples of Kirby diagrams of some simply-connected 4-manifolds; see Figure 17. One can obtain diagrams for connected sums of these by taking their disjoint union.

The complex projective plane $\mathbb{CP}^2$ with the complex orientation is represented by the unknot with framing $+1$. Indeed, if we remove a ball – thought of as a 4-handle – from $\mathbb{CP}^2$, we are left with the total space of the tautological line bundle over $\mathbb{CP}^1$. (A point of $\mathbb{CP}^1$ is a complex 1-dimensional subspac L of $\mathbb{C}^2$. The fibre of the tautological line bundle over L is L itself.) A generic section intersects the 0-section algebraically once as two complex lines in $\mathbb{CP}^2$ intersect once positively. Hence, the total space can be obtained by gluing two copies of $D^2 \times D^2$ along $S^1 \times D^2$ with framing $+1$. We identify one copy of $D^2 \times D^2$ with the 0-handle B^4. Then the other is a 2-handle attached to B^4 along the unkot $S^1 \times \{0\}$ with framing $+1$. We denote by $\overline{\mathbb{CP}}^2$ the complex projective plane with its orientation reversed. It can be represented by an unknot with framing -1.

FIGURE 18. Handle sliding the link component L_i with framing r_i over the component L_j with framing r_j.

Next, we describe a Kirby diagram for $S^2 \times S^2$. We write S^2 as the union of the hemispheres S^2_+ and S^2_-. Then

$$S^2 \times S^2 = (S^2_- \times S^2_-) \cup (S^2_+ \times S^2_-) \cup (S^2_- \times S^2_+) \cup (S^2_+ \times S^2_+).$$

We view $S^2_- \times S^2_-$ as a 0-handle, $S^2_+ \times S^2_-$ and $S^2_- \times S^2_+$ as 2-handles, and $S^2_+ \times S^2_+$ as a 4-handle. The 2-handles are attached along $\{0\} \times \partial S^2_-$ and $\partial S^2_- \times \{0\}$, respectively, which form a Hopf link. Both components of the Hopf link have framing 0, since the self-intersection of $S^2 \times \{p\}$ and $\{p\} \times S^2$ is 0 for $p \in S^2$.

By the Lickorish–Wallace theorem (Theorem 3.9.1), every closed, oriented 3-manifold Y admits a Kirby diagram without 1-handles. Let L be the framed link of this Kirby diagram, and we denote by M_L the corresponding 4-manifold with $\partial M_L = Y$. The boundary is unchanged under the following two Kirby moves:

(1) **Blow-up:** We add or remove an unknot with framing ± 1 that is unlinked from L. If L' is obtained by adding a $+1$-framed unknot to L, then $M_{L'} = M_L \# \mathbb{CP}^2$. Similarly, $M_{L'} = M_L \# \overline{\mathbb{CP}}^2$ if L' is obtained from L by adding a (-1)-framed unknot.

(2) **Handle slide:** We do a handle slide among the 2-handles. If L_i and L_j are the two components corresponding to the handles involved, with framings F_i and F_j, respectively, then we choose a band connecting L_i and F_j (i.e., an embedding $b \colon I \times I \hookrightarrow M_L$ such that $b(\{0\} \times I) \subset L_i$ and $b(\{1\} \times I) \subset F_j$, while the rest of $b(I \times I)$ is disjoint from L and F), and we replace L_i with $L_i' := L_i \#_b F_j$. For $k \neq i$, we set $L_k' := L_k$ and $F_k' := F_k$. The framing of L_i' is $r_i + r_j \pm 2\mathrm{lk}(L_i, L_j)$, depending on whether the band is orientation-preserving, where $r_k = \mathrm{lk}(L_k, F_k)$. For an illustration, see Figure 18.

While blow-up changes the underlying 4-manifold, handle slides do not.

There are two S^2-bundles over S^2. Indeed, since D^2 is contractible, the bundle is trivial over the Northern and Southern hemispheres, and the bundle is defined by the gluing map along the equator. This is given by an elment of $\pi_1(\mathrm{SO}(3)) \cong \mathbb{Z}_2$. We denote the non-trivial S^2-bundle over S^2 by $S^2 \mathbin{\widetilde{\times}} S^2$.

EXERCISE 4.1.3. Using Kirby calculus, show that

$$(S^2 \times S^2) \# \mathbb{CP}^2 \approx \mathbb{CP}^2 \# \mathbb{CP}^2 \# \overline{\mathbb{CP}}^2.$$

Furthermore, $S^2 \mathbin{\widetilde{\times}} S^2 \approx \mathbb{CP}^2 \# \overline{\mathbb{CP}}^2$.

The real power of Kirby calculus comes form the following result of Kirby [**33**]:

THEOREM 4.1.4. *Two Kirby diagrams without 1-handles represent diffeomorphic 3-manifolds if and only if they are related by a finite sequence of Kirby moves* (1) *and* (2).

There is also a Kirby calculus for 4-manifolds: Let X be a smooth 4-manifold, and f_0, f_1 two ordered Morse functions on it with a single index 0 critical point. Then we can connect them with a generic 1-parameter family of smooth functions f_t for $t \in I$ such that every f_t is ordered for generic t, and has a unique index 0 critical point. If we consider the Kirby diagrams of X corresponding to f_t, they change using the following moves:

(1) An isotopy of the 1- and 2-handles.
(2) A 1-handle slides over a 1-handle, or a 2-handle slides over a 1- or 2-handle.
(3) An index 1-2 or 2-3 birth or death.

In dotted circle notation, an index 1-2 birth results in an unknotted 2-handle, which is linked by a dotted circle. In general, one can cancel a 1- and a 2-handle whenever the 2-handle passes through the 1-handle exactly once.

When we are considering Kirby diagrams of closed 4-manifolds; i.e., the boundary of the 2-handlebody is a connected sum of copies of $S^1 \times S^2$ that we fill with 3-handles and a 4-handle, an index 2-3 birth corresponds to the appearance of an unknotted, 0-framed 2-handle.

For an in-depth study of 4-manifold topology from the point of view of Kirby calculus, see the books of Gompf and Stipsicz [**15**] and Akbulut [**2**].

4.2. The intersection form and the classification of 4-manifolds

Let X be a closed, connected, oriented, and simply-connected 4-manifold. Then $H_1(X) = H_3(X) = 0$. Hence $H_2(X)$ and $H^2(X)$ are free abelian by the universal coefficient theorem. We denote the fundamental class by $[X] \in H_4(X)$. Then the intersection form $Q_X \colon H^2(X) \times H^2(X) \to \mathbb{Z}$ of X is given by

$$Q_X(a, b) := \langle a \cup b, [X] \rangle$$

for a, $b \in H^2(X)$. This is a non-degenerate, symmetric bilinear form. It is *unimodular*; i.e., induces an isomorphism

$$H^2(X) \to \mathrm{Hom}(H^2(X), \mathbb{Z}) \cong H_2(X)$$

by Poincaré duality. Hence, its matrix in any basis is invertible over $\mathbb{Z}$; i.e., has determinant ± 1.

PROPOSITION 4.2.1. *Let X be a 4-manifold, and $s \in H^2(X)$ a cohomology class. Then there exists an oriented surface S properly embedded in X such that $[S] \in H_2(X)$ is Poincaré dual to s.*

PROOF. Since the Eilenberg–MacLane space $K(\mathbb{Z}, 2)$ is $\mathbb{CP}^\infty$, we have

$$H^2(X) \cong [X, \mathbb{CP}^\infty].$$

Then s corresponds to a continuous map $f_s \colon X \to \mathbb{CP}^\infty$ such that $f^*x = s$, where x is the generator of $H^2(\mathbb{CP}^\infty) \cong \mathbb{Z}[x]$, which is represented by the hypersurface $\mathbb{CP}^{\infty-1}$. We can homotope f such that it becomes smooth and transverse to $\mathbb{CP}^{\infty-1}$. Then we let $S := f_s^{-1}(\mathbb{CP}^{\infty-1})$. □

If A and B are surfaces Poincaré dual to a, $b \in H^2(X)$, then $Q_X(a, b)$ is the algebraic intersection number of A and B. This justifies the term "intersection form."

We can easily read off the intersection form of a 4-manifold from its Kirby diagram. Suppose that the 4-manifold X is given by a Kirby diagram without

1-handles and 3-handles; i.e., a framed link $L = L_1 \cup \cdots \cup L_n$ in S^3 with framing coefficient $r_i \in \mathbb{Z}$ for $i \in \{1, \dots, n\}$. Then we obtain a basis $b_1, \dots, b_n$ of $H_2(X)$ by capping off a Seifert surface of L_i with the core of the 2-handle attached along L_i. Furthermore, $Q_X(b_i, b_j) = \mathrm{lk}(L_i, L_j)$ if $i \neq j$ and $Q_X(b_i, b_i) = r_i$ by Lemma 3.4.8.

We can also consider the intersection form on $H^2(X; \mathbb{R})$. Nondegenerate, symmetric bilinear forms over *real* vector spaces are completely determined by their rank and signature. We write $b_2^+(X)$ and $b_2^-(X)$ for the dimensions of maximal positive and negative definite subspaces of $H^2(X; \mathbb{R})$. The signature of Q_X is $\sigma(X) = b_2^+(X) - b_2^-(X)$. Signature is a cobordism invariant, and it vanishes if and only if X is oriented 0-cobordant; see Section 1.7.

We say that Q_X is *positive definite* if $b_2^+(X) = b_2(X)$, *negative definite* if $b_2^-(X) = b_2(X)$, and is *indefinite* otherwise.

DEFINITION 4.2.2. The intersection form Q_X is *even* if $Q(x, x)$ is even for every $x \in H^2(X)$, and is *odd* otherwise. We call this the *type* of X.

DEFINITION 4.2.3. The Lie group $\mathrm{Spin}(n)$ is the connected double cover of $\mathrm{SO}(n)$ for $n \geq 2$. The oriented n-manifold M is *spin* if the structure group of TM can be lifted from $\mathrm{SO}(n)$ to $\mathrm{Spin}(n)$. A *spin structure* on M is an equivalence class of such lifts.

PROPOSITION 4.2.4. *The intersection form Q_X is even if and only if X is spin.*

PROOF. This is because the Spin structure of D^4 extends to a 2-handle if and only if the framing of the attaching 2-sphere is even, and the self-intersection of the corresponding element of $H_2(X)$ is precisely the framing. $\square$

Indefinite intersection forms have a simple classification:

THEOREM 4.2.5. *Two indefinite, unimodular, symmetric bilinear forms over $\mathbb{Z}$ are equivalent if and only if they have the same rank, signature, and type.*

Not every value of the triple rank, signature, and type can be realised:

PROPOSITION 4.2.6. *The signature of an even form is divisible by 8.*

However, this is the only restriction for indefinite forms. An important example of a definite even form of signature 8 is the E_8 form, given by the matrix

$$E_8 = \begin{pmatrix} 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 2 \end{pmatrix}. \tag{4.2.1}$$

The 4-manifold P_{E_8} with boundary defined by the Kirby diagram in Figure 19 has intersection form E_8. Its boundary is the Poincaré homology sphere Σ_P, which is the Brieskorn manifold $M(2, 3, 5)$; see Definition 3.9.3. By the work of Freedman [**11**], there is a topological 4-manifold Δ with boundary Σ_P that has the same homotopy groups as D^4. The E_8 *manifold* $M_{E_8} = P_{E_8} \cup \Delta$ is a closed topologyical 4-manifold with intersection form E_8.

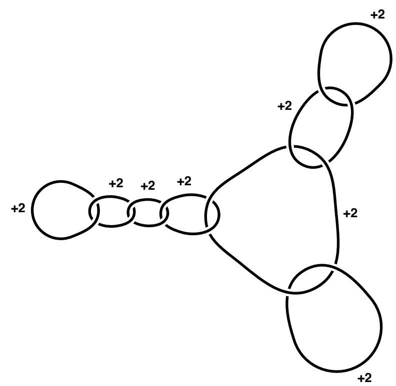

FIGURE 19. A Kirby diagram for a 4-manifold whose intersection form is the E_8 lattice.

The hyperbolic form

$$H = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

has rank 2 and signature 0. (Note that $Q_{S^2 \times S^2} = H$.) Hence, every symmetric, unimodular, indefinite, even form is isomorphic to $kE_8 \oplus lH$ for some $k \in \mathbb{Z}$ and $l \in \mathbb{Z}_+$. On the other hand, every indefinite odd form is isomorphic to $k(1) \oplus l(-1)$ for k, $l \in \mathbb{Z}_+$.

Not every symmetric, unimodular bilinear form can be the intersection form of a *smooth* 4-manifold, due to the following result of Rokhlin [**60**]:

THEOREM 4.2.7. *Let X be a closed, smooth, spin 4-manifold (i.e., Q_X is even). Then $\sigma(X)$ is divisible by 16.*

If M is a closed topological n-manifold, then the Kriby–Siebenmann class $\kappa(M) \in H^4(M; \mathbb{Z}_2)$ vanishes if M admits a PL structure. When $\dim(M) \geq 5$, the converse also holds. For further details, see the book of Kirby and Siebenmann [**34**]. When X is a spin 4-manifold,

$$\kappa(X) \equiv \sigma(X)/8 \mod 2.$$

A 4-manifold has a PL structure if and only if it has a smooth structure, and so a smooth 4-manifold has $\kappa(X) = 0$.

EXAMPLE 4.2.8. An example of a smooth spin 4-manifold with signature -16 is the K3 surface. It can be constructed as the solution to

$$x^4 + y^4 + z^4 + w^4 = 0$$

in $\mathbb{CP}^3$. Its intersection form is $-2E_8 \oplus 3H$.

From the point of view of complex geometry and algebraic geometry, there are many different K3 surfaces, but they are all diffeomorphic by the work of Kodaira.

It follows from Rokhlin's theorem (Theorem 4.2.7) that a closed topological 4-manifold with intersection form E_8 (e.g., M_{E_8}) does not admit a smooth structure. By a celebrated result of Freedman [**11**][**12**], the pair $(Q_X, \kappa(X))$ is a complete invariant of topological, simply-connected 4-manifolds X:

THEOREM 4.2.9. *Two closed, simply-connected topological 4-manifolds are homeomorphic if and only if they have isomorphic intersection forms and the same Kirby–Siebenmann class.*

Conversely, for any symmetric, unimodular bilinear form Q over $\mathbb{Z}$, there is a unique topological 4-manifold X with $Q_X = Q$ if Q is even, and two if Q is odd, distinguished by $\kappa(X)$.

Note that, when Q_X is odd and $\kappa(X) = 1$, then X does not admit a smooth structure. However, $\kappa(X) = 0$ does not guarantee that X is smoothable.

In Theorem 4.2.5, we saw that the indefinite forms have a simple classification. On the other hand, there is a wild world of definite forms. The following surprising result of Donaldson [**10**] implies that most of these forms do not arise as intersection forms of *smooth* 4-manifolds:

THEOREM 4.2.10. *Let X be a closed, smooth, simply-connected 4-manifold such that Q_X is positive (or negative) definite. Then Q_X is diagonalisable; i.e., isomorphic to $n(1)$ (or $n(-1)$).*

Combined with Freedman's theorem, we obtain that every smooth, definite 4-manifold is homeomorphic to either $\#_n\mathbb{CP}^2$ or $\#_n\overline{\mathbb{CP}}^2$.

By Theorem 4.2.5, indefinite forms are determined by their rank, signature, and type. In particular, the indefinite odd forms are $k(1) \oplus l(-1)$, so every 4-manifold with such an intersection form is homeomorphic to $k\mathbb{CP}^2\#l\overline{\mathbb{CP}}^2$ for k, $l > 0$.

What remains open is the geography problem for smooth, indefinite, even (i.e., spin) 4-manifolds. Namely, which of these intersection forms are represented by smooth 4-manifolds. Up to homeomorphism, they are all determined by their rank and signature. And we have seen that they have intersection forms $2kE_8 \oplus lH$ for $k \in \mathbb{Z}$ and $l > 0$. We can suppose that $k \geq 0$ by possibly reversing the orientation of the manifold. If $l \geq 3k$, then $2kE_8 \oplus lH$ is the intersection form of $kK3\#(l-3k)(S^2 \times S^2)$. The condition $l \geq 3k$ is equivalent to $b_2(X) \geq \frac{11}{8}\sigma(X)$. The $\frac{11}{8}$-conjecture states that this condition is also necessary for $2kE_8 \oplus lH$ to be the intersection form of a smooth, simply-connected 4-manifold. The best known result, due to Furuta [**14**], states that $b_2(X) \geq \frac{10}{8}\sigma(X)$ for every smooth, indefinite, even 4-manifold.

Another difficult question is the classification of smooth structures up to diffeomorphism on a given topological 4-manifold. We have seen that in higher dimensions, there are manifolds that admit more than one smooth structure, but the number is always finite. In dimensions below 4, every manifold admits a unique smooth structure. In contrast, Taubes has shown that $\mathbb{R}^4$ admits continuum many non-diffeomorphic smooth structures. Note that $\mathbb{R}^n$ has a unique smooth structure for $n \neq 4$.

If X is a closed topological 4-manifold, then it can be represented by a finite Kirby diagram, and hence admits at most countably infinite pairwise non-diffeomorphic smooth structures. The K3 surface, for example, admits infinitely many smooth structures, and so does $\mathbb{CP}^2\#k\overline{\mathbb{CP}}^2$ for $k \geq 3$. There is no known example of a smooth 4-manifold that admits only finitely many smooth structures. One of the most difficult open problems in topology is the smooth 4-dimensional Poincaré conjecture, which asks whether S^4 admits a unique smooth structure. To distinguish smooth structures, the main tool is the Seiberg–Witten invariant. For an introduction, see the book of Morgan [**49**]. However, this is only defined when $b_2^+(M) > 0$.

For a thorough and entertaining overview of 4-manifold topology, see the book of Scorpan [**64**].

Bibliography

[1] Ian Agol. The virtual Haken conjecture. *Doc. Math.*, 18:1045–1087, 2013. With an appendix by Agol, Daniel Groves, and Jason Manning.

[2] Selman Akbulut. *4-manifolds*, volume 25 of *Oxford Graduate Texts in Mathematics*. Oxford University Press, Oxford, 2016.

[3] Steven A. Bleiler, Craig D. Hodgson, and Jeffrey R. Weeks. Cosmetic surgery on knots. In *Proceedings of the Kirbyfest (Berkeley, CA, 1998)*, volume 2 of *Geom. Topol. Monogr.*, pages 23–34. Geom. Topol. Publ., Coventry, 1999.

[4] Gerhard Burde, Heiner Zieschang, and Michael Heusener. *Knots*, volume 5 of *De Gruyter Studies in Mathematics*. De Gruyter, Berlin, extended edition, 2014.

[5] Andrew Casson. Three-dimensional topology. unpublished lecture notes.

[6] Jean Cerf. Topologie de certains espaces de plongements. *Bull. Soc. Math. France*, 89:227–380, 1961.

[7] Jean Cerf. *Sur les difféomorphismes de la sphère de dimension trois* ($\Gamma_4 = 0$). Lecture Notes in Mathematics, No. 53. Springer-Verlag, Berlin-New York, 1968.

[8] Ralph L. Cohen. The immersion conjecture for differentiable manifolds. *Ann. of Math. (2)*, 122(2):237–328, 1985.

[9] Marc Culler, Nathan M. Dunfield, Matthias Goerner, and Jeffrey R. Weeks. SnapPy, a computer program for studying the geometry and topology of 3-manifolds. Available at http://snappy.computop.org (23/05/2021).

[10] S. K. Donaldson. An application of gauge theory to four-dimensional topology. *J. Differential Geom.*, 18(2):279–315, 1983.

[11] Michael H. Freedman. The topology of four-dimensional manifolds. *J. Differential Geometry*, 17(3):357–453, 1982.

[12] Michael H. Freedman and Frank Quinn. *Topology of 4-manifolds*, volume 39 of *Princeton Mathematical Series*. Princeton University Press, Princeton, NJ, 1990.

[13] D. B. Fuks and V. A. Rokhlin. *Beginner's course in topology*. Universitext. Springer-Verlag, Berlin, 1984. Geometric chapters, Translated from the Russian by A. Iacob, Springer Series in Soviet Mathematics.

[14] M. Furuta. Monopole equation and the $\frac{11}{8}$-conjecture. *Math. Res. Lett.*, 8(3):279–291, 2001.

[15] Robert E. Gompf and András I. Stipsicz. 4*-manifolds and Kirby calculus*, volume 20 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1999.

[16] C. McA. Gordon and R. A. Litherland. On the signature of a link. *Invent. Math.*, 47(1):53–69, 1978.

[17] Cameron McA. Gordon and John Luecke. Knots are determined by their complements. *J. Amer. Math. Soc.*, 2(2):371–415, 1989.

[18] Wolfgang Haken. Theorie der Normalflächen. *Acta Math.*, 105:245–375, 1961.

[19] Wolfgang Haken. Über das Homöomorphieproblem der 3-Mannigfaltigkeiten. I. *Math. Z.*, 80:89–120, 1962.

[20] A. J. S. Hamilton. The triangulation of 3-manifolds. *Quart. J. Math. Oxford Ser. (2)*, 27(105):63–70, 1976.

[21] Allen Hatcher. Notes on basic 3-manifold topology. https://pi.math.cornell.edu/~hatcher/3M/3Mfds.pdf, 2007.

[22] Allen E. Hatcher. The Kirby torus trick for surfaces. *arXiv:1312.3518*, 2022.

[23] Geoffrey Hemion. On the classification of homeomorphisms of 2-manifolds and the classification of 3-manifolds. *Acta Math.*, 142(1-2):123–155, 1979.

[24] John Hempel. 3*-Manifolds*. Princeton University Press, Princeton, N. J.; University of Tokyo Press, Tokyo, 1976. Ann. of Math. Studies, No. 86.

[25] Morris W. Hirsch and Barry Mazur. *Smoothings of piecewise linear manifolds.* Princeton University Press, Princeton, N. J.; University of Tokyo Press, Tokyo, 1974. Annals of Mathematics Studies, No. 80.

[26] William H. Jaco and Peter B. Shalen. Seifert fibered spaces in 3-manifolds. *Mem. Amer. Math. Soc.*, 21(220):viii+192, 1979.

[27] Klaus Johannson. *Homotopy equivalences of* 3*-manifolds with boundaries*, volume 761 of *Lecture Notes in Mathematics*. Springer, Berlin, 1979.

[28] Vaughan F. R. Jones. A polynomial invariant for knots via von Neumann algebras. *Bull. Amer. Math. Soc. (N.S.)*, 12(1):103–111, 1985.

[29] Jeremy Kahn and Vladimir Marković. Counting essential surfaces in a closed hyperbolic three-manifold. *Geom. Topol.*, 16(1):601–624, 2012.

[30] Jeremy Kahn and Vladimir Markovic. Immersing almost geodesic surfaces in a closed hyperbolic three manifold. *Ann. of Math. (2)*, 175(3):1127–1190, 2012.

[31] Louis H. Kauffman. State models and the Jones polynomial. *Topology*, 26(3):395–407, 1987.

[32] Michel A. Kervaire and John W. Milnor. Groups of homotopy spheres. I. *Ann. of Math. (2)*, 77:504–537, 1963.

[33] Robion Kirby. A calculus for framed links in S^3. *Invent. Math.*, 45(1):35–56, 1978.

[34] Robion C. Kirby and Laurence C. Siebenmann. *Foundational essays on topological manifolds, smoothings, and triangulations.* Princeton University Press, Princeton, N.J.; University of Tokyo Press, Tokyo, 1977. With notes by John Milnor and Michael Atiyah, Annals of Mathematics Studies, No. 88.

[35] Hellmuth Kneser. Geschlossen Flächen in dreidimensionalen Mannigfaltigkeiten. *Jahresbericht der Deutschen Mathematiker Vereinigung*, 38:248–260, 1929.

[36] P. B. Kronheimer and T. S. Mrowka. Khovanov homology is an unknot-detector. *Publ. Math. Inst. Hautes Études Sci.*, 113:97–208, 2011.

[37] W. B. R. Lickorish and K. C. Millett. The new polynomial invariants of knots and links. *Math. Mag.*, 61(1):3–23, 1988.

[38] William B. R. Lickorish. A representation of orientable combinatorial 3-manifolds. *Ann. of Math. (2)*, 76:531–540, 1962.

[39] Ciprian Manolescu. Lectures on the triangulation conjecture. In *Proceedings of the Gökova Geometry-Topology Conference 2015*, pages 1–38. Gökova Geometry/Topology Conference (GGT), Gökova, 2016.

[40] William W. Menasco and Morwen B. Thistlethwaite. The Tait flyping conjecture. *Bull. Amer. Math. Soc. (N.S.)*, 25(2):403–412, 1991.

[41] John Milnor. *Singular points of complex hypersurfaces.* Annals of Mathematics Studies, No. 61. Princeton University Press, Princeton, N.J.; University of Tokyo Press, Tokyo, 1968.

[42] John W. Milnor. A unique decomposition theorem for 3-manifolds. *Amer. J. Math.*, 84:1–7, 1962.

[43] John W. Milnor. *Morse theory.* Based on lecture notes by M. Spivak and R. Wells. Annals of Mathematics Studies, No. 51. Princeton University Press, Princeton, N.J., 1963.

[44] John W. Milnor. *Topology from the differentiable viewpoint.* The University Press of Virginia, Charlottesville, Va., 1965. Based on notes by David W. Weaver.

[45] John W. Milnor. On the 3-dimensional Brieskorn manifolds $M(p, q, r)$. In *Knots, groups, and 3-manifolds (Papers dedicated to the memory of R. H. Fox)*, volume 84 of *Ann. of Math. Studies*, pages 175–225. Princeton University Press, 1975.

[46] John W. Milnor and James D. Stasheff. *Characteristic classes.* Princeton University Press, Princeton, N. J.; University of Tokyo Press, Tokyo, 1974. Annals of Mathematics Studies, No. 76.

[47] Edwin E. Moise. Affine structures in 3-manifolds. V. The triangulation theorem and Hauptvermutung. *Ann. of Math. (2)*, 56:96–114, 1952.

[48] Edwin E. Moise. *Geometric topology in dimensions* 2 *and* 3. Springer-Verlag, New York-Heidelberg, 1977. Graduate Texts in Mathematics, Vol. 47.

[49] John W. Morgan. *The Seiberg-Witten equations and applications to the topology of smooth four-manifolds*, volume 44 of *Mathematical Notes.* Princeton University Press, Princeton, NJ, 1996.

[50] George D. Mostow. *Strong rigidity of locally symmetric spaces.* Princeton University Press, Princeton, N.J.; University of Tokyo Press, Tokyo, 1973. Annals of Mathematics Studies, No. 78.

[51] James R. Munkres. *Elementary differential topology*, volume 1961 of *Lectures given at Massachusetts Institute of Technology, Fall*. Princeton University Press, Princeton, N.J., 1966.

[52] James R. Munkres. *Topology*. Prentice Hall, Inc., Upper Saddle River, NJ, 2000. Second edition of [MR0464128].

[53] Kunio Murasugi. Jones polynomials and classical conjectures in knot theory. *Topology*, 26(2):187–194, 1987.

[54] Richard S. Palais. Local triviality of the restriction map for embeddings. *Comment. Math. Helv.*, 34:305–312, 1960.

[55] C. D. Papakyriakopoulos. On Dehn's lemma and the asphericity of knots. *Ann. of Math. (2)*, 66:1–26, 1957.

[56] Gopal Prasad. Strong rigidity of **Q**-rank 1 lattices. *Invent. Math.*, 21:255–286, 1973.

[57] Tibor Radó. Über den Begriff der Riemannschen Fläche. *Acta Sci. Math. Szeged.*, 2:101–121, 1925.

[58] Kurt Reidemeister. Elementare Begründung der Knotentheorie. *Abh. Math. Sem. Univ. Hamburg*, 5(1):24–32, 1927.

[59] Kurt Reidemeister. Homotopieringe und Linsenräume. *Abh. Math. Sem. Univ. Hamburg*, 11(1):102–109, 1935.

[60] Vladimir A. Rohlin. New results in the theory of four-dimensional manifolds. *Doklady Akad. Nauk SSSR (N.S.)*, 84:221–224, 1952.

[61] Dale Rolfsen. *Knots and links*, volume 7 of *Mathematics Lecture Series*. Publish or Perish, Inc., Houston, TX, 1990. Corrected reprint of the 1976 original.

[62] Horst Schubert. Knoten und Vollringe. *Acta Math.*, 90:131–286, 1953.

[63] Matthias Schwarz. *Morse homology*, volume 111 of *Progress in Mathematics*. Birkhäuser Verlag, Basel, 1993.

[64] Alexandru Scorpan. *The wild world of 4-manifolds*. American Mathematical Society, Providence, RI, 2005.

[65] Stephen Smale. Generalized Poincaré's conjecture in dimensions greater than four. *Ann. of Math. (2)*, 74:391–406, 1961.

[66] Norman Steenrod. *The topology of fibre bundles*. Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 1999. Reprint of the 1957 edition, Princeton Paperbacks.

[67] Morwen B. Thistlethwaite. A spanning tree expansion of the Jones polynomial. *Topology*, 26(3):297–309, 1987.

[68] René Thom. Quelques propriétés globales des variétés différentiables. *Comment. Math. Helv.*, 28:17–86, 1954.

[69] William P. Thurston. *Three-dimensional geometry and topology. Vol. 1*, volume 35 of *Princeton Mathematical Series*. Princeton University Press, Princeton, NJ, 1997. Edited by Silvio Levy.

[70] Bruce Trace. On the Reidemeister moves of a classical knot. *Proc. Amer. Math. Soc.*, 89(4):722–724, 1983.

[71] Friedhelm Waldhausen. Heegaard-Zerlegungen der 3-Sphäre. *Topology*, 7:195–203, 1968.

[72] Friedhelm Waldhausen. On irreducible 3-manifolds which are sufficiently large. *Ann. of Math. (2)*, 87:56–88, 1968.

[73] Andrew H. Wallace. Modifications and cobounding manifolds. *Canad. J. Math.*, 12:503–528, 1960.

[74] H. Whitney. The self-intersections of a smooth n-manifold in $2n$-space. *Ann. of Math.*, 45:220–246, 1944.