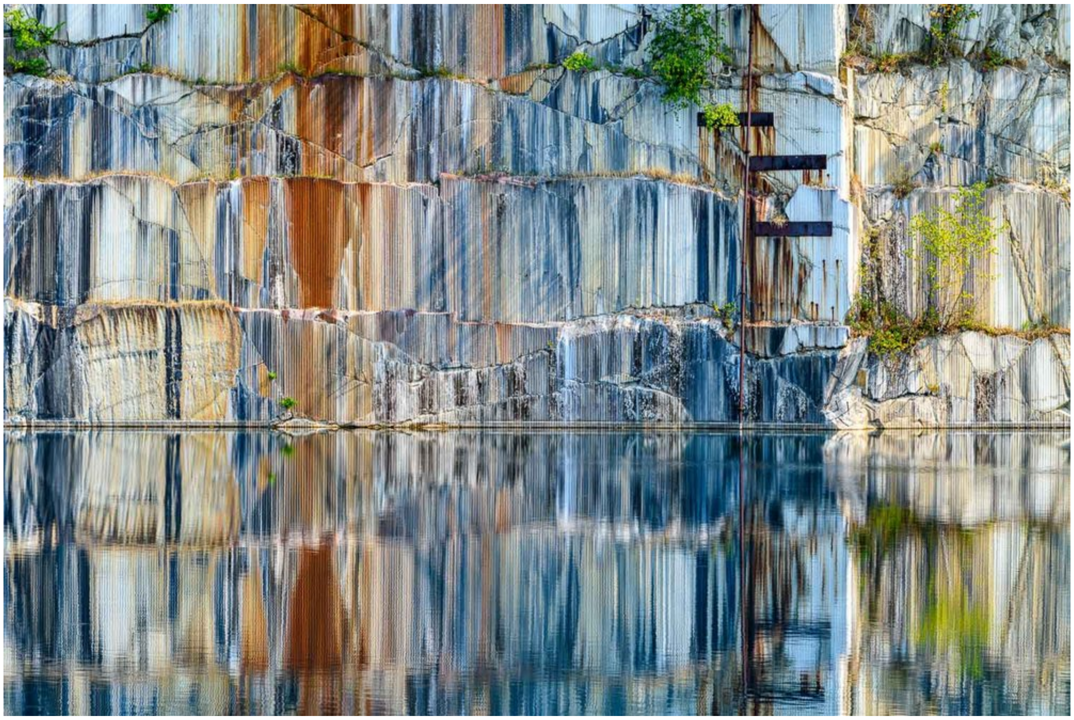


Geometric Topology

Simon Salamon

Draft lecture notes
September 2024

These notes originate from the KCL modules 6ccm327a and 7ccm327b, taught with common lectures in 2021–2023.

They are subject to correction and amendment as the 2024 semester proceeds, and will not correspond exactly to the lectures in 2024/25. Some topics from the notes will be omitted in lectures, and others may be covered and explained in more detail.

The notes should therefore be regarded as a supplement, and not an alternative, to the lectures.

The overall concept of the module owes a big debt to the book by Gilbert and Porter, though these notes skim over many chapters in that book.

The author learnt a lot of the material on knots from Warwick lecture notes by Brian Sanderson and Aberdeen lecture notes by Richard Hepworth, both were available online.

Cover photograph courtesy of Kip King, Norwich, Vermont

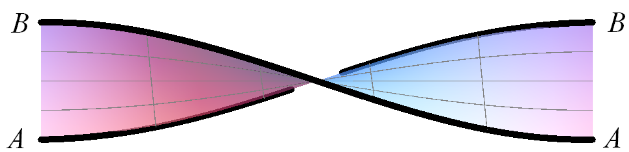
Contents

0 Introduction

This section has been included because it served as the basis of two introducory lectures recorded in 2021, and available on the Keats page.

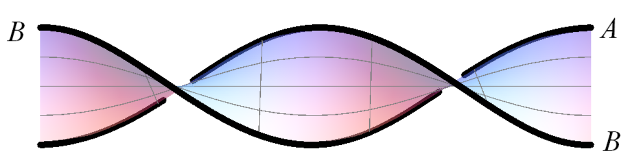
0.1 Knots and links

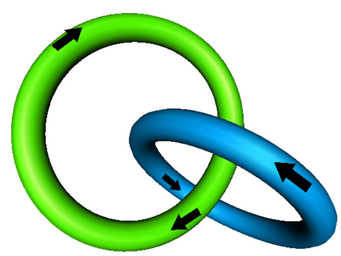
Möbius band. This is formed by joining the end of a strip of paper twisted once, so that A joins to A and B to B. The image is the logo of a commercial buildings maintenance company, seen in Surrey Street. What does its boundary look like in space?



This question is answered by cutting the strip along the midline: one gets a single strip (twisted twice, but that is irrelevant as we are regarding the strip as the boundary curve). Thus, the boundary is the **unknot** – a single unkotted loop. One can also predict this result without experimenting, by drawing a **knot diagram** for the boundary. What happens if we insert more twists?

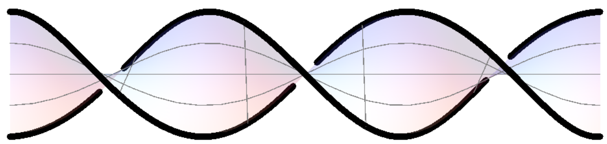
Hopf link. The boundary of a 2-twisted band has two components. Applying a suitable homeomorphism $\mathbb{R}^3 \to \mathbb{R}^3$ (deformation of space), we get two circles whose non-separability can be detected by a non-zero **linking number**:

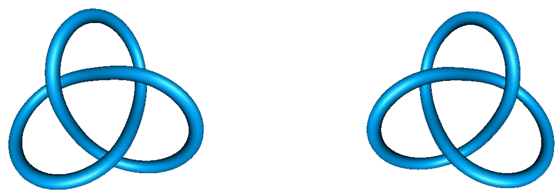




Ignore the arrows, which will only concern us later in the module, when we discuss oriented links.

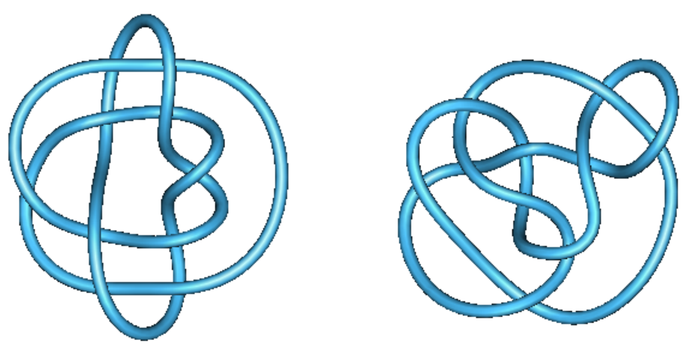
Two trefoil knots. The boundary of a 3-twisted band is an everyday knot, homeomorphic to either to the left-handed or right-handed version:



$L3_1$  $R3_1$

The correct notion of knot equivalence is **ambient isotopy** or continuous deformation. It is easy to accept that a trefoil knot is not isotopic to its mirror image, though it will take time before we can prove this mathematically. As space curves represented in blue, their torsions (computed using the Serret-Frenet equations) have different signs, but this module will not make use of differential calculus.

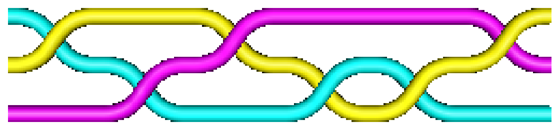
Perko pair. These are the two knots shown. Rather than thinking of the images as curves in space, wwe may regard them as plane diagrams together with crossing data, so when two strands cross we know which one is on top:

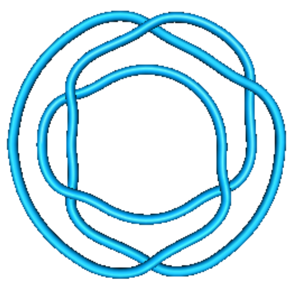


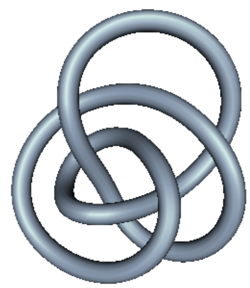
These two knots are in fact ambient isotopic, a fact recognized in 1973 by the lawyer after whom the pair is named. This can be demonstrated by applying a sequence of **Reidemeister moves** to pass from one diagram to the other.

The pair had been listed as distinct knots in earlier tables, which now display 165 different 'prime' knots with a minimum of 10 crossings. The Perko knot is catalogued as the 161st knot that can be drawn with 10 (and no less) crossings, so is catalogued with the symbol 10_{161}. However, there are still two distinct knots, because (like the trefoil) the mirror images are not ambient isotopic!

Braids to knots. One way to construct one or more knots is to close a braid with $n \geqslant 2$ strands. The example below gives rise to a single knot, though it takes a bit of effort to 'catalogue' it:







The first (left-hand) diagram is not alternating (the crossinsg do not go over-under-over-under etc), and the same knot can be drawn with less than 6 crossings (right).

We shall not consider braids in this module, but there is a fascinating relationship with prime numbers.

Jones polynomial. This was discovered in 1984, and led to big advances in knot theory after that. It's actually a Laurent polynomial. The trefoil knots have

$$\begin{aligned} V(L3_1) &= -t^4 + t^3 + t \\ V(R3_1) &= -t^{-4} + t^{-3} + t^{-1}. \end{aligned}$$

The definitions confirm that taking a mirror image of a knot corresponds to swapping

$$t \leftrightarrow t^{-1}.$$

Theorem. If a knot K has a 'reduced' alternating diagram with c crossings then c equals the difference between highest and lowest powers of t in $\mathbf{V}(K)$.

The Perko knot has

$$\mathbf{V}(10_{161}) = t^{-3} + t^{-6} - t^{-7} + t^{-8} - t^{-9} + t^{-10} - t^{-11},$$

and no alternating diagram. The form of $\mathbf{V}$ shows that it is not equivalwnt to its mirrow image.

0.2 Surfaces

A surface or (better) a 2-manifold is a topological space in which each point has a neighbourhood homeomorphic to (an open disk in) $\mathbb{R}^2$.

Classification theorem. A compact surface without boundary is determined by whether or not it is *orientable* and its Euler number or characteristic

$$\chi = V - E + F.$$

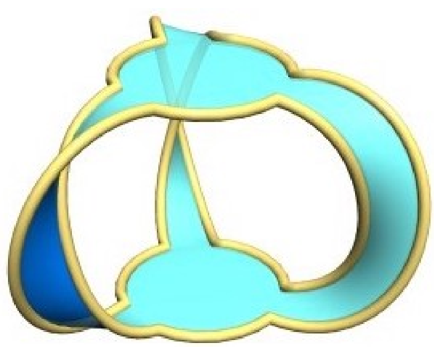
This can be computed from a legitimate subdivision of the surface, such as a triangulation (in which the regions are triangles).

$$\begin{array}{ll} \text{oriented:} & \chi = 2, 0, -2, -4, -6, \ldots \\ \text{non-oriented:} & \chi = 1, 0, -1, -2, -3, -4, \ldots \end{array}$$

Surfaces without boundary can be visualized in space only if they are oriented, though one can also construct surfaces abstractly. If the surface is oriented, then $\chi = 2 - 2g$ is even, where $g \geqslant 0$ is the *genus*. The sphere S has genus 0, and the torus T (surface of a doughnut with one hole) has $g = 1$. Place the symbols S and T in the right boxes below, and add the non-orienteable surfaces P (the real projective plane), and K (the Klein bottle):

	2	1	0	-1	-2	-3
orientable		$\varnothing$		$\varnothing$		$\varnothing$
non-orientable	$\varnothing$					

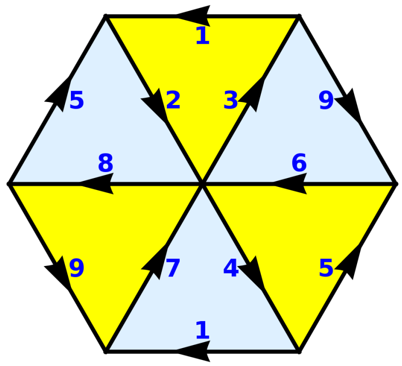
A Seifert surface. A trefoil knot is bounded by an orientable surface that is *homeomorphic* to a compact surface or 2-manifold M from which a disk has been removed (even though one cannot discern M from the blue area). In which box is M?



It is oriented because it is 2-sided in space. One can answer then the question by computing the Euler number $\chi = V - E + F$.

Converting the twisted bands into rectangles each with 4 vertices, we get $V = 12$, $E = 18$ and $F = 2 + 3 = 5$. Thus $\chi = -1$ and the surface has $\chi = 0$. It is the torus T!

Triangulation. What surface do we get if we identify parallel opposite edges of a hexagon? The consistent orientations of parallel arrows makes it oriented.

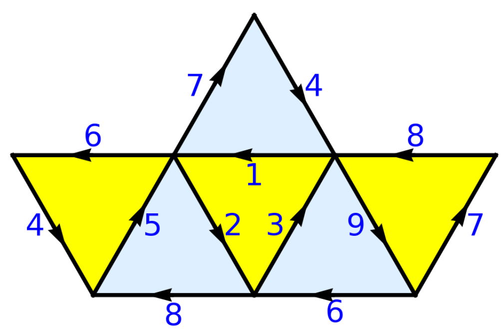


$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

$$\chi = V - E + F = 3 - 9 + 6 = 0$$

A torus again!

Cutting and pasting. One can verify that $\chi = 0$ more simply by taking $F = 1$, $E = 3$ and (the only thing to check) $V = 2$. The advantage of the triangles is that we can reaarnge them (to form a boat), and see the torus from the way the edges around the peerimeter are identified:



0.3 Fundamental group

Fix a point x_0 in a topological space X. An element of the fundamental group $\boldsymbol{\pi_1}(X, x_0)$ is represented by a path starting and ending at x_0. Two such loops are deemed to be equivalent if they are **homotopic** – one can be deformed into the other.

Using *covering maps*, we shall prove the

Theorem. Any loop in the unit circle $S^1 \subset \mathbb{C}$ based at 0 is homotopic to the loop

$$s \longmapsto e^{2\pi iks}, \qquad s \in [0,1],$$

for some $k \in \mathbb{Z}$, which is the 'winding number', as in complex analysis.

The fundamental group is a *group*; the group operation is given by going round the first loop and then the second loop. One needs to prove that this operation satisfies the group axioms: associativity, existence of an identity and inverses.

It follows that $\boldsymbol{\pi_1}(S^1)$ is the infinite cyclic group

$$F_1 = \{x^a : a \in \mathbb{Z}\}$$

where the symbol x represents the loop $s \mapsto e^{is}$. This is isomorphic to the integers $\mathbb{Z}$, but most fundamental groups we shall encounter are not abelian, so it is best to use multiplicative notation and think of $\boldsymbol{\pi_1}(S^1)$ as the **free group** F_1 on one generator.

As a topological subspace of $\mathbb{R}^3$ or as a quotient of a rectangle, the torus T^2 is homeomorphic to the product $S^1 \times S^1$ of two crcles. It follows that

$$\boldsymbol{\pi_1}(T^2) \cong \mathbb{Z} \times \mathbb{Z} = \mathbb{Z}^2.$$

This is not the same as the free group F_2 on two generators, which consists of all words

$$a^{n_1} b^{n_2} a^{n_3} \cdots b^{n_k}$$

where each n_i is an integer (possibly 0). To get $\boldsymbol{\pi_1}{}_1(T)$ we abelianize F_2 by imposing the relation $ab = ba$.

Oriented surfaces of genus $g \geqslant 2$ have non-abelian fundamental groups. The sphere has trivial fundamental group, whereas the projective plane (and the space of rotations in 3-space) has $\boldsymbol{\pi_1} \cong \mathbb{Z}_2$.

Complement of a knot. Another invariant of a knot is the fundamental group of its complement. If U is the unknot (thought of as a circle in space), then

$$\boldsymbol{\pi_1}(\mathbb{R}^3 \backslash U) \cong F_1,$$

since any loop is really determined by how many times it warps around U. For either trefoil knot, we shall see that

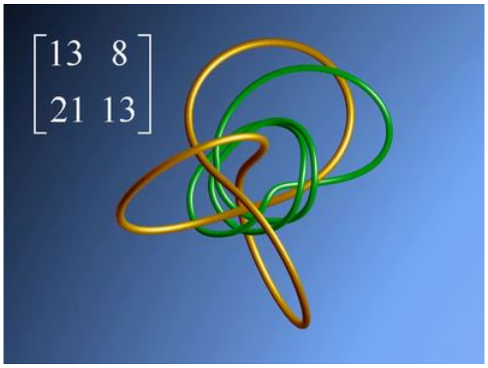
$$\boldsymbol{\pi_1}(\mathbb{R}^3 \backslash 3_1) = \left\langle a, b : a^3 = b^2 \right\rangle.$$

This is an infinte non-abelian group, but if one imploses the relations $b^2 = e$ and $bab = a^2$, one obtains the group S_3 bor order 6.

Advanced fact. The space of lattices in $\mathbb{R}^2$ can be identified with both the coset space $\mathsf{SL}(2,\mathbb{R})/\mathsf{SL}(2,\mathbb{Z})$ and

$$\{(g_2, g_3) \in \mathbb{C}^2 : g_2^3 - 27g_3^2 \neq 0\}.$$

Up to scaling this is $S^3 \backslash 3_1$. Knots in this 3-manifold arise as periodic orbits in the Lorenz attractor:



The images are taken from work of Ghys and Leys.

1 Review of topology

A topological space is a set X with a chosen family (the 'topology') of **open sets** satisfying some simple conditions (which we shan't list on this page).

Example. The usual topology of $\mathbb{R}^n$ is defined as follows. A subset U of $\mathbb{R}^3$ is open iff for each $x \in U$ one can fit a solid round ball with positive radius and centre x inside U.

This notion allows one to bypass technicalities and simplify definitions in analysis:

Definitions. A mapping $f\colon X \to Y$ is **continuous** if the inverse image of every open subset of Y is open in X.
A mapping $f\colon X \to Y$ is a **homeomorphism** if it is bijective and both f, f^{-1} are continuous (so U open $\Leftrightarrow$ $f(U)$ open). Two spaces X, Y are called **homeomorphic** if there exists such a map, and this is the natural equivalence relation on topological spaces.

Homeomorphic spaces are 'the same' or (to use the language of algebra) 'isomorphic', but note the necessity to impose in general that f^{-1} be continuous.

1.1 Induced topologies

Let X be a topological space. Any subset S of X is a topological space in its own right with the **subset topology**. For this, a subset of S is declared open iff it equals $U \cap S$ for some open subset U of X.

Example. The 2-sphere S^2, consisting of points **x** in $\mathbb{R}^3$ a unit distance from the origin, is a topological space. Indeed, it is a compact surface or 2-manifold (a concept to be defined later in the course).

If $\sim$ is an equivalence relation on X then the set

$$Q = X/\!\sim$$

of equivalence classes is a topological space with the **quotient topology**. A subset U of Q is declared open iff $q^{-1}(U)$ is open in X, where $q\colon X \to Q$ maps an element to its equivalence class.

More examples. (i) The quotient group $\mathbb{R}^2/\mathbb{Z}^2$ is homeomorphic to a torus formed by revolving a circle around a disjoint axis in $\mathbb{R}^3$; the quotient topology coincides with the subspace topology.
(ii) The real projective plane $\mathbb{RP}^2$ is the quotient of S^2 by the equivalence relation that identifies **x** and $-$**x**. (This can be written $S^2/\mathbb{Z}_2$ because the multiplicative group $\{1, -1\}$ acts on S^2 and the equivalence classes are the orbits.)

1.2 Connected spaces

Definition. A space X is **connected** if there is no proper subset U such that both U and $X \backslash U$ are open.

Theorem. A subset of $\mathbb{R}$ is connected iff it is an interval.

In practice, it is easier to apply the

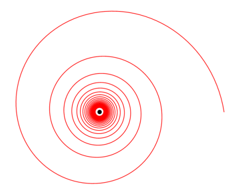
Definition. A space X is **path-connected** if any two points $x, y \in X$ can be joined by a path, i.e. there exists a continuous map $\alpha\colon [0,1] \to X$ with $\alpha(0) = x$ and $\alpha(1) = y$.

Exercises. (i) If X is path-connected then it is connected.

(ii) The subset

$$\{\tfrac{2\pi}{t}(\cos t, \sin t) : t \geqslant 2\pi\} \sqcup \{(0,0)\}$$

of $\mathbb{R}^2$ is connected but not path-connected:



(iii) If X is connected and $f\colon X \to Y$ is continuous then $f(X)$ is connected. The intermediate value theorem is a corollary.

1.3 Compact spaces

Definition. A space X is **compact** if from any covering of $X = \bigcup_\alpha U_\alpha$ by open sets we can select a finite subcovering, i.e. a finite number of the U_α whose union is X.

A subset C of X is called **closed** if its complement $X \backslash C$ is open. In $\mathbb{R}^n$, being closed is equivalent to containing all its limit points. Beware that a surface is sometimes called closed to indicate that it has no boundary, but it could still be closed in the above sense provided its boundary limit points form part of the surface (think of a cylinder).

Examples. $\mathbb{R}$ is not compact, since $\{(-n, n) : n \geqslant 1\}$ has no finite subcovering. Nor are bounded part-open intervals like $(0,1)$ or $(0,1]$.

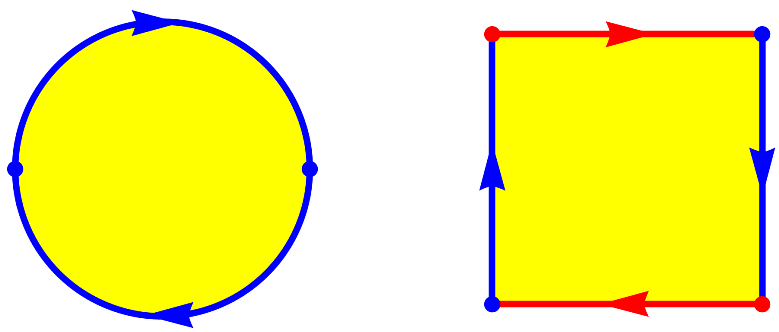
Heine-Borel theorem. A subset S of $\mathbb{R}^n$ is compact iff it is both closed and bounded.

Exercise. If X is compact and $f\colon X \to Y$ is continuous then $f(X)$ is compact. If $f\colon [a,b] \to \mathbb{R}$ is continuous it follows that its image is a closed interval $[c,d]$. As a corollary, f is bounded and attains its bounds.

1.4 The projective plane

Let $X = \mathbb{R}^3\backslash\{(0,0,0)\}$, and write $\mathbf{u} \sim \mathbf{v}$ to mean that $\mathbf{u}, \mathbf{v} \in X$ are scalar multiples. Then $X/\sim$ is the real projective plane, denoted $\mathbb{RP}^2$ or (in this course) P.

Any point of $\mathbb{RP}^2$ is represented by a unit vector $\mathbf{v} \in S^2$, but $\pm\mathbf{v}$ define the same point. So $\mathbb{RP}^2$ is homeomorphic to '$S^2/\mathbb{Z}_2$'. A point of the latter can be represented by (x, y, z) with $z \leqslant 0$, or by a point (x, y) with $x^2+y^2 \leqslant 1$. That point is unique unless $x^2+y^2 = 1$ in which case $(x, y) \sim (-x, -y)$, furnishing the disk model (left), equivalent to the square model (right):



We shall return to this model in §6.1.2.

2 Mathematical knots

2.1 Knots and their diagrams

Consider the circle

$$S^1 = \{(\cos t, \sin t) : t \in [0, 2\pi)\} \subset \mathbb{R}^2.$$

This set is a special case of the n-dimensional sphere S^n, consisting of all points of unit distance from the origin in $\mathbb{R}^{n+1}$. One can also regard S^1 as the set

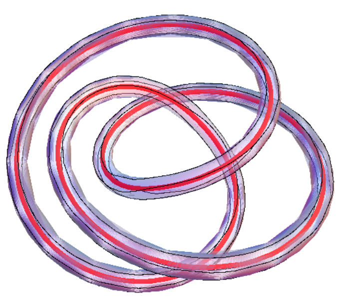
$$\{e^{it} : t \in [0, 2\pi)\}$$

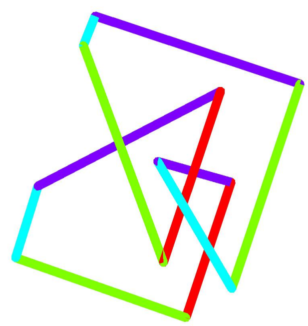
of complex numbers of unit modulus.

Definition. A **knot** is the image of a **continuous injective** mapping $f\colon S^1 \to \mathbb{R}^3$.

In practice, we shall assume that f is continuously differentiable and that $f'(t)$ is never zero. This ensures that the knot is smooth and can be surrounded by a tube of some fixed radius (below left).

A knot K is **tame** if there exists a homeomorphism of $\mathbb{R}^3$ that maps K onto the image of such a smooth knot, or (equivalently) one that maps K onto a piecewise linear knot formed by a finite union of line segments (right):

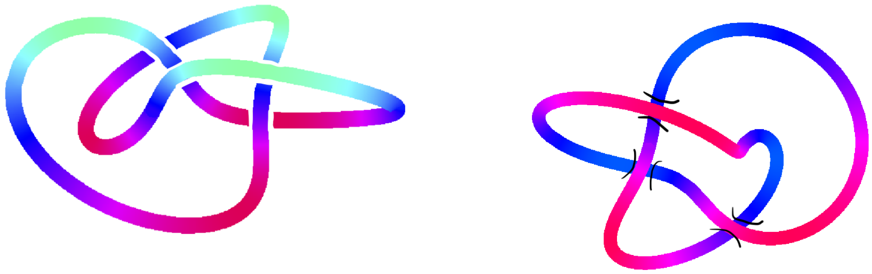




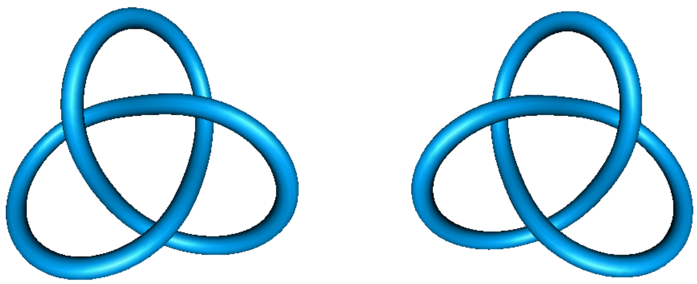
Restricting to tame knots avoids pathologies such as strings of knots getting smaller and smaller to some limits.

Given a knot K (a continuous curve without self intersections) in $\mathbb{R}^3$, we can project it onto any plane. Such a projection $\pi\colon K \to \mathbb{R}^2$ gives a valid diagram provided it is $1:1$ apart from a finite number $c \geqslant 0$ of points of $\mathbb{R}^2$ where it is $2:1$ and the branches are 'transversal'. These are the **crossings**, each of which has an **underpass** and **overpass**.

Both diagrams below are projections of a trefoil knot inscribed on a torus. In practice, the strands in a knot diagram are thickened for clarity, but in theory they are just single curves. The nature of each crossing can be emphasized by breaking the strands, or in some other way:



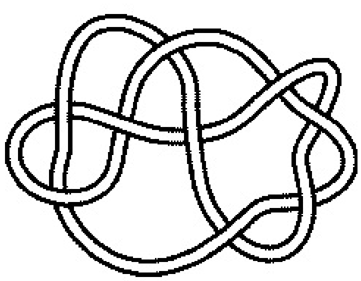
However, the most economical way to represent trefoil knots is by means of 3-leaved diagrams, here duplicated from p6:



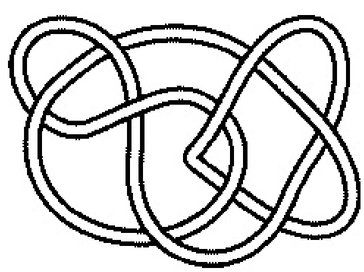
In fact, there are two trefoil knots, a left one and a right one, denoted $L3_1$ and $R3_1$. They are mirror images, but one cannot be deformed to the other in space. This chiral phenomenon is true for most knots, see §2.2.1 belows.

2.1.1 Arcs and regions

A knot diagram is divided into c **arcs** or strands, each of which starts and ends at an underpass. A **bridge** is an arc with at least one overpass. A diagram is **alternating** if overpasses and underpasses alternate in trasversing the diagram. The next diagrams appear in Rolfsen's knot tables, they each have $c = 9$ crossings (not 36!), and only the first is alternating:



9_{14}



9_{45}

The **shadow** of a knot diagram is a planar graph all of whose c vertices have degree 4. It must therefore have $e = 4c/2 = 2c$ edges and Euler's formula implies that there are

$$2 - c + e = c + 2$$

regions, including the outside.

2.1.2 Ambient isotopy

It is natural to regard two knots as being equivalent ('the same') if one can be manoeuvred into the other by hand, and regarding them as made of elastic rope so length does not matter. Technically, we are fixing S^1 and continuously varying the injective mapping $f\colon S^1 \to \mathbb{R}^3$.

However, the preferred mathematical notion of equivalence is given by the

Definition. Two knots K_0, K_1 are **ambient isotopic** if there exists a continuous map $F\colon \mathbb{R}^3 \times [0,1] \to \mathbb{R}^3$ such that $F_t = F(\bullet, t)$ is a homeomorphism for all t, F_0 is the identity and $F_1(K_0) = K_1$.

It is a theorem that both notions of equivalence coincide for tame knots.

Because F_0 and F_1 are connected by a path of homeomorphisms, F_1 will preserve the orientation of space. For example, F_1 cannot be a reflection. Typically, a knot and its reflection (in some plane) will not be ambient isotopic.

A knot diagram D represents a unique ambient isotopy class of knots, because we can construct a knot in space that projects to its diagram in a horizontal plane, using 'height' to faithfully render the crossings. Any two knots constructed in this way will be ambient isotopic.

But the same ambient isotopy class of knots can be represented *many different* diagrams. Even if we do not alter the knot in space, it can be projected to many different planes, obtained by applying a rotation and/or translation to any given plane.

2.2 Orientation and writhe

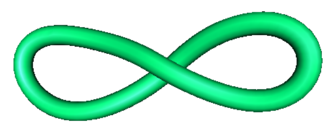
A knot defined by $f\colon S^1 \to \mathbb{R}^3$ is oriented by transferring a clockwise or anti-clockwise direction on S^1 to the image. Each knot diagram crossing x can then be assigned a positive or negative 'writhe' sign or integer, denoted $w(x) \in \{-1, +1\}$.

$$+1 \quad \begin{array}{c}\uparrow\\ \longrightarrow\\ \uparrow\end{array} \qquad\qquad -1 \quad \rightarrow\Big\uparrow\rightarrow$$

So for positivity, 'x axis wins' or (rotating 45^{o} clockwise) 'matrix diagonal wins'.

Definition. The **writhe** $w(D)$ of an oriented diagram D equals

$$\sum_x w(x),$$



where we sum over all c crossings x of D.

Example. The 'eight-twist' above right is unknotted, yet $w(D) = -1$. A change of orientation of a knot would reverse all the arrows above, so does not affect the writhe.

2.2.1 Mirror images

Given a knot K in $\mathbb{R}^3$, we can reflect it in any plane (mirror), which may ormay not intersect the knot. The resulting knot mK is well defined up to ambient isotopy (because the composition of any two reflections is an isometry with positive determinant and itelf defines an ambient isotopy of K).

Proposition. If D is a diagram for K then mK can be represented by a diagram D' with the same shadow, but in which all the crossings of D have been reversed, so $w(D') = -w(D)$.

To see this, choose Cartesian coordinates so that K lies in the slice $0 < z < 1$, and D is defined by projection to $z = 0$. If we reflect in this plane, the projection of mK to $z = -1$ will have the same shadow $\{(x, y) : (x, y, z) \in K\}$ but with crossings reversed.

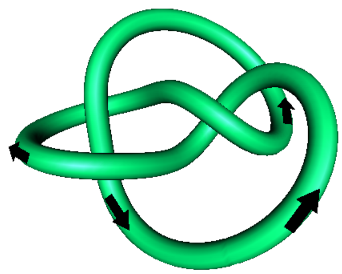
Definition. If K is ambient isotopic to mK then K is called **achiral** or **amphicheiral**, otherwise **chiral**.

Examples. (i) The trefoil knot is chiral, so there are really two distinct knots: $L3_1$ and $R3_1 = m(L3_1)$, though it is not easy to *prove* that these are not ambient isotopic.

(ii) The figure-eight knot is the simplest knot after the trefoil. It has an alternating diagram D with 4 crossings and

$$w(D) = -1 - 1 + 1 + 1 = 0,$$

with crossings counted from left to right:



The knot 4_1 is in fact achiral. Under certain conditions, zero writhe is a necessary condition for a diagram to represent an achiral knot, but the writhe is *not* invariant under ambient isotopy.

Achirality is rare: the next achiral knot in the tables is 6_3, though there are 5 (out of 21) in the list of (non-composite) knots whose diagrams have 8 crossings: $8_3, 8_9, 8_{12}, 8_{17}, 8_{18}$.

A separate question is whether there exists an ambient isotopy of a given knot K to itself that reverses orientation. In this case K is called **reversible** or **invertible**. Strictly speaking, a knot is a map from $S^1 \subset \mathbb{R}^2$, and so has a distingusihed orientation (direction of travel). But obviously we can reverse this. If K stands for a knot together with a give orientation, then rK denotes the same knot with the opposite orientation, and we are asking whether $K \approx rK$.

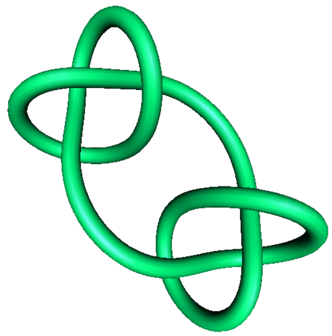
If we take a symmetric 3-leaved (say, left) trefoil diagram and look at it from the other side of the paper, the diagram willlook the same, but the orentation will be reversed. Thus, $L3_1 \approx r(L3_1)$, so $L3_1$ is reversible, as is $R3_1$. All knots with diagrams having less than 8 crossings are reversible, but 8_{13} is not. Given an oriented knot K, we can ask which (if any) of $K, mK, rK, rmK = mrK$ are ambient isotopic. None of them are for the pretzel knot $P(7, 5, 3)$ (see 3.4.3).

A knot K is **composite** if it is formed by 'sticking together' two other knots. This means that it has a diagram which can be divided into two halves of the plane with a separating line that cuts only two strands. If these strands are joined on either side, the resulting diagram cannot represent the unknot. An exampleis shown on page p31. We can write

$$K = K_1 \# K_2 \quad \text{or } K_1 + K_2$$

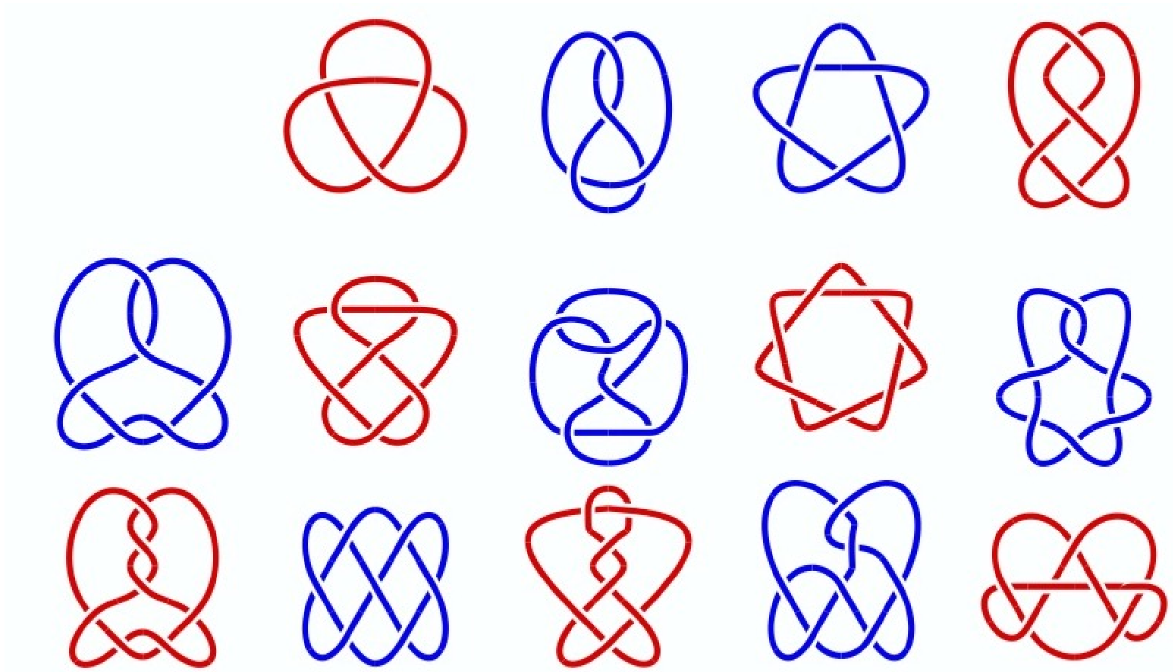
with each knot summand K_1, K_2 non-trivial. However, the ambient isotopy class of K will depend on choosing an orientation for each summand (if it is not the case that $K_1 \approx rK_1$ and $K_2 \approx rK_2$).

Example. The square or reef knot below is ambient isotopic to $R3_1\#L3_1 \approx L3_1\#R3_1$, and therefor achiral.

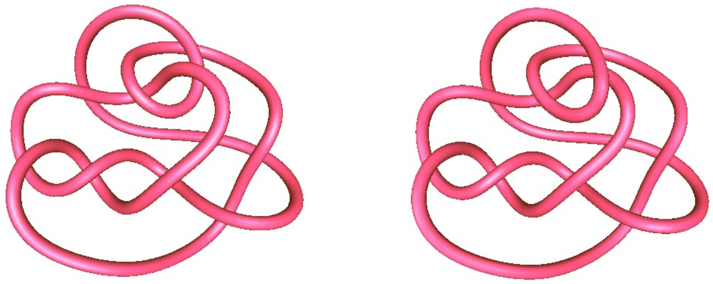


Bt the granny knot $L3_1\#L3_1$ at the bottom of p31 is not.

If K is not composite, it is called **prime**. Here are diagrams of prime knots with $\leqslant 7$ crossings:



Example. One of the two knots below is the **unknot**, i.e. isotopic to a circle, which has a diagram with no crossings (we should have inserted it top left in the table). The other is (ambient isotopic to) a trefoil knot.



2.2.2 Links

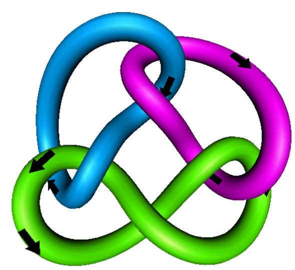
A link consists of a disjoint union $L = K_1 \sqcup \cdots \sqcup K_n$ of a finite number of knots in $\mathbb{R}^3$. A connected component of the link is a knot, so a knot is a link with one component. A link is oriented by assigning an orientation to each component.

Let $D = D_1 \sqcup D_2$ be a projection of (diagram for) L, where D_i is a diagram of K_i.

Definition. The **linking number** between K_i and K_j is

$$\ell_{ij} = \tfrac{1}{2}\sum_x w(K_i, K_j),$$

with the sum taken over all crossings x of D_i with D_j.

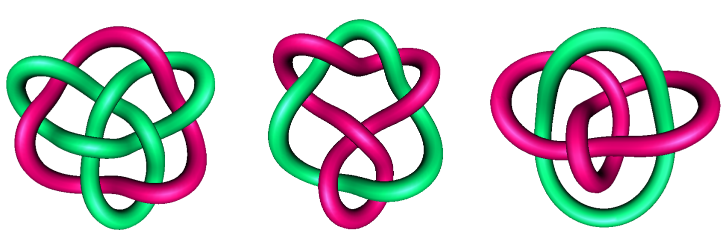


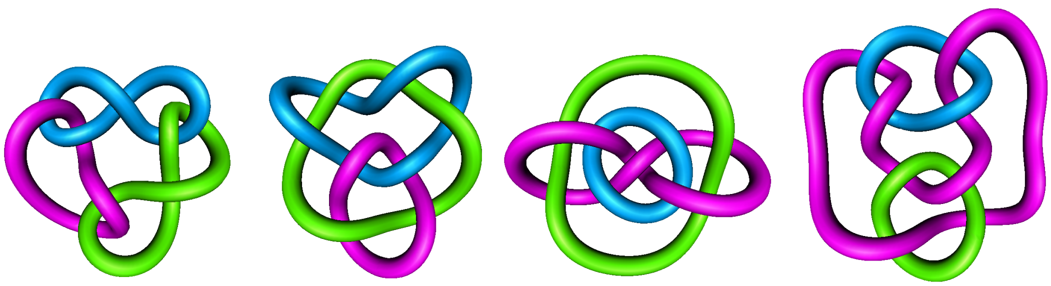
Example. Take $1 =$green, $2 =$blue, $3 =$mauve, on the right. With the given orientations,

$$\ell_{12} = -1, \quad \ell_{23} = \ell_{31} = 1.$$

We shall see that a non-zero linking number does detect components that cannot be separated. The writhe is more subtle and (for links) does depend on orientations.

Below are more links, most of whose components are unknots.

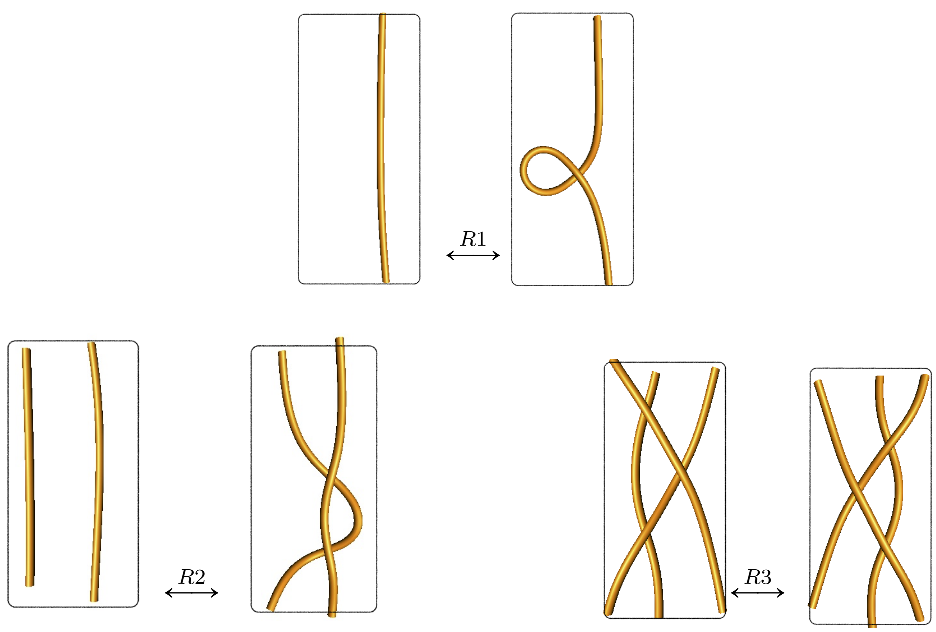




2.3 Reidemeister moves

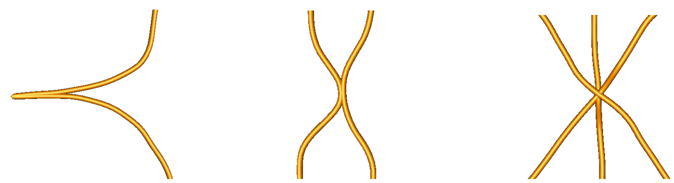
Such a move is performed on a portion of the diagram D of a knot K, by 'unplugging' the box on the left and replacing it by the one on the right, or vice versa. No other strands can be present in those regions.

Each move produces a diagram D' that is obviously the projection of some knot K' that is ambient isotopic to K. Instances of the three moves are sketched, but there are other valid versions with evident crossing changes.



2.3.1 Deformations and singularities

In carrying out the Reidemeister moves (R-moves for short), one is naturally allowed to deform the strands in part or all of the plane, provided the nature of the crossings is preserved. Technically, this corresponds to applying an orientation-preserving homeomorphism of the plane to the underlying diagram. Its use is always permitted, and (if it is needed call attention to such a deformation) it is indicated by R0.



Crossings in a knot diagram are based on the notion of an **ordinary double point**. The R-moves represent transitions that result from deformations of other types of singularities that are not allowed in knot diagrams: (1) cusps, (2) tangents, (3) triple points, shown above. Such configurations are modelled (up to R0 moves) on the equations

$$(1)\ x^3 = y^2, \quad (2)\ x^2 = y^4, \quad (3)\ x^3 = xy^2$$

2.3.2 Kurt Reidemeister's theorem

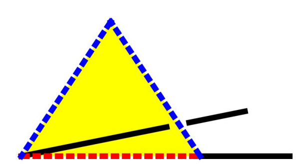
Suppose that a link diagram D' is obtained from D by carrying out a finite sequence of R-moves. Then we say that D, D' are **isotopic diagrams** and write $D \sim D'$.

Assume that two knots or links K, K' project to respective diagrams D, D'. It is clear that $D \sim D'$ implies that K and K' are ambient isotopic (interpret one R-move at a time). It will be of enormous theoretical importance to know that the converse is true:

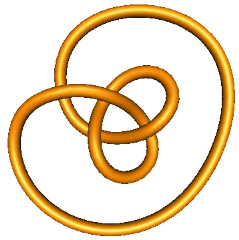
Theorem [Reidemeister, 1932]. If K and K' are ambient isotopic then $D \sim D'$.

One can understand why this is true by considering what happens to a diagram as one carries out an ambient isotopy in space. This could be a mere rotation (which is equivalent to moving the projection plane). In addition to R0 moves, singularities of type (1), (2), (3) may develop, but these are 'resolved' by the respective R-moves.

Reidemeister's proof (that we omit) consisted in studying 'triangle moves' on piecewise linear knots.



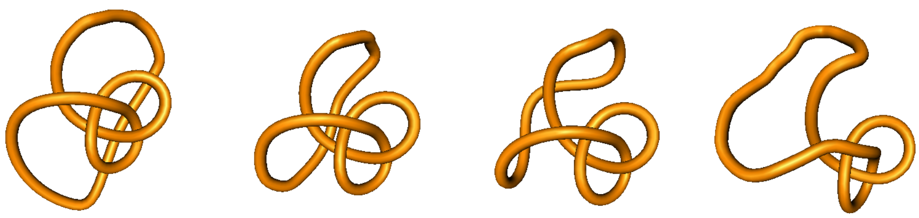
Example. The knot above is obviously a trefoil because we can pass the big loop across underneath (or over) from right to left and then untwist it.



$w(D) = +4$

Reidemeister's theorem assures us that this 'macro move' can be achieved by a sequence of R-moves.

In fact, it can be achieved by moves R2,R2,R3 (to arrive at the first diagram below), then R3,R3,R2 (to arrive at the second) then R3 (the third), then R1,R1, and finally (not shown) R1, not to mention R0 at every stage:



2.3.3 Effect of R-moves on the writhe

It is obvious that applying R1 to twist a single arc introduces a crossing and changes the writhe of the whole knot diagram by ± 1. (The writhe inside the previous right-hand R1 box is -1.) On the other hand,

Lemma. Moves R2 and R3 do not change the writhe of a diagram.

For R2, this is because (however the strands are oriented) the two crossings in the right-hand box have opposite writhe signs. In the case of R3, the central crossing is unchanged. On the left we have NW and SW crossings, and on the right, NE and SE crossings. Using these compass points also to represent their signs, we have NW=SE and SW=NE. So the writhe inside each box is the same.

Later on, we shall need the stronger equivalence relation of **regular isotopy** between diagrams D, D'. This means that D' can be obtained from D by a sequence of moves of type R0, R2, R3, and is indicated by $D \approx D'$.

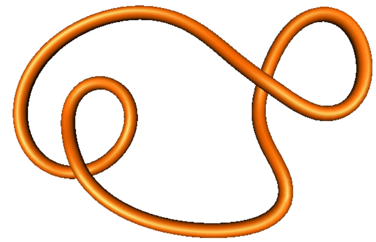
Corollary. If $D \approx D'$ then $w(D) = w(D')$.

2.3.4 Example: cancelling twists

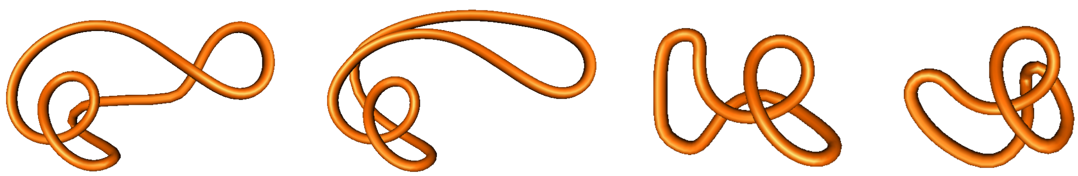
Here is a diagram D with two twists and zero writhe that is **regularly isotopic** to the diagram of the unknot with no crossings:

$$D \approx O.$$

The possibility of eliminating twists without using R1 is related to a procedure in higher-dimensional topology called the 'Whitney trick'.



The regular isotopy from D to O is accomplished by means of moves R2,R0,R0,R3, and (not shown) R2,R2:



2.3.5 Back to linking number

Let $L = K_1 \sqcup K_2$ be an oriented link with 2 components (so each is an oriented knot). Let D be a projection of (i.e. diagram for) L. We defined the linking number between the two knots as

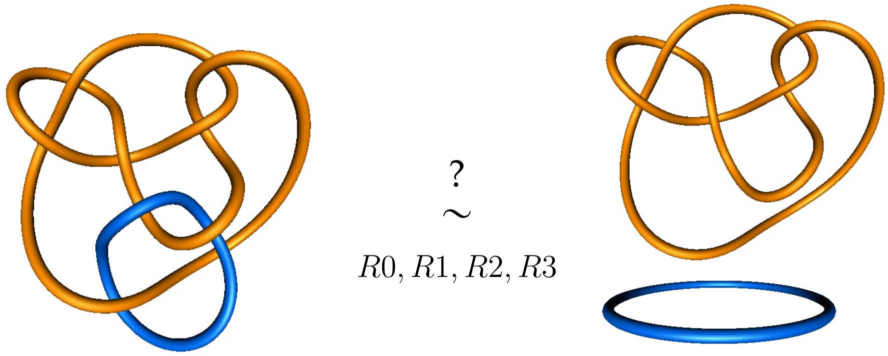
$$\ell = \ell_{12} = \tfrac{1}{2}\sum_x w(K_1, K_2),$$

where the sum is taken over all crossings x of a strand of K_1 with one of K_2.

The quantity ℓ will not change when we apply an isotopy to D, because R2 and R3 preserve the writhe, and ℓ is unaffected by R1 moves (since these twists only add or subtract crossings on the *same* component). The sign of ℓ will change if we change the orientation of exactly one of K_1, K_2.

Corollary. The absolute value $|\ell|$ is invariant by isotopy, and is therefore an ambient isotropy invariant of L. If $\ell \neq 0$ then the two knots cannot be separated in space.

Example. To illustrate the corollary, consider the two links below:



If the left-hand link is ambient isotopic to the separated knots **in space** then (by Reidemeister's theorem) the two link **diagrams** are isotopic, meaning that they differ by a sequence of R-moves. But this would imply that $\ell = \frac{1}{2}(+1+1+1+1) = 2$ (on the left) equals 0 (because there are no crossings between K_1, K_2 on the right), contradiction.

The converse to the last Corollary is false. The two components of the Whitehead link have $\ell = 0$, but they cannot be separated. This can be proved using the concept of colouring.

2.4 Numerical invariants

Definitions. Let K be a knot, or a link with more than one component.

The **crossing number** $\mathsf{cr}(K)$ is the least number of crossings needed in *any* diagram D of K. It is the basis of traditional knot tables. If $\mathsf{cr}(K) < 3$ then $\mathsf{cr}(K) = 0$ and K is trivial, i.e. ambient isotopic to an unknot like the circle.

A **minimal** diagram is one (as in the knot tables) with exactly $\mathsf{cr}(K)$ crossings.

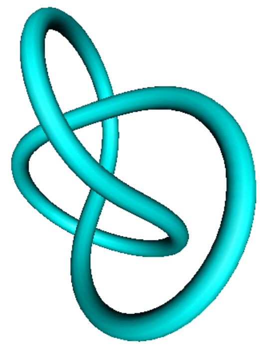
The **unknotting number** $\mathsf{u}(K)$ is the least number of crossing-reversals in *any* diagram D of K needed to convert D to the projection of an unknot. If K is a knot with $\mathsf{cr}(K) \leqslant 7$ then $u(K) \leqslant 2$ except that $u(7_1) = 3$. The fact that $u(8_{10}) = 2$ was only verified in 2005, and u of some knots with $\mathsf{cr} = 10$ is still unknown.

The **bridge number** $\mathsf{br}(K)$ is the least number of bridges in any diagram of K.
If K is a knot with $0 < \mathsf{cr}(K) \leqslant 7$ then $\mathsf{br}(K) = 2$, but $\mathsf{br}(8_{10}) = 3$.

Warning. Many minimal diagrams have more than $\mathsf{br}(K)$ bridges (for example, if they are alternating). Moreover, $u(K)$ is not necessarily realized by a *minimal* diagram of K, which makes it hard to determine the unknotting number.

2.4.1 Some known results

An alternating diagram of a trefoil knot 3_1 has 3 bridges, but a diagram with 2 is easily obtained by modifying 2 crossings of the figure-eight knot:



Lemma. If D is a diagram of a knot with c crossings then we need reverse at most $c/2$ to obtain the unknot, so $u(K) \leqslant \lfloor c/2 \rfloor$. (See sheet 2.)

Lemma. If $\mathsf{br}(K) = 1$ then K is trivial. (This can be shown using the so-called DT code of a knot, later.)

Recall the notion of the composite $K_1 \# K_2$ of two knots, and the images on pages p20, p31.

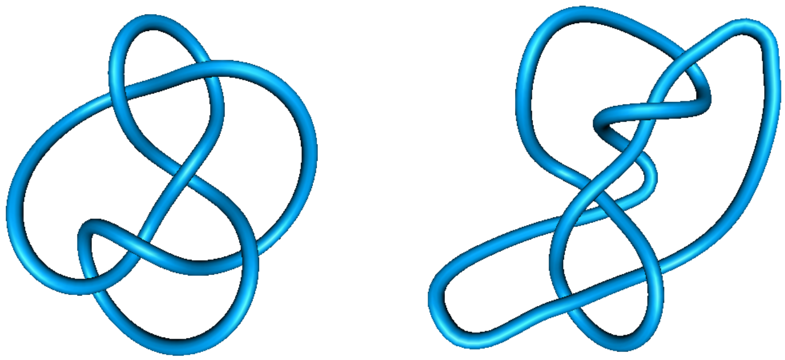
Theorem [Scharlemann, 1985]. If $u(K) = 1$ then K is prime. Equivalently, the composite $K_1 \# K_2$ of two non-trivial knots must have $u \geqslant 2$.

Theorem [Schubert, 1954]. $\mathsf{br}(K_1 \# K_2) = \mathsf{br}(K_1) + \mathsf{br}(K_2) - 1$.

2.4.2 Maxima and minima

Treat a portion of the image of a knot projection $f\colon [0, 2\pi) \to S^1 \to \mathbb{R}^3 \to \mathbb{R}^2$ with $f(t) = (x(t), y(t))$ as a graph. It has a turning point where $dy/dx = 0$ (or $dy/dt = 0$).

Example. A standard diagram of 6_3 has 3 local maxima and minima, but is isotopic to one with 2 of each, all absolute:

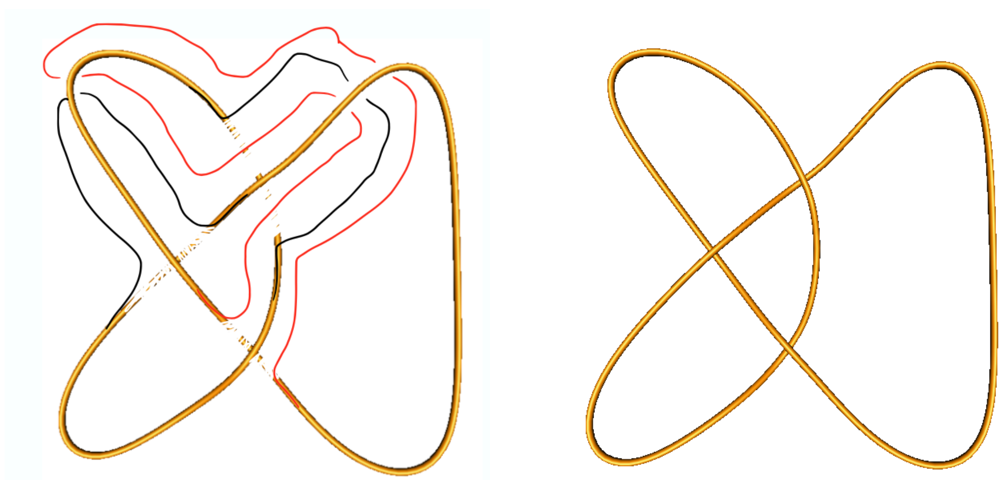


By passing crossings upwards, one can then eliminate all those below the highest one, by an algorithm in which underpasses are successively moved upwards, converting the top hoops into bridges. The new diagram becomes complicated, but one deduces the

Theorem. $\mathsf{br}(K) = n$ iff K has a diagram with n (and no fewer) local maxima.

There is in fact an algorithm for 'lifting bridges'.

To carry out the process for the diagram of $L3_1$ bottom left, one needs a total of 5 crossings to secure the 2 bridges. The black strands are mere deformations of the golden ones, but crossings involving red strands also require moves R2,R3. On the other hand, one needs at least 9 crossings to secure 2 bridges for $R3_1$ bottom right!



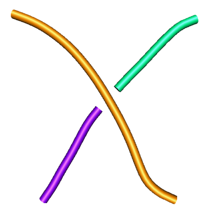
3 Colouring

Let L be a knot or a link with more than one component, and let $n \geqslant 2$ be a positive integer.

3.1 The concept

Definition. L is n-colourable if it possesses a diagram whose arcs can be labelled with **at least two distinct integers** from $\{0, 1, \ldots, n-1\}$ such that at each crossing

$$\text{underpass label} + \text{underpass label} = 2(\text{overpass label}) \mod n.$$



No knot is 2-colourable because both underpassing arcs would necessarily have equal labels (either 0 or 1) at each crossing, so (by tracing around the knot) only one label would ever be seen. But any link with at least two components is 2-colourable (just label all arcs of the one component 0, and everything else 1).

It is easy to see that a link is 3-colourable or **tri-colourable** if at each crossing either all three colours (meaning $0, 1, 2$) are distinct (as in the image above) or they are all the same.

3.1.1 Invariance by R-moves

By checking the effect of Reidemeister moves on coloured crossings, one deduces the

Theorem. If D is n-colourable and $D \sim D'$ then D' is n-colourable.

We can therefore say that a knot or link in space is n-colourable if any one of its diagrams is. In other words, this concept is an **invariant** of ambient isotopy, even though colouring a knot in space makes no sense, at least without our perception of it!

Proof of the theorem. The colouring equation is set up to be invariant by the R-moves, which we consider one at a time. Bear in mind that in applying an R-move from left to right or vice versa, the colours of the strands entering the boxes are fixed because they must agree with those of the rest of the link diagram outside the boxes.

The coefficient '2' deals with R1: a twist will automatically satisfy the colouring equation because all colours (i.e. integers modulo n) are equal.

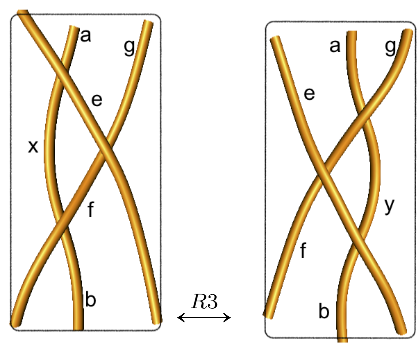
Move R2 (from left to right) adds an arc inside the box, whose colour x can be chosen freely to satisfy the one equation $a + x = b$.

Move R3 is the only tricky case. For the setup in the left box below, take a crossing with overpass $\searrow$ coloured e and underpasses f, g, and arcs $\downarrow$ underneath a, x, b.

Studyiong the effect of R3 on colouring is a bit more involved. The colouring equations become

$$\begin{array}{lll} f+g=2e, & a+x=2e, & x+b=2f, \\ f+g=2e, & a+y=2g, & y+b=2e. \end{array}$$

If the left-hand ones (first row) are satisfied, we must prove that we can choose y to satisfy the right-hand ones.



So define $y = 2g - a$, and note that $y + b = 2g - a + b = 2g + 2f - 2e = 2e$, exactly as required. Other versions of R3 can be proved in a similar manner.

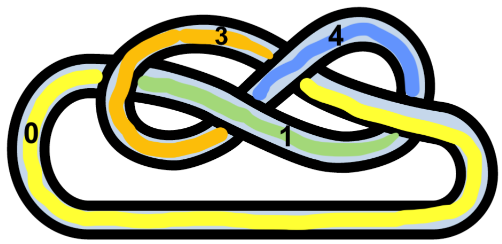
3.1.2 A 5-colouring of 4_1

First, we record the

Lemma. If a knot or link is n-colourable, it is also m-colourable if n divides m.

For if $m = nk$, a colouring equation $n|(a + b - 2d)$ implies that $m|(ka + kb - 2kd)$, which is the equation in which all the colours have been multiplied by k.

Another observation is that (for fixed n), we can add any constant integer to all the labels, and the colouring equations are still satisfied. This means that we are free to choose any arc and colour it '0'. Here we have coloured the figure-eight knot with $n = 5$; since there are 4 arcs, we cannot use all 5 colours.



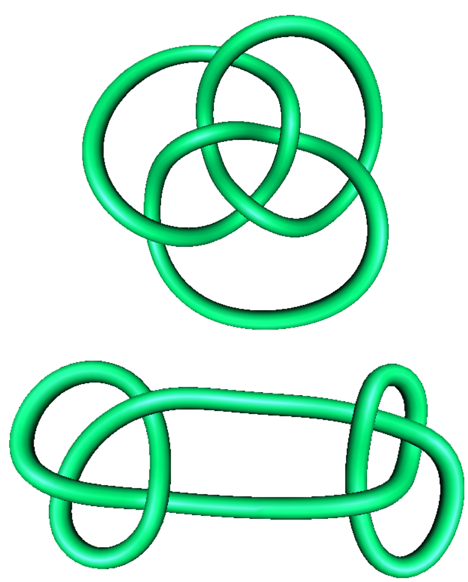
The trefoil knot 3_1 is 3-colourable. It is also easy to check that 3_1 cannot be 5-coloured. Similarly, the figure-eight knot 4_1 cannot be 3-coloured.

Corollary [if it was not already obvious]. 3_1 is not ambient isotropic to 4_1.

3.1.3 Exercises and applications

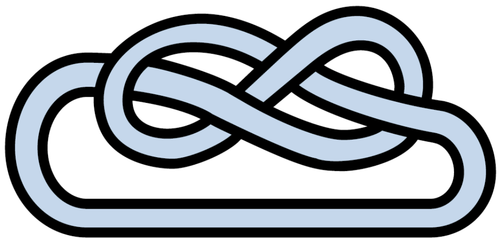
The unknot is not n-colourable for any n.
K is n-colourable iff its mirror image mK is.
The trefoil knot 3_1 is n-colourable iff $3|n$.
The figure-eight knot 4_1 is n-colourable iff $5|n$.
The knot 7_7 is n-colourable iff $3|n$ or $7|n$.
The Borromean rings are n-colourable iff $2|n$.
The composite 'granny knot' $L3_1\#L3_1$ (bottom) is 3-colourable.

We can now distinguish ambient isotopy classes of many such knots and links!



3.2 Colouring matrices

Label the crossings in the figure-eight diagram 1,2,3,4 from left to right.



Label its arcs by their overcrossings, giving the **colouring matrix** A_+:

$$A_+ = \begin{pmatrix} 2 & -1 & 0 & -1 \\ -1 & 2 & -1 & 0 \\ -1 & 0 & 2 & -1 \\ 0 & -1 & -1 & 2 \end{pmatrix},$$

whose rows represents crossings and columns arcs. If the arcs are coloured x_1, x_2, x_3, x_4 modulo n then we need

$$A_+ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{pmatrix} n,$$

for some integers x_i, b_j. Since $A_+(1,1,1,1)^\top = \mathbf{0}$, we can take $x_4 = 0$, but the other x_i cannot all be 0. Notice that the column vectors of A and (because the diagram is alternating) the row vectors of A add up to **0**.

Cofactors. Since the fourth row/equation is a linear combination of the others, it suffices to consider the top 3×3 block A of A_+ and solve

$$\begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ -1 & 0 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = n \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}, \quad \text{or} \quad A\mathbf{x} = n\mathbf{b}.$$

Now

$$\mathbf{x} = nA^{-1}\mathbf{b} = \frac{n}{\det A}\widetilde{A}\mathbf{b} = \frac{n}{5}\begin{pmatrix} 4 & 2 & 1 \\ 3 & 4 & 2 \\ 2 & 1 & 3 \end{pmatrix}\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}.$$

We now see that if $n = 5$ possible colourings are generated by the columns of $\widetilde{A}$, the transpose of the matrix of cofactors of A. In fact, $\widetilde{A}$ has rank 1 modulo 5, so there 4 possible colourings with $x_4 = 0$, generated by taking $\mathbf{b} = (1, 0, 0)^\top$, namely

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 4 \\ 3 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 \\ 1 \\ 4 \end{pmatrix}, \begin{pmatrix} 2 \\ 4 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

3.2.1 The determinant

First, suppose that L is an alternating link, that is, one that possesses some diagram D that is alternating. (Recall that the term 'link' includes 'knot'.)

Define the colouring matrix A_+ of D as before, by assigning the overpassing arc to each crossing and setting 2 down the diagonal position. Beware though that if a given arc passes the same crossing more than once, the numbers $2, -1, -1$ might become $1, -1$ or $2, -2$. Each arc contains exactly one overpass. So not only do the entries of each row (crossing) sum to zero, but also the entries of each column (arc) sum to 0. In particular, the row vectors (or crossing equations) sum to zero.

Proposition. Let M be a $c \times c$ matrix whose row and column vectors both sum to zero. Then all the first minors (determinants of submatrices of size $(c-1) \times (c-1)$) are equal up to sign.

We shall see that we can always define a colouring matrix so as to use the Proposition for any link L, alternating or not. The **determinant** of L will then be defined to equal the absolute value of any first minor A (equivalently, cofactor) of A_+.

Proof of the previous proposition using minors. Let M' be the matrix obtained from M by adding 1 to every entry of M. Fix (i, j) and let M'' be the matrix obtained from M' by adding to row i of M' to all its other rows, and then adding to the new column j all its other new columns. The entries in row i or column j of M'' are then n except that $(M'')_{ij} = n^2$. Let M''' be the matrix obtained from M'' by subtracting $1/n$ times row i from all the other rows.

Let $M_{i,j}$ be the submatrix of M formed by deleting row i and column j. Row i of M'' equals $(n \cdots n^2 \cdots n)$, so M''' coincides with M everywhere except in row i, and column j (where all entries are 0 except for the n^2). So if we compute its determinant by expanding along column j we obtain

$$\det(M''') = (-1)^{i+j} n^2 \det(M_{i,j}).$$

The rules for determinants tell us that $\det(M''') = \det(M')$ and the right-hand side is independent of (i, j). $\square$

3.2.2 A list of determinants

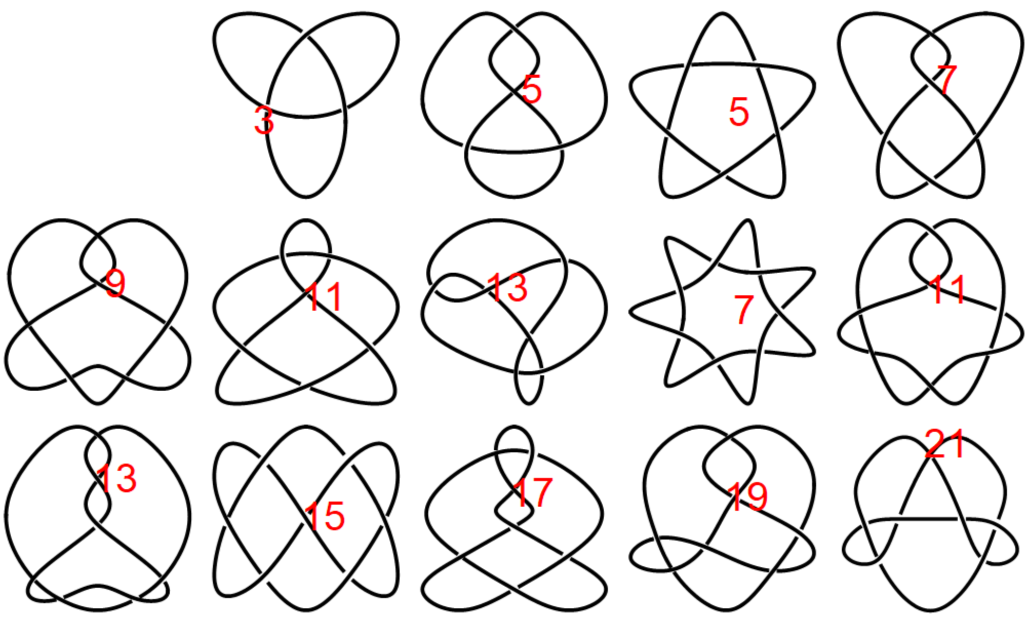
After introducing some technicalities, we shall show that the determinant of a general link is well defined, and see that L **is** n**-colourable iff** n **and** $\det A$ **have a non-trivial common factor.**

It follows that:

(0) If $\det A = 0$ then L is colourable for any n (impossible for a knot, whose determinant must be odd, possible for a link).

(1) If $|\det A| = 1$ then L is colourable for no n (possible even for a non-trivial knot).

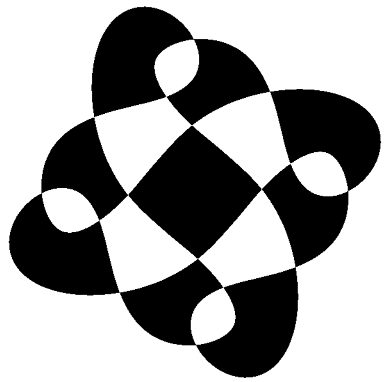
Here are some knot determinants:



3.3 Geometry of diagrams

Let D be diagram of a link L with c crossings. We shall assume that the diagram is **connected** and it contains **no closed arc**. Then its shadow is a planar graph with c vertices and $c+2$ regions, including outside (see p16).

We next state without proof the fairly obvious fact that the shadow of a knot diagram can be 'chess-boarded':



Chess lemma. Black and white can be assigned to the regions of D in such a way that the same colours only meet at a vertex and not along an edge.

We shall use this result to show that:

(i) a chess-boarded shadow distinguishes an alternating diagram,

(ii) a colouring matrix can be defined so that the row vectors sum to zero, and

(iii) to implement a quick way of computing the determinant.

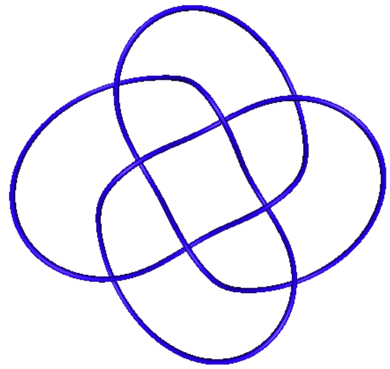
3.3.1 Alternating knots

Let S be the shadow of a knot diagram with a chess-boarding. Orient S and start anywhere. At the first crossing one reaches, make the strand an overpass if black is on the right as you arrive, otherwise make it an underpass. Since black and white alternate, this rule will determine an alternating diagram with the property that (if, as usual, we regard an overpass as the x-axis) quadrants Q1 and Q3 are shaded black.

Corollary. The shadow of any knot diagram has two alternating diagrams associated to it, corresponding (up to ambient isotopy) to a knot K and its mirror image mK.

It is convenient, whenever possible, to chess-board a shadow so that white is on the outside. This then distinguishes one of K and mK. For example, making the three leaves of a trefoil shadow black determines $R3_1$.

Any knot up to and excluding 8_{19} can be represented by an alternating diagram. Such a knot is therefore determined by its shadow up to reflection. The absence of over and underpasses in a knot diagram assumes it is alternating. If 8_{19}'s shadow is rendered alternating it becomes a diagram of 8_{18}:

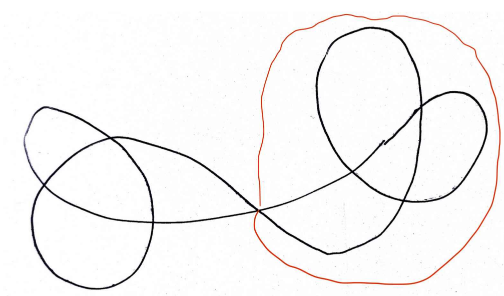


3.3.2 Reduced diagrams

For future reference, it will help to restrict to diagrams that have no twists that can easily be eliminated (either by an R1 move on a single arc, or by a 'macro move'):

Definition. The diagram of a link is **reduced** if no crossing is an **isthmus**, one that is bounded by 3 (rather than 4) regions.

An isthmus arises from a 'macro' twist contained in some region, and defines a closed curve $f(S^1) \subset \mathbb{R}^2$, shown red below left. On the right is the isthmus of Sestri Levante, surrounded by the blue Mediterrean sea.



By Schoenflies' theorem (a strengthening of Jordan's), $f(S^1)$ is unknotted and there is a homeomorphism $F\colon \mathbb{R}^2 \to \mathbb{R}^2$ such that the image of $F \circ f$ is a circle, the boundary of a disk that can be rotated in space to eliminate the twist. This is the idea behind the proof of the

Theorem. Any link is ambient isotopic to one with a reduced diagram.

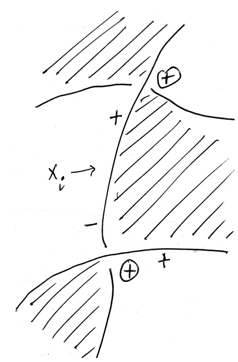
3.4 The Goeritz method

3.4.1 Back to the colouring matrix

Let D be a reduced diagram of a knot or link L. In addition, suppose that D is **connected** and contains **no closed arc**. The colouring equations are then encoded in a matrix A_+ of size $c \times c$, as in §3.2.

Given a chess-boarding of D, assign a **chess sign** (also called Goeritz index) $+1$ to a crossing if (with x axis as the overpass) quadrants Q1, Q3 are black, otherwise -1. For $+1$ use the entries $2, -1, -1$ in the row of the colouring matrix, otherwise $-2, 1, 1$.

Lemma. In this setup, the integers in each column of A_+ (defined by an arc) sum to 0.

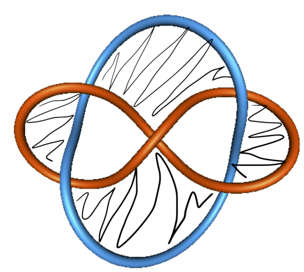


Proof. Consider 'subarcs' of D, equivalently edges x_i of the shadow S. Each colouring equation involves exactly four of them:

$$x_1 - x_2 + x_3 - x_4 = 0,$$

where x_i has a positive sign iff it approaches the crossing with black on its right. The overpass will consist of either $\{x_1, x_3\}$ or $\{x_2, x_4\}$. Each subarc will be assigned opposite signs from its two crossings, and all the crossing equations add up to 0. □

3.4.2 Computing the determinant



There is an easier way to compute the determinant of a knot. Take the chess-boarding of the Whitehead link shown right. Label the white regions inner left 1, right 2, outside 3. Define a 3×3 matrix $G_+ = (g_{ij})$ (named after Lebrecht Goeritz, who obtained a PhD under Reidemeister in 1933) by

$$g_{ij} = \begin{cases} \text{sum of any crossings* common to regions } i, j, & \text{if } i \neq j, \\ \text{minus the sum of all crossings* bounding region } i, & \text{if } i = j. \end{cases}$$

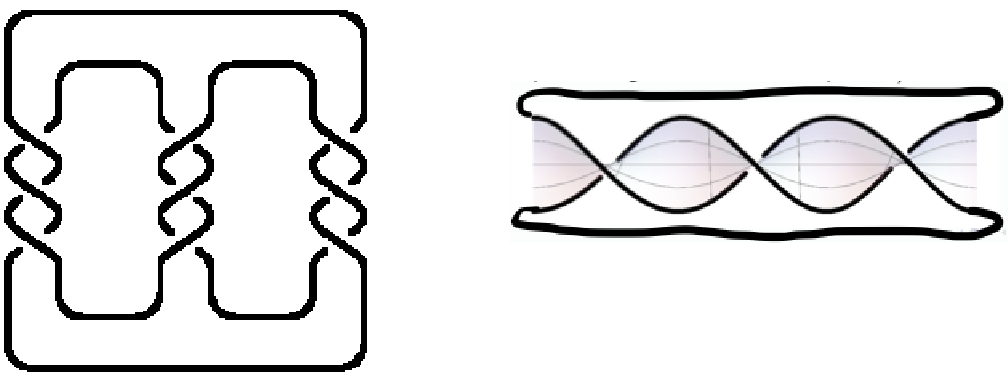
*In taking the sum, the crossing should be assigned its chess sign (all positive here), so G_+ is symmetric and its rows and columns add up to zero. We then compute the determinant using the previous proposition. Deleting last row and column yields

$$G = \begin{pmatrix} -3 & 1 \\ 1 & -3 \end{pmatrix}, \qquad \det G = 8.$$

It can be shown that $|\det G|$ agrees with the link determinant computed from the colouring matrix A. We shall not justify this but recommend its use in practice.

3.4.3 Pretzel links

If p, q, r are integers, the (triple) **pretzel link** $P(p, q, r)$ is formed by taking three twists with p, q, r crossings, and combining them in the form of a flat 'pretzel'. This provides a diagram D in which the twists are shown from left to right. If D is chess-boarded with white on the outside then a positive/negative integer means that all the crossings in that twist have positive/negative chess signs. With this convention, the one shown top right is $P(-3, 3, -3)$ (which happens to coincide with 9_{46}), whereas the trefoil knot $R3_1$ on th right is $P(-1, -1, -1)$:



In space, one can think of the three twists more symmetrically as dangling vertically from points of a circular light fitting. That shows that a cyclic permutation of the three integers leads to ambient isotopic links. If p is even/odd then the two strands enter and leave the twist in the same/reversed order on D. It follows that $P(p, q, r)$ is a knot if at least two of p, q, r are odd, otherwise it is a link with 2 or 3 components.

Proposition. The determinant of the link $P(p, q, r)$ equals $|qr + rp + pq|$.

Justification with the Goeritz method. Chess-boarding a diagram like that of $P(-3, 3, -3)$ shown leaves two interior white regions, one bounded by $|p| + |q|$ crossings and the other by $|q| + |r|$. The resulting Goeritz matrix

$$G = \begin{pmatrix} -p - q & q \\ q & -q - r \end{pmatrix}$$

takes account of the chess signs. Its determinant is $pq + pr + qr$. □

A a corollary, $P(-2, 3, 5)$ and $P(-3, 5, 7)$ (for example) have determinant one, and (like the unknot) cannot be coloured for any modulus n.

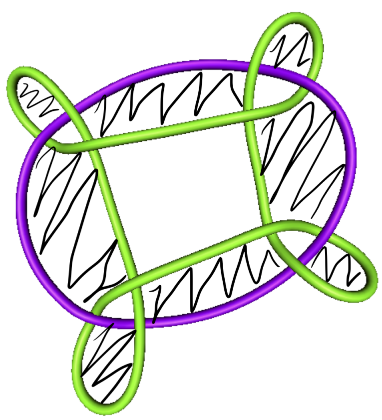
Unless p, q, r all have the same sign, the chess signs will also vary in sign and the usual diagram of $P(p, q, r)$ cannot be alternating. The first non-alternating knot in the tables is 8_{19}, which happens to coincide with $P(3, 3, -2)$ and is a knot that can, in common with $3_1, 5_1, 7_1$, be inscribed on a torus, see the French logo and p42 below!

FORUM DE
L'ENTREPRENEURIAT

3.4.4 More examples

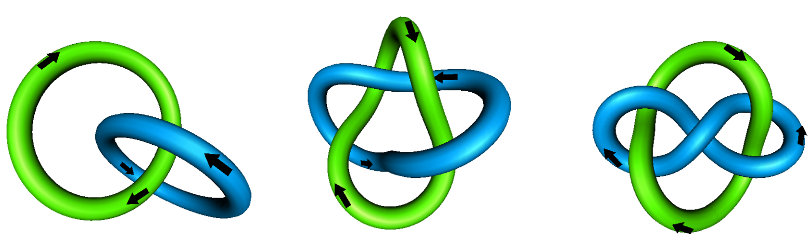
Recall that no knot can be 2-coloured (because both underpasses at a crossing must be 0 or 1, so there can only be one colour throughout the component). So a knot always has odd determinant.

Any link with at least 2 components can be 2-coloured (one knot 0, the rest 1). If a link possesses a diagram with 2 components that can be separated then its determinant is zero, but the converse is false: the link below has $\boldsymbol{\Delta} = 0$:



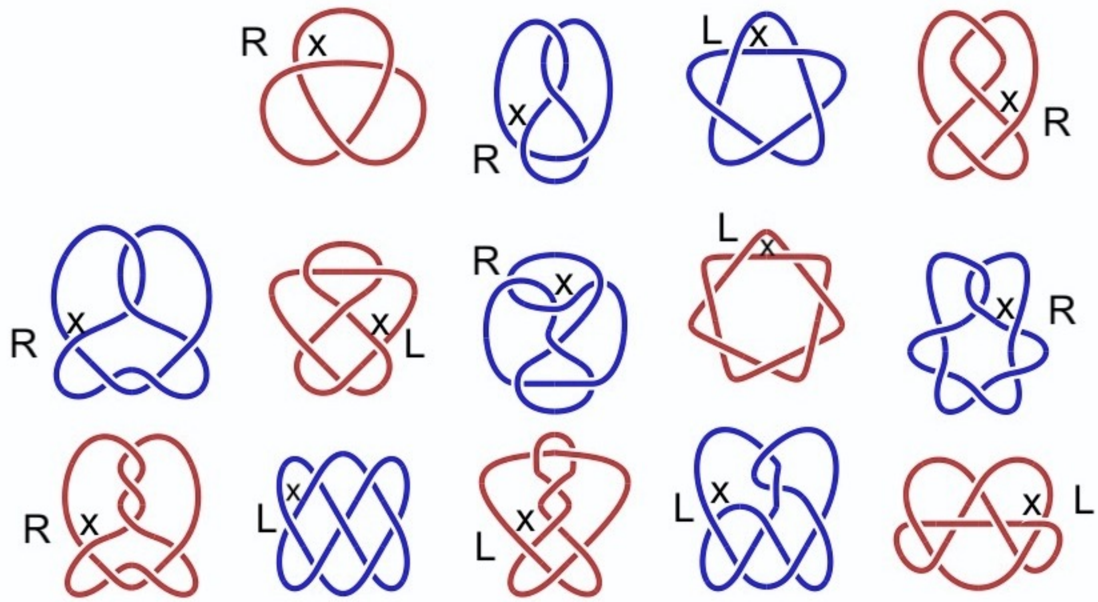
$$G = \begin{pmatrix} -3 & 0 & 1 & 0 & 0 \\ 0 & 3 & -1 & 0 & 0 \\ 1 & -1 & 0 & -1 & 1 \\ 0 & 0 & -1 & 3 & 0 \\ 0 & 0 & 1 & 0 & -3 \end{pmatrix}$$

The determinants for the 2-component links on sh1 are $2, 4, 8$ respectively (ignore the arrows):



3.4.5 Left and right

If a knot diagram is alternating, wwe could call it 'rght-handed' if all its chess sgns are positive when it is chess-boarded with white outside. This then gives us a scheme whereby we can distinguish between left and right version of the knots in the table on p20, which is consistent with the nomenclature for the trefoil knots $L3_1$, $R3_1$. (An 'x' marks a crossing where we have checked the chess sign, but they must all be the same.)



For the two achiral entries, the two versions are necessarily ambient isotopic: $L4_1 \approx R4_1$ and $L6_3 \approx R6_3$.

Exercise. Use the diagram for 6_3 to visualize this knot in $\mathbb{R}^3$. Find a linear transformation of $\mathbb{R}^3$ (with suitable origin) that establish the equivalence $L3_1 \approx R3_1$.

3.5 Knot DT codes

Let D be an oriented diagram with c crossings. Choose a starting point so that the first crossing encountered is an underpass, and chess-board the diagram so that white is on the right at the start. Number the crossings

$$1, 2, 3, \ldots, c, c+1, c+2, \ldots, 2c$$

as they are encountered in succession around the circuit. If a crossing is labelled with an odd number the first time it is encountered (like 1), it will have an even label the next time, and vice versa. This is an application of the existence of a chess-boarding. Thus, the set of crossings determines a bijective mapping $f\colon \{1,3,5,7,\ldots\} \to \{2,4,6,8,\ldots\}$, which we abbreviate by the ordered list

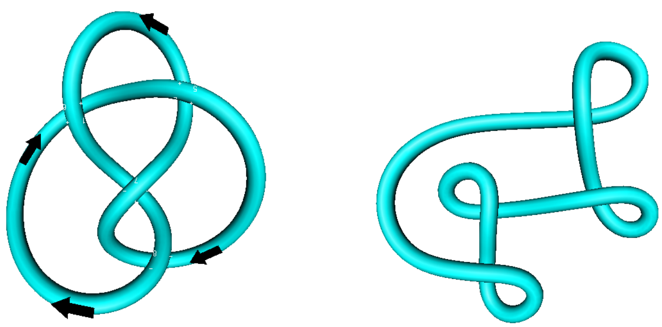
$$f(1)\ f(3)\ f(5)\ f(7)\ \ldots$$

If D is alternating, the overpasses will all be labelled by the even integers $2,4,\ldots,2c$. In general, one adds a minus sign to the even integer if it labels an **underpass** (for example -2 if one starts 'under, under, …').

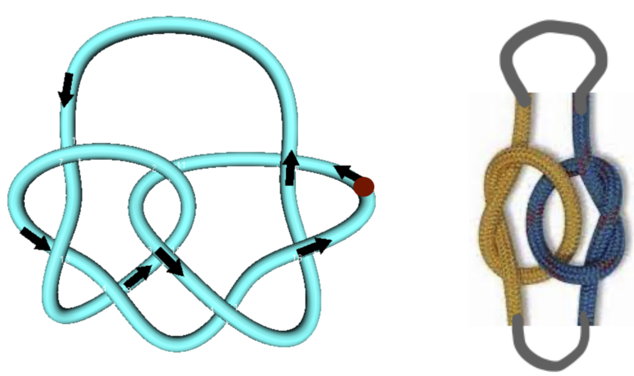
The resulting list of even integers is called the **Dowker-Thistlethwaite code** or **DT** code of the knot. Of the $2^c c!$ possibilities, many will give rise to an unknot and few will correspond to a knot K whose crossing number $\mathrm{cr}(K)$ (recall p26) equals c.

3.5.1 Examples: 4_1 and 8_{19}

With the given orientation below, it does not matter before which underpass we start: we always return to the initial crossing after 5 more subarcs and the DT-code is $6\ 8\ 2\ 4$. With the opposite orientation it is $4\ 6\ 8\ 2$. The code $2\ 4\ 6\ 8$ gives the unknot:



The knot 8_{19} is the first non-alternating one in the Rolfsen table. It is formed by linking together two open trefoil knots and then closing them, as shown.



(The second image courtesy of 'The Happy Mariner'.) It clear from either diagram that 8_{19} is the pretzel knot $P(3,-2,3)$. Thus, it has determinant 3 and is 3-colourable. It is sometimes called the 'true lovers knot' (TLK), but the diagram of its shadow on p35 reveals a 4-fold symmetry.

Starting from the red dot in the left-hand diagram, the DT code is

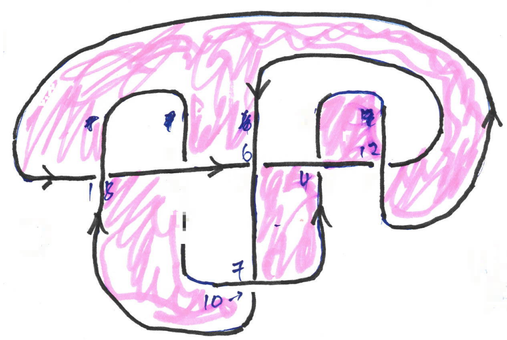
$$6\ \ -10\ \ 16\ \ 12\ \ 14\ \ -2\ \ 8\ \ 4.$$

Minus signs record that circuit markers 10 and 2 occur at underpasses.

3.5.2 Reconstructing a diagram

This (reduced alternating) diagram D of a knot K is the result of realizing the DT code

$$8\ \ 6\ \ 12\ \ 10\ \ 2\ \ 4.$$



One begins with the subarcs in a horizontal line leading to crossings $1,2,3,\ldots$ whose counterparts $8,10,6\ldots$ are indicated initially by vertical struts waiting to be joined up.

After crossing $5/12$, one has to loop back to arrive at $3/6$, and here there is a choice. In D the loop is on top, which assigns to $3/6$ a negative writhe sign; looping bottom would have effectively reflected the diagram and changed *all* the writhe signs. An extra horizontal line is then needed to deal with crossing $7/10$. The rest of D is determined until one reaches $5/12$, but placing the final big loop will not affect the isotopy class.

In this way, we see that the same code gives rise to both a knot and its mirror image.

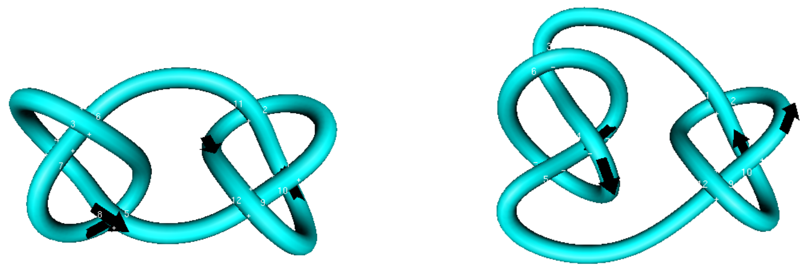
3.5.3 Composite examples

We have shown that, given a DT code, it is possible to construct an associated knot diagram. It is known that a DT code of a prime knot (like in the tables) will uniquely specify the ambient isotopy class of the knot up to reflection.

On the other hand, the DT code

$$4\ \ 6\ \ 2\ \ 10\ \ 12\ \ 8$$

is a valid specification of alternating diagrams for the composite knots $L3_1\#R3_1$ and $R3_1\#R3_1$ illustrated overleaf, even though theses two are *not* mirror images.



The composite nature of the knots can be recognized by observing that the DT code pairs numbers within the two subsets $\{1,2,3,4,5,6\}$ and $\{7,8,9,10,11,12\}$ of crossing labels. The example exploits the ambiguity in reconstructing a knot and its mirror image.

3.6 The colouring group

Let A_+ be the colouring matrix determined by a chess-boarding of D, as before.

Definition. The **determinant** of D is $|\det A|$ where A is a submatrix of A_+ obtained by deleting any one row and one column.

The previous results guarantee that this non-negative quantity does not depend on A. Since we have obtained it in a systematic way from the system of colouring equations, which is unchanged by R-moves, $|\det A|$ depends only on the ambient isotopy class of the link L. We may therefore talk about the **determinant of a link** or knot.

Definition. Having chosen A, the **colouring group** C of D is the abelian group generated by the symbols $x_1, \ldots, x_{c-1}$ subject to the relations given by the rows of A.

It can be shown that the isomorphism class of C likewise only depends on the ambient isotopy class of the link, and that $|C| = |\det A|$ provided A is invertible.

Example. It follows from the matrix calculation on p32 that the knot 4_1 has $C \cong \mathbb{Z}_5 = \mathbb{Z}/5\mathbb{Z}$ a cyclic group of order 5. We shall revisit this example next.

3.6.1 Diagonalization of integer matrices

Let D be the diagram of 4_1 that we previously coloured on p32 with

$$A = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ -1 & 0 & 2 \end{pmatrix}.$$

Then

$$C = \{x_1, x_2, x_3 : 2x_1 - x_2 = 0,\ -x_1 + 2x_2 - x_3 = 0,\ -x_1 + 2x_3 = 0\},$$

where the x_i should now be thought of as mere symbols, not integers. More abstractly, C is the quotient $\mathbb{Z}^3/(A^\top\mathbb{Z}^3)$ of the set of all integer column vectors by the subgroup generated by columns of $A^\top$. By applying elementary row operations to A we can make it upper triangular and deduce that

$$C = \{x_1, x_2, x_3 : x_1 = 2x_3,\ x_2 = 4x_3,\ 5x_3 = 0\} = \{x : 5x = 0\} \cong \mathbb{Z}_5,$$

which coincides with what we discovered at the start of §3.2.

Lemma [Smith Normal Form, SNF]. Suppose that A is a square matrix with integer entries. There exist integer matrices P, Q with determinant 1 such that PAQ is diagonal with entries $d_1, d_2, \ldots$.

The proof of this lemma (that crops up in the matrix presentation of modules) uses elementary row *and column* operations. We omit the details.

3.6.2 A colouring theorem

Theorem. Let D be a connected reduced diagram of some knot or link L. Let $\mathbf{\Delta} = |\det A|$ denote its determinant.
(i) If $\mathbf{\Delta} = 0$ then L is n-colourable for any $n \geqslant 2$.
(ii) If $\mathbf{\Delta} \neq 0$ then L is n-colourable iff $\gcd(\mathbf{\Delta}, n) > 1$.

Proof. We can assume that A is obtained from A_+ by deleting the last row and column. Use SNF to write $PAQ = B$ where $B = \mathsf{diag}(d_1, \ldots, d_{c-1})$. Allowing P to have determinant ± 1, we can assume that each d_i in the SNF is a non-negative integer. Set $\mathbf{y} = Q^{-1}\mathbf{x}$. The system of equations for n-colourability can be written

$$A\mathbf{x} = \mathbf{0} \mod n \quad \Rightarrow \quad P^{-1}B\mathbf{y} = \mathbf{0} \mod n \quad \Rightarrow \quad B\mathbf{y} = \mathbf{0} \mod n,$$

so $d_i y_i = 0 \mod n$ for $1 \leqslant i \leqslant c-1$. We are assuming that $x_c = 0$.
(i) If $\mathbf{\Delta} = 0$ then $d_i = 0$ for some i and we can take $y_i = 1$ and all other $y_j = 0$.
(ii) If $\gcd(\mathbf{\Delta}, n) > 1$ then n has a prime factor $p > 1$ in common with some d_i. Thus, $d_i = pe$ and $n = pf$, and we can take $y_i = f$ and $y_j = 0$ if $j \neq i$. Conversely if there is a solution with $y_i \neq 0$ mod n then $\gcd(d_i, n) > 1$. □

The SNF of the Goeritz matrix G will also determine the colouring group, so typically the diagonal form of A will have lots more 1's than that of G.

3.6.3 Finite abelian groups

Any finite abelian group is known to be the product of cyclic groups. We are using SNF to establish this theorem in our situation, although we are working with additive notation (so zero is the identity, and 'product' becomes 'direct sum').

The proof of the colouring theorem actually determines the structure of the set of solutions to the colouring equations, and so the colouring group C. The latter is an abelian group with addition. By converting (changing basis) from the symbols x_j to y_i, we see that

$$C \;\cong\; \mathbb{Z}_{d_1} \oplus \cdots \mathbb{Z}_{d_{c-1}}, \qquad d_i \geqslant 0,$$

where $\mathbb{Z}_d = \mathbb{Z}/d\mathbb{Z}$ is a cyclic group of order (size) d, except that $\mathbb{Z}_0 = \mathbb{Z}$ denotes the infinite cyclic group. Note that $\mathbb{Z}_1 = \{0\}$ is the identity group, so if $d_i = 1$ there is effectively no contribution to the direct sum above.

Corollary. If $\boldsymbol{\Delta} \neq 0$ then $|C| = \boldsymbol{\Delta}$. But if $\boldsymbol{\Delta} = 0$ then C is infinite.

There are different ways of decomposing a finite abelian group $C = \bigoplus \mathbb{Z}_{d_i}$ as a direct sum of cyclic groups. One can always arrange that $d_i | d_{i+1}$ to get a unique description. Or (since $\mathbb{Z}_{pq} \cong \mathbb{Z}_p \oplus \mathbb{Z}_q$ if $\gcd(p,q) = 1$), one can take each d_i to be a prime power.

4 Knot polynomials

In this section, we shall define the Kauffman bracket $\boxed{D}$ (which is a Laurent polynomisl in a variable denoted A) of a knot diagram D. We shall then use it, together with the writhe $w(D)$, to define the Jones polynomial $\mathbf{V}(K)$ of a knot K, which is a Laurent polynomal in a variable denoted by t (sometimes one uses q).

It is important to note that both $\boxed{D}$ and $w(D)$ depend on the choice of diagram D representing K, whereas $\mathbf{V}(K)$ depends only on the ambient isotopy class of K.

4.1 The Kauffman bracket

Recall that a regular isotopy between link diagrams D, D' is defined by a sequence of moves R0,R2,R3. Let O denote a diagram of the unknot with no crossings (a 'circle'), and let $O\,D$ denote the disjoint union of plane diagrams.

Proposition. The rules

$$\begin{cases} \text{(i)} & \boxed{O} = 1 \\ \text{(ii)} & \boxed{O\,D} = -(A^2 + A^{-2})\boxed{D} \\ \text{(iii)} & \boxed{\times} = A\,\boxed{\asymp} + A^{-1}\boxed{)(} \end{cases}$$

allow one to associate to each link diagram D a Laurent polynomial $\boxed{D}$ in A (so one can have positive and negative powers of A and constants) that depends only on the regular isotopy class of D, so that $D \approx D' \Rightarrow \boxed{D} = \boxed{D'}$ (see p24).

The usual notation for the (Kauffman) bracket is $\langle D \rangle$, but using boxes will make it easier to surround parts of diagrams. In either case, it is customary not to include the variable, which is always denoted 'A'.

Example. The bracket of the disjoint union of $n \geqslant 1$ circles equals $(-A^2 - A^{-2})^{n-1}$.

4.1.1 A word about splittings

Each boxed term in rule (iii) denotes the Kauffman bracket of an entire diagram with the strands joined inside the boxes as shown, and no changes outside the boxes. The rule relates the bracket of 3 diagrams, in which a given crossing X is eliminated ('split') in its two possible planar ways.

If we regard the overpass as the x-axis (in either direction), a **positive splitting** (whose resulting diagram has a coefficient A^{+1}) corresponds to opening up a channel between the odd quadrants. It is therefore achieved by placing a 'splitting marker' between Q1 and Q3, which can be imagined to be the line with Cartesian equation $y = +x$. Similarly, a **negative splitting** corresponds to joining the even quadrants so as to insert the splitting marker $y = -x$.

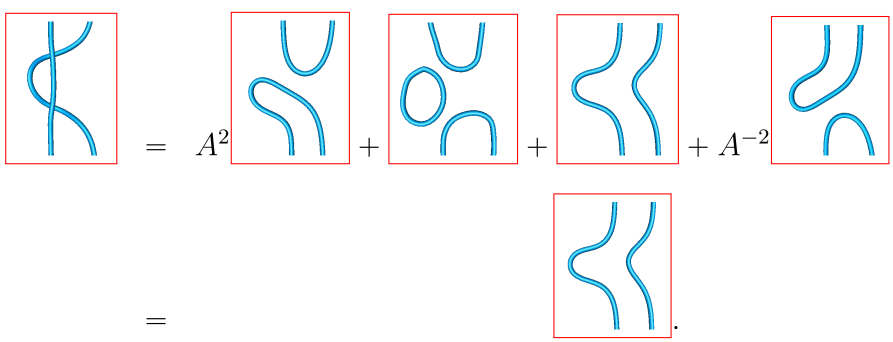
Note that orientation plays no role in the definition of the bracket. In particular, the crossing X in (iii) has no sign associated to it, and (by rotating the page) any crossing can be viewed as if it looks like X. However, if the underlying diagram is chess-boarded then X acquires a chess sign (p42), and joining the black regions will result a splitting whose sign equals that chess sign (positive iff Q1,Q3 are black).

4.1.2 Invariance under R2

If a diagram has more than one crossing, the order in which one applies (iii) is irrelevant. For example, one can apply (iii) to each of 2 crossings of a diagram D to decompose $\boxed{D}$ into the sum of 4 terms, with coefficients

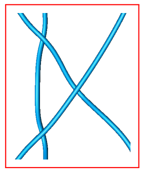
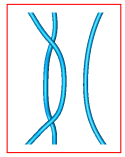
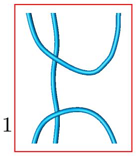
$$A^2, \quad AA^{-1} = 1, \quad A^{-1}A = 1, \quad A^{-2}.$$

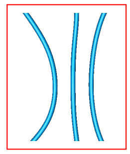
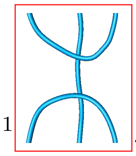
Resolving the next 2 crossings gives 4 terms, but (ii) causes 3 of these to cancel out. Hence, the double underpass can be removed without changing the bracket polynomial:



4.1.3 Invariance under R3

Apply (iii) to the diagonal crossing below (the overpass happens to be SW–NE now), then R2 invariance to the diagram with coefficient A. This gives a symmetrical configuration, which establishes R3 invariance.

 $= A$  $+ A^{-1}$ 

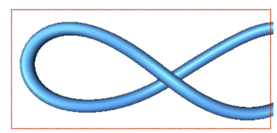
$= A$  $+ A^{-1}$ .

Starting with a diagram with c crossings, we can keep 'splitting' so as to end up with 2^c diagrams, each one D_i (the result of a **state**) with no crossings. Thus, $D_i \approx n_i O$ and

$$\boxed{D_i} = (-A^2 - A^{-2})^{n_i - 1}$$

where each n_i $(1 \leqslant i \leqslant 2^c)$ is a positive. See §4.3.

4.1.4 Effect of R1



Let D' be a link diagram incorporating the twist shown, and let D be the diagram in which R1 has been used to eliminate the twist (keeping the ends fixed). The latter is negative in the sense that its writhe sign is -1, chosen to conform with our diagram of (iii). We want to relate $\boxed{D'}$ (shown) and $\boxed{D}$.

If the crossing X is split positively (so a splitting marker is placed horizontally), we get a single connected arc inside the box, and so D altogether. If X is split negatively (vertical splitting marker), we get an extra circle resulting in a diagram $D'' \sim OD$. So

$$\boxed{D'} = A\boxed{D} + A^{-1}\boxed{D''} = A\boxed{D} + A^{-1}(-A^2 - A^{-2})\boxed{D} = -A^{-3}\boxed{D}.$$

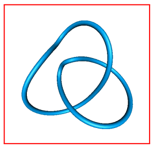
Now $w(D') = w(D) - 1$, and it follows that

$$(-A)^{-3w(D')}\boxed{D'} = (-A)^{-3w(D)}\boxed{D}.$$

Given the diagram D of a link L, the quantity $(-A)^{-3w(D)}\boxed{D}$ is invariant by moves R0,R1,R2,R3. It therefore depends only on the ambient isotopy class of L: it is (up to a change of variable) the **Jones polynomial** of L.

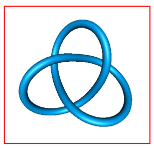
4.1.5 Examples: Hopf link and trefoil diagrams

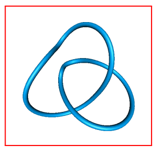
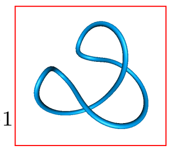
Splitting the 2 crossings of a diagram of the Hopf link gives 2^2 states and



$$= A^2\boxed{O\!\!\!\circ} + 2\boxed{O} + A^{-2}\boxed{OO} = (A^2 + A^{-2})(-A^2 - A^{-2}) + 2 = -A^4 - A^{-4}.$$

This allows us to determine the bracket of a diagram of the right-handed trefoil $R3_1$:



$= A$  $+\, A^{-1}$  (both twists here are negative)

$$= A(-A^4 - A^{-4}) + A^{-1}(A^{-3})^2 \ = \ -A^5 - A^{-3} + A^{-7}.$$

Exercises. (i) The bracket of the 3-leaved diagram of $L3_1$ is $A^7 - A^3 - A^{-5}$.

(ii) If mD is the diagram D in which all crossings have been reversed then $\boxed{mD}$ is obtained from $\boxed{D}$ replacing A by A^{-1}.

4.2 The Jones polynomial

We have already defined this implicitly for the diagram D of an **oriented** link L, by combining the writhe and Kauffman bracket, so the result is unchanged by all R-moves.

The Jones polynomial $\mathbf{V}(L)$ (at times denoted $\mathbf{V}_L(t)$) is in fact a Laurent polynomial in a variable $t^{1/2}$ that is substituted in place of A^{-2}. It was introduced by Vaughan Jones in 1984, but it was quickly understood that it can be derived from Kauffman's bracket:

Definition. If D is a diagram for L then $\mathbf{V}(L) = (-A)^{-3w(D)}\boxed{D}$ with $A = t^{-1/4}$.

$\mathbf{V}(L)$ an invariant of the isotopy class of L as an **oriented link**, but orientation is irrelevant if L is a knot.

We will see that if L has an odd number of components then $\mathbf{V}(t)$ is actually Laurent polynomial in t. Otherwise it will be $t^{1/2}$ times such a Laurent polynomial. It has integer coefficients.

Exercise. The Jones polynomial of the mirror image of a link is found by substituting t^{-1} in place of t.

4.2.1 Examples of knot polynomials

Unlike the bracket, $\mathbf{V}$ is a true knot invariant, so its value can be computed using *any* valid diagram, which is very convenient. Previous formulae imply that $\mathbf{V}$ of any unknot equals 1, and

$$\begin{aligned}\mathbf{V}(L3_1) &= (-A)^{+9}\boxed{D_L} = -A^9(A^7 - A^3 - A^{-5}) = -t^{-4} + t^{-3} + t^{-1} \\ \mathbf{V}(R3_1) &= (-A)^{-9}\boxed{D_R} = -A^{-9}(A^{-7} - A^{-3} - A^5) = -t^4 + t^3 + t\end{aligned}$$

in accordance with the previous exercise.

Exercise. An easier-to-memorize formula reflects the fact that 4_1 is achiral:

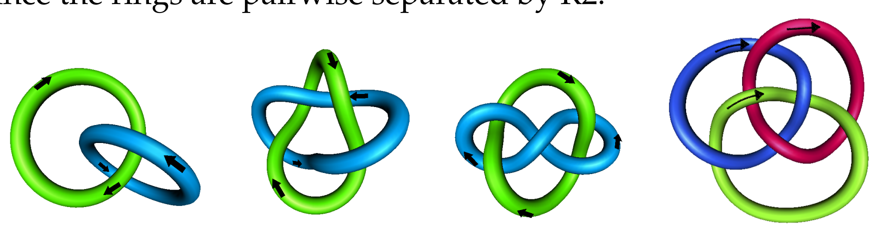
$$\mathbf{V}(4_1) = t^{-2} - t^{-1} + 1 - t + t^2.$$

Results to come. (i) If L is a link with $\ell \geqslant 1$ components then $\mathbf{V}_L(1) = (-2)^{\ell-1}$, so a knot K has $\mathbf{V}_K(1) = 1$. This is easily proved using the skein relation below.

(ii) If a knot K has a (connected, reduced) alternating diagram with c crossings, then the difference between the highest and lowest powers of t (the 'Laurent span') in $\mathbf{V}(K)$ equals c.

4.2.2 Examples of link polynomials

The first three oriented links have respective writhes $2, -4, 1$ (see sheet 1). The Hopf link below is right-handed. The underlying Solomon link is 'right-handed', but has been given a left-handed orientation. The centre crossing of the Whitehead link is here right-handed. The Borromean rings are oriented so that the central 'triangle' is cyclic, but $w = 0$ since the rings are pairwise separated by R2.



With the chosen configurations and orientations, one can show that

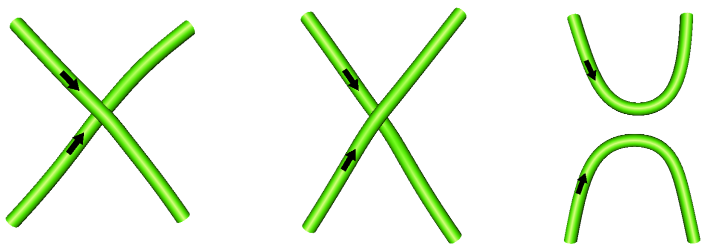
$\mathbf{V}(\text{right Hopf link}) = -t^{1/2} - t^{5/2}$
$\mathbf{V}(\text{Solomon link}) = -t^{-9/2} - t^{-5/2} + t^{-3/2} - t^{-1/2}$
$\mathbf{V}(\text{Whitehead link}) = -t^{-3/2} + t^{-1/2} - 2t^{1/2} + t^{3/2} - 2t^{5/2} + t^{7/2}$
$\mathbf{V}(\text{Borromean rings}) = -t^{-3} + 3t^{-2} - 2t^{-1} + 4 - 2t + 3t^2 - t^3$

4.2.3 The skein relation for V

Skein, pronounced 'skane', as in *a skein of yarn on a sewing bobbin*, *a skein of hair*, *a skein of geese*, *a skein of incoherent proofs*.

Surround a crossing of an **oriented** link diagram D by a box and orient the page so that arrows enter on the left and leave on the right, and it looks like L_+ or L_- below. There are exactly 3 ways to arrange the strands in the box with at most one crossing, those shown. In particular, there is only *one* way to split an **oriented crossing**.

There is a well-known relation between the Jones polynomials of the links one can form by rearranging the strands in such a box:



L_+ $\quad$ L_- $\quad$ L_0

Proposition. Set $z = t^{1/2} - t^{-1/2}$. Then

$$\boxed{t^{-1}\mathbf{V}(L_+) - t\mathbf{V}(L_-) = z\mathbf{V}(L_0)}$$

Exercise. If $L_\pm$ are knots, then L_0 is a link with two components. This helps explain the appearance of half-integral powers of t as coefficients of $\mathbf{V}(L_0)$.

Proof of the proposition. The skein relation is proved by splitting the two crossings using Kauffman's rule (iii). Let $w_0 = w(L_0)$ denote the writhe of the diagram representing L_0. The writhe sign of the featured crossing in $L_\pm$ is ± 1, so the writhe of the entire diagram representing $L_\pm$ equals $w_0 \pm 1$. We know that

$$A^4\boxed{L_+} = A^5\boxed{\asymp} + A^3\boxed{)(} \quad \text{and} \quad A^{-4}\boxed{L_-} = A^{-3}\boxed{)(} + A^{-5}\boxed{\asymp}$$

Therefore

$$\begin{aligned} t^{-1}\mathbf{V}(L_+) &= (-A)^{-3w(L_+)}A^4\boxed{L_+} &= (-A)^{-3w_0}\left(-A^2\boxed{\asymp} - \boxed{)(}\right) \\ t\mathbf{V}(L_-) &= (-A)^{-3w(L_+)}A^{-4}\boxed{L_-} &= (-A)^{-3w_0}\left(-\boxed{)(} - A^{-2}\boxed{\asymp}\right) \end{aligned}$$

Subtracting,

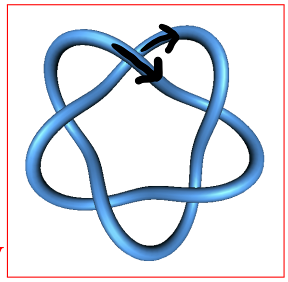
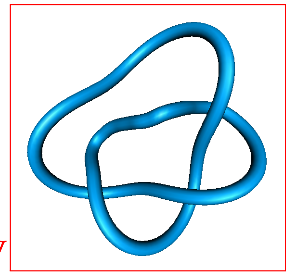
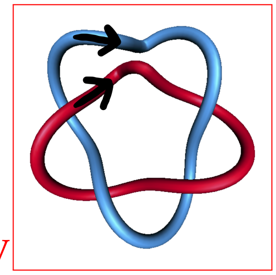
$$t^{-1}\mathbf{V}(L_+) - t\mathbf{V}(L_-) = (-A^2 + A^{-2})(-A)^{-3w_0}\boxed{\asymp} = z\mathbf{V}(L_0).$$

We have used the change of variable $t = A^{-4}$ freely. $\square$

4.2.4 Examples: the Jones polynomials of 5_1 and of a Solomon link

We shall apply the skein relation in order to compute the Jones polynomial of the knot 5_1 represented below left. This diagram has writhe $+5$, and its crossings are of the type seen in L_+. Set $z = t^{1/2} - t^{-1/2}$ again (purely as an abbreviation).

Focus on the top crossing; when it is reversed, apply R2 to get the middle diagram:

$t^{-1}\mathbf{V}$  $-\, t\,\mathbf{V}$  $=\, z\,\mathbf{V}$ 

Therefore

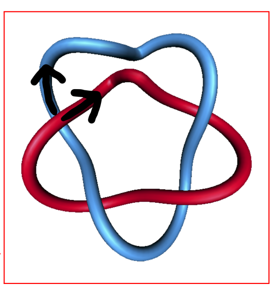
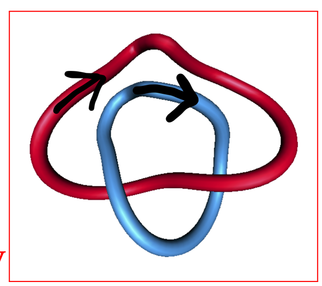
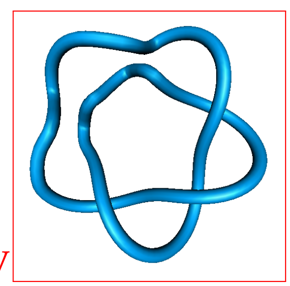
$$\mathbf{V}(5_1) = t^2\mathbf{V}(R3_1) + t\,z\,\mathbf{V}(S)$$

where S is the 'right-handed' Solomon link on p50, but this time with a right-handed orientation (so $w = +4$).

We will see next that $\mathbf{V}(S)$ can be expressed in terms of the polynomials

$$\mathbf{V}(RH) = -t^{5/2} - t^{1/2}, \quad \mathbf{V}(R3_1) = -t^4 + t^3 + t$$

of the right-handed trefoil knot and Hopf link. Focus on the top left crossing:

$t^{-1}\mathbf{V}$  $-\, t\,\mathbf{V}$  $=\, z\,\mathbf{V}$ 

Therefore

$$\mathbf{V}(S) = t^2\mathbf{V}(RH) + t\,z\,\mathbf{V}(R3_1) = -t^{11/2} + t^{9/2} - t^{7/2} - t^{3/2}.$$

Exercise. Why is this merely t^6 times the polynomial on p50?

Putting it all together,

$$\mathbf{V}(5_1) = -t^7 + t^6 - t^5 + t^4 + t^2.$$

4.2.5 A list of Jones knot polynomials

3_1	3	$-t^4+t^3+t$
4_1	5	$t^2-t+1-t^{-1}+t^{-2}$
5_1	5	$-t^7+t^6-t^5+t^4+t^2$
5_2	7	$-t^6+t^5-t^4+2t^3-t^2+t$
6_1	9	$t^4-t^3+t^2-2t+2-t^{-1}+t^{-2}$
6_2	11	$t^5-2t^4+2t^3-2t^2+2t-1+t^{-1}$
6_3	13	$-t^3+2t^2-2t+3-2t^{-1}+2t^{-2}-t^{-3}$
7_1	7	$-t^{10}+t^9-t^8+t^7-t^6+t^5+t^3$
7_2	11	$-t^8+t^7-t^6+2t^5-2t^4+2t^3-t^2+t$
7_3	13	$-t^9+t^8-2t^7+3t^6-2t^5+2t^4-t^3+t^2$
7_4	15	$-t^8+t^7-2t^6+3t^5-2t^4+3t^3-2t^2+t$
7_5	17	$-t^9+2t^8-3t^7+3t^6-3t^5+3t^4-t^3+t^2$
7_6	19	$-t^6+2t^5-3t^4+4t^3-3t^2+3t-2+t^{-1}$
7_7	21	$t^4-2t^3+3t^2-4t+4-3t^{-12}+3t^{-2}-t^{-3}$
8_{19}	3	$-t^8+t^5+t^3$

$t=1$ gives $\ell=1$. The second column shows determinants, also obtained ($\pm$) by setting $t=-1$!

A green polynomial denotes an achiral knot.

The red polynomials are those of torus knots and are 'missing' one or more powers of t. 8_{19} is the first non-alternating knot.

4.3 Summing over states

Fix a diagram D of a link L.

One can compute the Kauffman bracket of D by choosing one of its crossings to split using (iii), and then applying the same step to each of the two resulting diagrams. Proceeding in this way we finish with 2^c diagrams, each of which has no crossings left, and is therefore a disjoint union of 'circles' (closed curves, some possibly nested). Because the changes only take place around each crossing, the order in which we choose the crossings is irrelevant, as in the proof on p47 of R2 invariance.

A choice of splittings for *every* crossing is called a **state**. A splitting is positive iff it the diagram that results is paired with A, see p46. So a state can be regarded as a map

$$s\colon \{1,2,\ldots,c\} \longrightarrow \{+1,-1\}$$

that assigns to each crossing the sign in which it is to be split. There are 2^c such states. One can indicate by $s(D)$ the diagram that results when s is applied to D, and by $|s|$ (pronounced 'mod s') the number of circles in $s(D)$. Each state involves p positive crossings and n negative ones, with $p+n=c$ and

$$p-n=\sum_{i=1}^{c} s(i).$$

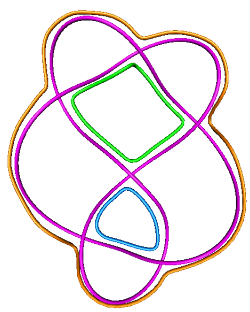
4.3.1 The sum of all fears

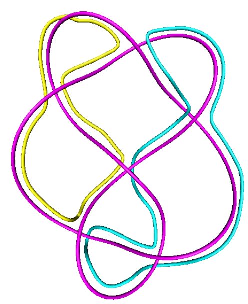
Repeated application of rule (iii) yields the 'big sum'

Proposition. $\boxed{D} = \sum_{2^c \text{ states}} A^{p(s)-n(s)}(-A^2 - A^{-2})^{|s|-1}.$

There are two special states, namely s_+ with $s_+(i) = +1$ for all i (so all splittings are positive), and likewise s_- (all splittings negative).

Below left, we see the effect of s_+ on a mauve diagram of 6_2, and $|s_+| = 3$. Right, we see a state s with $p = n = 3$ and $|s| = 2$. One can also verify that $|s_-| = 5$. (If it is difficult to see the overpasses, note that the diagrams are alternating, and all the chess signs are $+1$ when they are chess-boarded with white on the outside!).





Exercise. In general, if s, s' are states whose values (splittings) disagree at exactly one crossing, then $|s'| = |s| \pm 1$.

4.3.2 Alternating links

Suppose that a link has an alternating diagram D that is positively chess-boarded with B/W black/white regions. A positive splitting of every crossing will unite all the black regions and result in unknots around the white ones, so $|s_+| = W$. Similarly, $|s_-| = B$.

Lemma. If D is **reduced** (definition on p36) then the highest power of A in $\boxed{D}$ will arise from s_+ and Kauffman term

$$A^c(-A^2 - A^{-2})^{W-1} = \pm A^{c+2W-2} \pm \dots$$

Proof. Set $h = c + 2W - 2$. Let s' denote a state with exactly one negative splitting. We know that $|s'| = |s_+| \pm 1$, so the highest power of A in the Kauffman term of s' will either equal h or $h - 4$. We must prove it is the latter, otherwise there is a danger that terms in A^h could cancel each other out. But we claim that the hypothesis 'reduced' forces $|s'| = |s| - 1$. For, having made a positive splitting, the resulting two arcs must form part of *two distinct circles*, for otherwise the splitting would be an isthmus, with an arc joining one side to the other, following the coast of a white sea, cf. p36. If that splitting is changed to negative, the white regions in Q2,Q4 are united, two circles become one, and $|s'| = |s_+| - 1$. □

4.3.3 Laurent span

This is the difference between the highest and lowest powers in a Laurent polynomial, so we can talk about the A-span in $\boxed{D}$, or the t-span in $\mathbf{V}(L)$.

Suppose that D is both reduced and *connected*. Then it has $B + W = c + 2$ regions by Euler's formula, see p16. We have seen that the highest power of A in $\boxed{D}$ arises from $h = c + 2W - 2$. Similarly, the lowest power will be $\ell = -c - 2B + 2$. So the A-span is

$$h - \ell = 2c + 2(W + B - 2) = 2c + 2(c + 2 - 2) = 4c.$$

To pass from the Kauffman bracket $\boxed{D}$ to the Jones polynomial $\mathbf{V}(K)$ we multiply by $(-A)^{-3w(D)}$ (which does not affect the span), and replace A by $t^{-1/4}$. In conclusion,

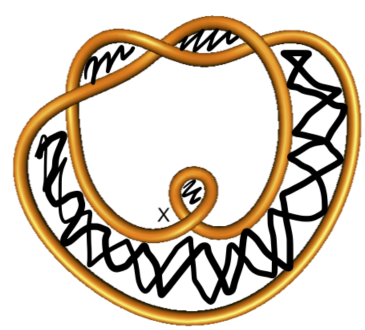
Corollary. If a link L has a connected reduced alternating diagram with c crossings, c equals the t-span of $\mathbf{V}(L)$, and so in these circumstances c **depends only on the ambient isotopy class of** L.

Examples:

$$\mathbf{V}(6_2) = t - 1 + 2t^{-1} - 2t^{-2} + 2t^{-3} - 2t^{-4} + t^{-5}$$
$$\mathbf{V}(\text{right Hopf}) = -t^{5/2} - t^{1/2} \quad \text{but} \quad \mathbf{V}(OO) = -t^{-1/2} - t^{1/2}.$$

4.3.4 A bad example

The diagram D represents the ambient isotopy class of the trefoil knot $R3_1$, since the bottom crossing X (which creates an isthmus) can be eliminated by an R1 twist.



Let $Z = -A^2 - A^{-2}$. Since $s_+(D)$ unites all the black regions, it consists of $W = 2$ circles, i.e. $|s_+| = 2$. This gives the leading term $A^c Z^{2-1} = A^4 Z = -A^6 - \cdots$ in the big sum. The crucial question is what happens when one changes one of the splittings to be negative.

If X, it would split off the isthmus leading to 3 circles, and a term $A^2Z^2 = A^6 + \cdots$. If one of the other 3 crossings, we get back to one circle, so the big sum only ever has two terms in A^6, which cancel out. In fact

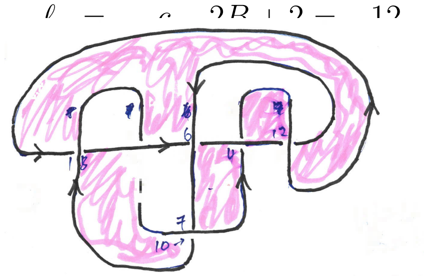
$$\boxed{D} = A^4Z + A^2(Z^2+3) + A^0(6Z) + A^{-2}(4Z^2) + A^{-4}Z^3 = A^2 + A^{-6} - A^{-10}.$$

Since $w(D) = 3 - 1 = 2$, the Jones polynomial $\mathbf{V}(R3_1)$ is $A^{-6}\boxed{D} = t + t^3 - t^4$.

4.3.5 Example form DT code

Because there were no minus signs in the DT code on p42, the reconstructed diagram is alternating. After chess-boarding, it has $W = 4$ white regions and $B = 4$ 'black' (meaning pink) ones. It follows from §4.3.3 that the highest and lowest exponents of A in the Kauffman bracket are

$$\begin{aligned} h &= c + 2W - 2 = 12, \\ \ell &= -c - 2B + 2 = -12 \end{aligned}$$



Moreover, since $W - 1 = B - 1$ is odd, they both appear with coefficient -1:

$$\boxed{D} = -A^{12} \pm \cdots - A^{-12}.$$

The six crossings (from left to right) have writhe signs $---+++$, so $w(D) = 0$, and

$$\mathbf{V}(K) = -t^{-3} + \cdots - t^3.$$

It follows from the list of Jones polynomials on p53 that K is ambient isotopic to 6_3. The fact that 6_3 is achiral explains why $\mathbf{V}(K)$ is unchanged when t is replaced by t^{-1}.

4.4 Historical remarks

Conway's knot in Cambridge

In 1987, Kauffman, Murasugi & Thistlethwaite improved the Corollary on p55 by relating $\mathbf{V}$ to spanning trees in a graph arising from the chess-boarding, so as to prove the

Theorem. If a link L has a reduced alternating diagram D with c crossings then $c = \mathrm{cr}(L)$ is its crossing number.

This theorem establishes was one of P.G. Tait's conjectures, dating from 1898.

It is an open questions as to whether any knot with $\mathbf{V}(K) = 1$ is necessarily trivial, i.e. ambient isotopic to an unknot? There do exist non-trivial *links* with $\mathbf{V}(L) = 1$. There *are* examples of distinct knots K_1, K_2 with $\mathbf{V}(K_1) = \mathbf{V}(K_2)$, for example

$$-t^4 + 2t^3 - 2t^2 + 2t + t^{-2} - 2t^{-3} + 2t^{-4} - 2t^{-5} + t^{-6}$$

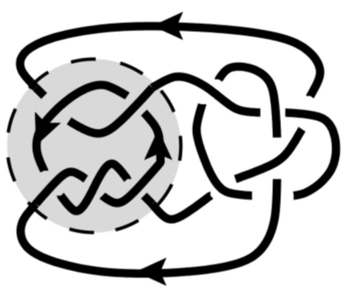
is the Jones polynomial shared by the Conway and K-T knots, each with $\mathsf{cr} = 11$.

To cite an easier result, the Jones polynomial behaves well with respect to the connect sum of links, namely $\mathbf{V}(K_1 \# K_2) = \mathbf{V}(K_1)\mathbf{V}(K_2)$. In particular, this confirms that there are exactly three ambient isotopy classes formed by connecting two trefoil knots.

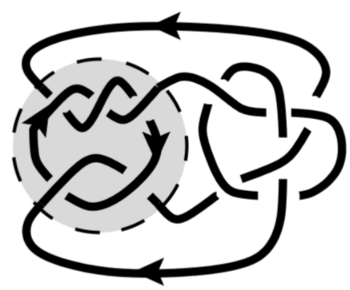
4.4.1 Tangles and mutants

The notion of a tangle was introduced by John Conway around 1970, as a means to construct knots. A 2-tangle is part of a diagram contained in a disk, which exits the boundary at 4 points, which here could be labelled NE,SE,SW,NW. Rational tangles have a sequence of twists inside the disk, as in a pretzel knot, and (remarkably) can be characterized by an associated continued fraction.

Given a knot K incorporating a tangle, a **mutant** can be formed by (for example) rotating in space only that part in the disk:



Conway
($11n34$, genus $= 3$)



Kinoshita--Terasaka
($11n42$, genus $= 2$)

(Images courtesy S. Chmutov.) The two are distinguished by the least genus of the oriented surfaces that they bound (see §5.4.2, or by the fundamental group of their complements in $\mathbb{R}^3$ (see §11.3).

4.4.2 The Alexander polynomial

The first 'polynomial' to represent the ambient isotopy class of a link L was discovered by J.W. Alexander in 1923. It was (and often still is) introduced by labelling regions of a diagram and examining colouring-type equations

$$x_1 - tx_2 + tx_3 + x_4 = 0 \mod \mathbf{\Delta}_L(t),$$

where (x_2, x_4) is an oriented overpass, and $\mathbf{\Delta}_L(t)$ plays the role of the determinant. The Alexander polynomial $\mathbf{\Delta}(L) = \mathbf{\Delta}_L(t)$ is no longer an integer, but a Laurent polynomial in a variable t, well defined up to multiplying by an integer power t^k. The determinant of L is recovered by taking $t = -1$ (so what we previously called $\mathbf{\Delta}$ is now $|\mathbf{\Delta}_L(-1)|$).

Conway showed in 1969 that that the Alexander polynomial satisfies a simpler version of the skein relation, namely

$$\mathbf{\Delta}(L_+) - \mathbf{\Delta}(L_-) = z\,\mathbf{\Delta}(L_0), \quad \text{where } z = t^{1/2} - t^{-1/2} \text{ again.}$$

Although neither $\mathbf{V}(L)$ nor $\mathbf{\Delta}(L)$ can be obtained from the other, a 2-variable 'HOMFLY' polynomial defined by an enhanced skein relation generalizes both of them.

5 Intermission

5.1 Surfaces from knots

Let D be a knot diagram with c crossings, and give it a chess-boarding. This defines a 'cloth' surface $\mathscr{M}$ in $\mathbb{R}^3$ by interpreting each crossing vertex as hiding a 180^{o} twist in the cloth. In practice, one could manufacture such a surface by first taking disjoint pieces of cloth to match the black regions, and then joining them with a slim ribbon twisted by 180^{o} (like this ) for each crossing.

If the ribbon is attached to one region, its free end should be twisted anti-clockwise iff the crossing has chess sign $+1$. Provided the twist matches the strands of the crossing in this way, the boundary of $\mathscr{M}$ will be the knot we started with.

5.1.1 Orientation

The surface $\mathscr{M}$ is called **two-sided** in space if one can consistently distinguish two sides throughout, so that if one side is coloured black and the other (say) grey, the two colours only meet at the surface's boundary if there is one. This is possible if all crcuits from one region back to itself pass through an *even* number of crossings. This is the case for the diagram with two black regions on p59); we shall see that the resulting surface is homeomorphic to a torus minus a disk.

Two-sided surfaces are **orientable**, meaning that there is a well-defined notion of clockwise rotation at any point. Any surface without boundary in $\mathbb{R}^3$ (such as a sphere or torus) has this property because it has an inside and an outside. But orientability is a characteristic of the abstract surface, not just the way it sits inside space.

The trefoil diagram with three black regions gives a one-sided surface. It is in fact a 3-twisted Möbius band in $\mathbb{R}^3$, though it is homeomorphic to the usual 1-twisted Möbius band. One-sided surfaces are **non-orientable**: it is possible to carry around the surface a small set of Cartesian axes so that y points (as usual) to the left of x at the start, but to the right of x after a path returning to the start.

5.2 Euler characteristic

Suppose that a surface $\mathscr{M}$ has been subdivided using a total of V vertices, E edges (homeomorphic to finite intervals), and F faces (these could be triangles or other regions homeomorphic to a disk).

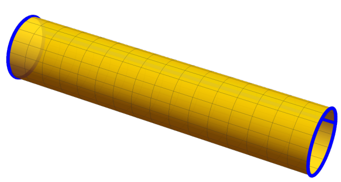
Theorem. $\chi = V - E + F$ is independent of the choice of subdivision, and is therefore a topological invariant of $\mathscr{M}$. It is called the **Euler characteristic** or number.

We shall see a proof, after we have given a combinatorial definition of a surface.

A version of the theorem applies to a planar graph, which actually deteermines subdicvision of the sphere S. In this case, F is the number of regions including the outside and $\chi = 2$, which corresponds to the Euler characteristic of S.

For S itself, one can choose any vertex and take its complement to be its face, to yield $(V, E, F) = (1, 0, 1)$. A less economic subdivision would consist of 2 faces (the north and south hemispheres), 2 vertices (say in Iquitos and Singapore) and the 2 half-equators, so $(2, 2, 2)$. *Any* subdivision gives $\chi = 2$.

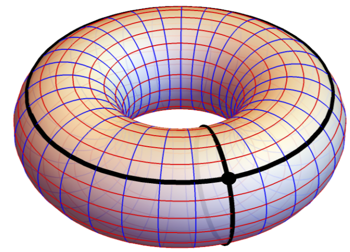
Of course, S is a surface without boundary. The next simplest orientable surface without boundary is obtained by adding a handle to S, and the resulting surface is topologically a torus T. (The symbol T represents the homeomorphism class, rather than any given model of the torus.)



A more symmetrical-looking torus in $\mathbb{R}^3$ can be constructed by identifying the ends of a flexible cylinder. Topologically, a cylinder is equivalent to a sphere minus two disks; each time a disk is removed, χ drops by 1, so $\chi(\text{cylinder}) = 2-1-1 = 0$. This is consistent with the model of cylinder obtained from a rectangle by identifying the edges top and bottom, which leaves $(V, E, F) = (2, 3, 1)$ and again $\chi = 2$.

The closed edges left and right of the rectangle correspond to the boundary circles of the cylinder that we need to further identify to obtain the torus. All 4 vertices of the rectangle have now become one, so $\chi = 1 - 2 + 1 = 0$.

If we choose the dimensions correctly, we can map the rectangle continuously **onto** the surface of revolution in $\mathbb{R}^3$ so that the united edges map to the two orthogonal circles intersecting in a single vertex.



5.3 A preview of classification

Recall that a subset of $\mathbb{R}^3$ is **compact** iff it is bounded (finite in extent) and closed (contains its limit points), see p13. A compact surface (being closed) must include its boundary if it has one. The boundary (itself being compact) must be a link consisting of $r \geqslant 0$ knots, trivial or otherwise.

We shall see that there are three quantities that can be used to distinguish a compact surface $\mathscr{M}$ topologically:

(i) whether or not it is orientable (Yes/No),

(ii) the Euler characteristic, an integer $\chi \leqslant 2$,

(iii) the number $r \geqslant 0$ of boundary components it has.

The precise result will be formulated combinatorially in the sequel.

If $\mathscr{M}$ is an orientable surface without boundary then it is homeomorphic to a 'torus with g holes', equivalently a sphere with g handles attached to it. To attach a handle one must first remove two disks (reducing χ by 2) and then add a curved cylinder (leaving χ unchanged). So $\chi(\mathscr{M}) = 2 - 2g$, where $g \geqslant 0$ is the so-called genus. But if there are r boundary components, we have

$$\chi(\mathscr{M}) = 2 - 2g.$$

5.4 Seifert's algorithm

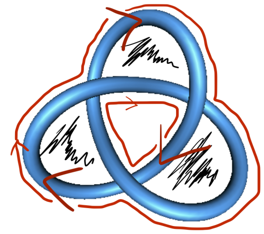
For a cloth surface, $\chi(\mathscr{M}) = B - c$ where B is the number of black regions. This is because each black region is a face, and each crossing ribbon  amounts to subtracting an edge from each face and adding a rectangle, so $\chi \mapsto \chi - 2 + 1$ reduces by 1. The twist is irrelevant here, χ only depends on the shadow of the diagram. The standard diagram of 4_1 gives a non-orientable surface with $\chi = 3 - 4 = -1$. However,

Theorem. Any knot is the boundary of some **orientable surface** in $\mathbb{R}^3$.

Proof. Choose an oriented diagram for the knot, and define a 'Seifert state' σ to split each crossing as in the right-hand side of the skein relation, see p51. We are left with a certain number $|\sigma|$ of oriented circles, each of which is the boundary of a disk in the plane. Paste in a 180^{o}-twisted ribbon between pairs of them to correctly interpret each crossing. (To do this, stack the disks vertically if one is inside another.) This will ensure that the boundary of the resulting surface is faithfully encoded by the diagram. Each disk acquires an orientation from its boundary circle, and the ribbons will consistently propagate the orientation from disk to disk. □

5.4.1 Examples: the trefoil and figure-eight knots

The 3-leaved diagram D (right) of $R3_1$ has writhe $+3$. If D is chess-boarded with white on the outside, the chess signs are all positive. That implies that $\sigma = s_+$ and $|\sigma| = W = 2$.



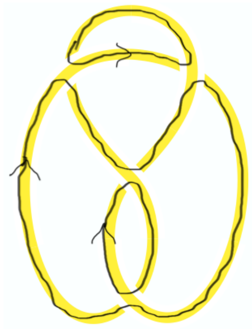
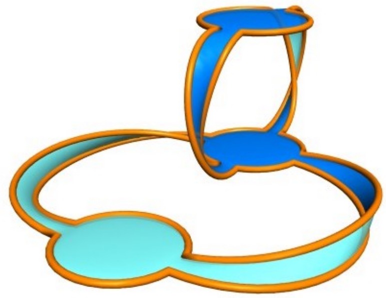
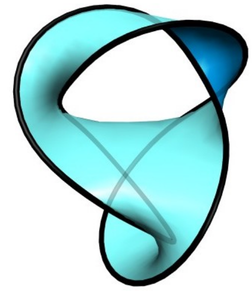
The resulting Seifert surface $\mathscr{S}$ is formed from a disk for each circle and a twisted band to interpret each crossing, so that the boundary of $\mathscr{S}$ is a trefoil knot.

$\mathscr{S}$ is orientable, because it is 2-sided in $\mathbb{R}^3$.

$\chi(\mathscr{S}) = 12 - 18 + 5 = -1$ (6 vertices on each disk, add 6 edges to cap off the strips), or more simply $2-3$ (two disks and 3 ribbons).

The classification of surfaces will tell us that $\mathscr{S}$ **is homeomorphic to a torus minus a disk**. If, on the other hand, the ribbons are not twisted, we merely have a sphere with 3 disks removed, yet again $\chi = 2 - 3 = -1$.

Orient the diagram (below left) of 4_1. There is then only one way to split each crossing preserving the orientation. This gives a state σ consisting of 3 circles, one nested inside another, leading to two images of the resulting Seifert surface, the second deformed from the disks and ribbons of the first.

It is easy to compute the Euler characteristic of the surface $\mathscr{S}$ produced by Seifert's algorithm in general. Each disk contributes $+1$ to χ, and each ribbon -1. So $\chi(\mathscr{S}) = |\sigma| - c$, and the examples above (centre and right) again have $\chi = -1$.

5.4.2 Genus of a knot

Definition. The **genus** $g(K)$ of a knot K is the least genus of any orientable surface that it bounds. It is therefore a knot invariant.

If $g(K) = 0$ then K bounds a disk and is the unknot. It is also known that:
(i) $g(K)$ is achieved by Seifert's algorithm if the knot is alternating, but not in general;
(ii) the genus of a connect sum is additive:

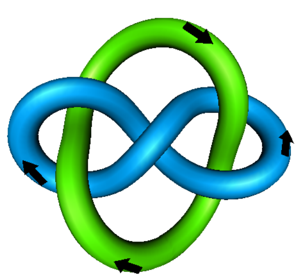
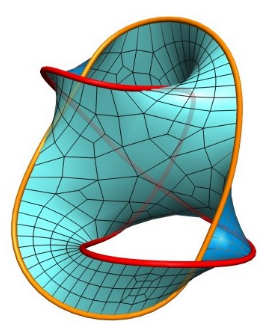
$$g(K_1 \# K_2) = g(K_1) + (K_2).$$

The connect sum requires one to first select an orientation for each knot, and in general its ambient isotopy class will depend on this choice.

Example. The Seifert surface $\mathscr{S}$ of the Whitehead link has

$$\chi(\mathscr{S}) = |\sigma| - c = 3 - 5 = -2,$$

which implies that $-2 = 2 - 2g - r = -2g$, and $g = 1$. Therefore $\mathscr{S}$ is homeomorphic to a torus minus two disjoint disks, though this is not obvious from its picture!

5.4.3 A Seifert surface for 6_3

This image was used as a logo for the 2023 module pages. Like other images on previous pages, it was compiled with the software SeifertView.

6 Surfaces without boundary

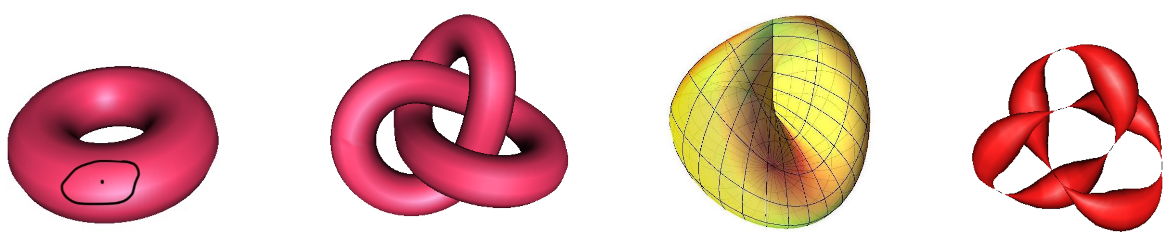
6.1 Definition and examples

In this section, we shall be concerned with connected compact surfaces, **without boundary**. As already explained, These come in two families: orientable and non-orientable. The former ones can be embedded (i.e., mapped injectively and continuously) into $\mathbb{R}^3$, the others (assuming there is no boundary) can not. We shall classify both types.

Definition. A surface without boundary (or 2-manifold) is a (Hausdorff) topological space, each point of which lies in an **open set** that is homeomorphic to a disk in $\mathbb{R}^2$.

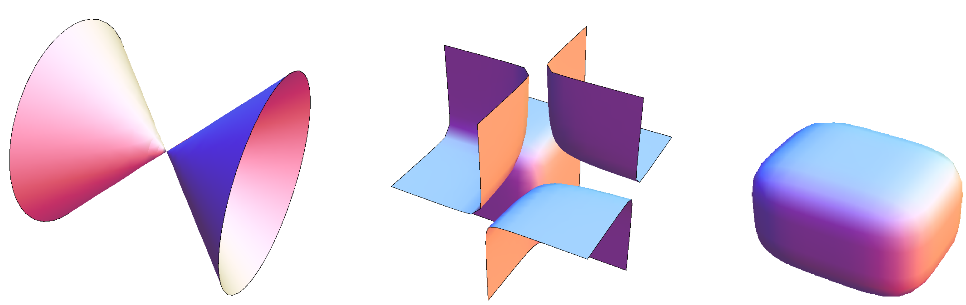
6.1.1 Examples

The last two below do not have this 'locally Euclidean' property: the third has a segment that resembles the intersection of two planes, the fourth (a trefoil of sausages) has 6 points that each resemble the vertex of a double cone. The first two are homeomorphic, and therefore deemed equivalent in our forthcoming classification.



One can describe surfaces in space using Cartesian equations.

Exercise. The equations overleaf all describe subsets of $\mathbb{R}^3$. Which are surfaces? Which are connected and which are compact? Which equations match the three pictures?



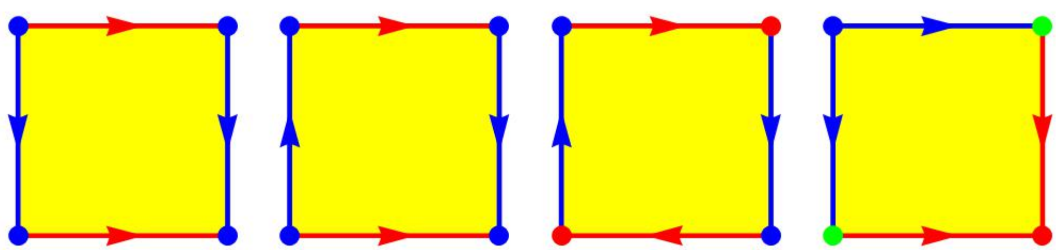
Answers can be found at the end of this section.

(i) $|x| + |y| + |z| = 1$
(ii) $yz + zx + xy = 0$
(iii) $xyz = 1$
(iv) $x^8 + y^8 + z^8 = 1000$
(v) $\dfrac{1}{x} + \dfrac{1}{y} + \dfrac{1}{z} = 0.$

6.1.2 Abstract models for four surfaces

If enough cuts are made on a surface, so it is subdivided by a graph of vertices and edges, the pieces can be deformed and laid out as plane figures. The surface can then be reconstructed by 'sewing' along the cuts* (see p68). In four simple cases, a single square suffices with the cuts defining pairs of edges.

Left to right are models of the torus (as on p60), Klein bottle, projective plane, sphere:



$aba^{-1}b^{-1}$ $\quad$ $aba^{-1}b$ $\quad$ $abab$ $\quad$ $abb^{-1}a^{-1}$

By labelling the edges, and reading clockwise from top left, these models can be described by the 'words' displayed. The last two can be simplified into AA and AA^{-1}.

6.1.3 Embedded surfaces in $\mathbb{R}^3$

A compact surface $\mathscr{M}$ is **embedded** in $\mathbb{R}^3$ if there is a continuous injective map

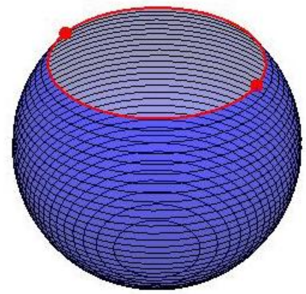
$$\mathscr{M} \longrightarrow \mathbb{R}^3.$$

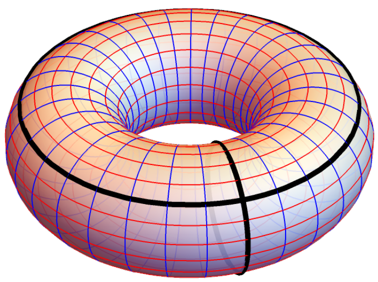
It will then be homeomorphic to its image.

As a topological space, the unit sphere

$$S^2(1) = \{(x, y, z) : x^2 + y^2 + z^2 = 1\} \subset \mathbb{R}^3$$

is what we get when we take a blue disk, push it into a bag, and fasten its top boundary up with a 'zip'. The symbol S stands for any surface homeomorphic to $S^2(1)$, for example the octahedron $|x| + |y| + |z| = 1$.





We have already seen the torus parametrized by a rectangle $\mathscr{R}$ (equivalently, square). All 4 vertices of $\mathscr{R}$ map to common point p on the torus, and the 4 edges of $\mathscr{R}$ are identified in pairs and mapped to 2 circles intersecting in p.

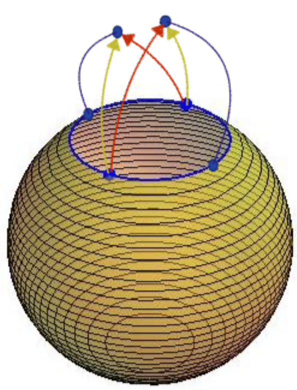
6.1.4 Immersed surfaces in $\mathbb{R}^3$

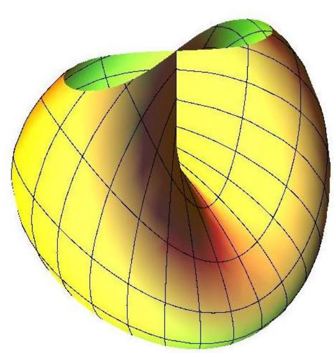
The Klein bottle K is formed by passing a cylinder through itself. It can be mapped almost injectively into $\mathbb{R}^3$, but there is a circle of self-intersection that actually represents two disjoint circles on K.

The projective plane P can be immersed in $\mathbb{R}^3$ as a sphere with **crosscap**, parametrized by

$$\{(\cos u \sin 2v,\ \sin u \sin 2v,\ \cos^2 v - cos^2 u \sin^2 v)\}.$$

It has a segment of self-intersection, where the image passes through itself:





We have truncated the top of the crosscap in the image merely to show the inside. These immersed surfaces in $\mathbb{R}^3$ are one-sided in the sense that the image of a continuous path in P or K can travel to all sides.

6.1.5 Classification

The next few lectures will occupy us with developing examples and techniques that will lead to a proof of the

Classification theorem (v1). Any connected compact surface **without boundary** is homeomorphic to one of the following:
(i) a sphere S, or a sphere with $g \geqslant 1$ handles, equivalently a 'torus with g holes',
(ii) a sphere with $h \geqslant 1$ crosscaps.

In case (i), the surface $\mathscr{M}$ can be embedded in $\mathbb{R}^3$, and is **orientable**. The integer g is called its **genus**, and S itself corresponds to $g = 0$. One can regard a surface of genus g as the 'connected sum' (in analogy to knots) of g tori, and one often writes

$$\mathscr{M} = \underbrace{T\#\cdots\#T}_{g} = gT$$

where the equals sign really means 'homeomorphic to'.

In (ii) the surface must contain a Möbius band and is **non-orientable**. Any such can be regarded as the connected sum of h copies of the projective plane, so $\mathscr{M} = hP$. We shall see that the Klein bottle has $h = 2$, so we write $K = P\#P = 2P$. A key step in the proof of the classification theorem is to show that $K\#P = T\#P$.

Answers to the exercise. Multiplying (v) by xyz gives (ii); both equations describe a double cone, not a surface. (iv) is the compact 'ellipsoidal box', (iii) is an unbounded surface with 4 components, which is a 'smoothing' of 3 intersecting planes; (i) is an octahedron (not shown) made of 8 triangles, one in each octant.

6.2 Subdividing surfaces

Given a surface (possibly with boundary), we cut it into smallish 'triangles' (possibly with curved sides) with matching directed edges, which can be labelled ($\xrightarrow{a}$, $\xrightarrow{b}$, etc) so the same label only occurs on two adjacent triangles. Such a **triangulation** can always be chosen* with the following properties:

(i) each edge belongs to at most two triangles;

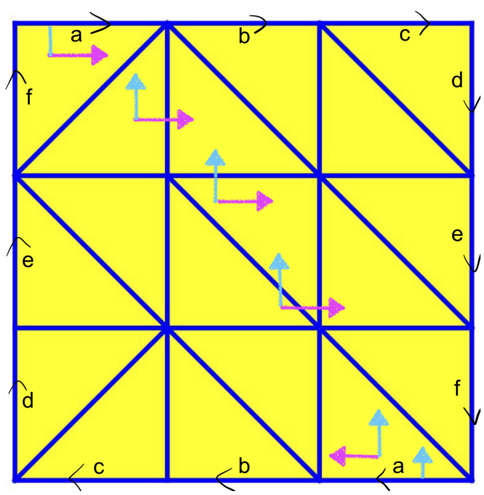
(ii) the set st(v) ('star') of triangles sharing a vertex v forms a polygon homeomorphic to a closed disk (this would also hold for point on the boundary if there were one);

(iii) two triangles are either disjoint or meet in one common edge or a single vertex.

Properties (i) and (ii) guarantee the locally Euclidean property. Each edge must belong to *exactly* two triangles assuming (as currently) that there is no boundary.

*A rigorous proof of the existence of a triangulation was the final step [T. Radó, 1925] in the classification of surfaces. It relies on Schoenflies' theorem stating that any closed curve in the plane bounds a disk (also cited on p36).

Example: Triangulating the projective plane. If we start with the rectangle or square model, we can divide it into 2 triangles, but many more are needed to prevent triangles having two isolated vertices in common. Below, we see that 18 triangles suffice, but the minimum for P is known to be 10.

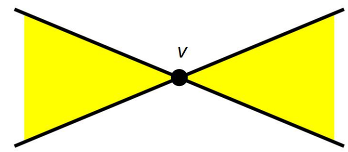


Here, the boundary is encoded by the word $abcdef\,abcdef$, equivalent to $ADAD$ or even AA (as on p66). A right-handed set of axes can be propagated from one triangle to another, but when it passes across the boundary of the square, it becomes left-handed. So there is no consistent sense of (clockwise/anti-clockwise) rotation, and the surface $\mathscr{M}$ is said to be **non-orientable**. This is like having a one-sided surface in $\mathbb{R}^3$, but the concept does not require one to visualize $\mathscr{M}$ in space.

6.2.1 Polygonal surfaces

Once a surface $\mathscr{M}$ is triangulated, the triangular pieces can be separated and then partially re-assembled **in the plane**, adding one triangle at a time by pairing 2 **edges**.

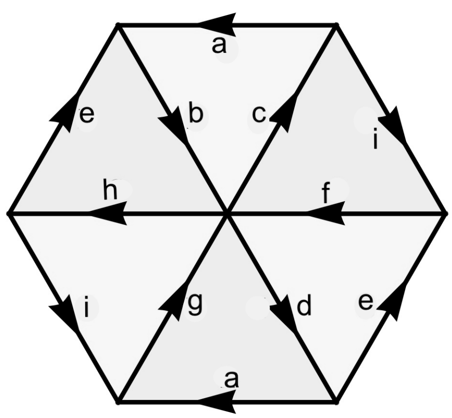
When this process is complete, we can avoid isolated vertices like v shown, because rule (ii) implies that a punctured neighbourhood of v is connected, so we can always find triangles filling in above and/or below v, with their edges identified.



The order of matchings can result in different configurations but, assuming $\mathscr{M}$ is **connected**, we can always suppose that the final result is a single (filled) **polygon**. Its edges will then be labelled, like the squares on p66.

In practice, one can relax the rules for triangulation, and divide a surface into a finite union of polygons, whose boundary edges are matched in pairs. Each polygon has a perimeter code consisting of a word W (such as $abca^{-1}d$). The symbol a^{-1} indicates that a and a^{-1} have an opposite sense of rotation around the polygon, though the inverse is omitted (by the lecturer) if the symbol accompanies an arrow.

6.2.2 A hexagonal model of a torus



$$\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$$

This 6-sided polygon $\mathscr{P}$ is triangulated with edges labelled in accordance with the rows and columns (read upwards) of the matrix above. The perimeter code

$$a^{-1}ie^{-1}ai^{-1}e$$

determines a surface $\mathscr{M}$, which (we shall explain) is a topological **quotient** of $\mathscr{P}$.

Note that:

(i) No occurrence of $\ldots x \ldots x \ldots$ means that $\mathscr{M}$ has a consistent sense of (clockwise/anticlockwise) rotation, so is **orientable**.

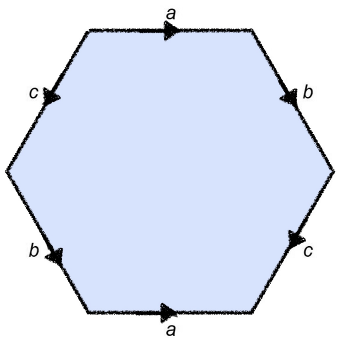
(ii) Move 3 triangles so as to transform the hexagon into an 8-sided polygon with perimeter code $\underbrace{dh^{-1}}\,\underbrace{g^{-1}f}\,\underbrace{hd^{-1}}\,\underbrace{f^{-1}g} = DGD^{-1}G^{-1}$, which is equivalent to that of a torus on p66. (The solution is in the second introductory video).

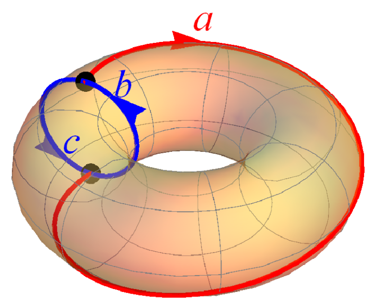
A Dehn twist. We now know that the quotient of the hexagon with perimeter code $a^{-1}ie^{-1}ai^{-1}e$ (or $abca^{-1}b^{-1}c^{-1}$ with a change of letters) is homeomorphic to the torus.

Consequently, there must exist a continuous surjective mapping

$$\mathbb{R}^2 \supset \mathscr{P} \xrightarrow{q} \mathscr{M} \subset \mathbb{R}^3,$$

where $\mathscr{M}$ is a torus of revolution. The map q can be understood by cutting the torus along the blue circle, and then twisting it by 180° before sewing it back together:





A technique used by Max Dehn around 1910 to generate the so-called mapping class group of a surface.

6.3 Surfaces as topological quotients

Given a connected surface $\mathscr{M}$ without boundary, recall that we can make a model of it as a polygon $\mathscr{P}$, with its boundary or perimeter $\partial\mathscr{P}$ consisting of $2n$ sides. A perimeter code consists of a word with $2n$ letters occurring in pairs. The edges of each pair need to be 'sewn' or 'glued' so as to eliminate the boundary of $\mathscr{P}$ and recover $\mathscr{M}$.

More formally, the code defines an equivalence relation on $\mathscr{P}$ whose classes have size 1, 2 or more. An interior point of $\mathscr{P}$ is only equivalent to itself. An interior point of an edge a is equivalent to exactly one other point on the other edge labelled a or a^{-1}.

A vertex will be equivalent to at least one other point, unless it occurs in the middle of aa^{-1} or $a^{-1}a$, which (if a 2-gon on its own) represents the sphere S.

$\mathscr{M}$ can be defined as the set $\widehat{\mathscr{P}}$ of equivalence classes, and there is a *surjective* mapping, called the **projection**,

$$q\colon \mathscr{P} \to \mathscr{M}.$$

The image $\mathscr{M}$ is not merely a set, but a topological space: a subset U of $\mathscr{M}$ is declared open iff $q^{-1}(U)$ is open in $\mathscr{P}$ (and so of the form $\mathscr{P} \cap V$ where V is an open subset of $\mathbb{R}^2$). This makes q continuous.

Example. Let $x \in \partial\mathscr{P}$ so that $m = q(x) = \{x, x'\}$. Then an open subset $q^{-1}(U)$ of $\mathscr{P}$ containing x must contain a small semicircular region around both x and x'.

6.3.1 Normal form

Our task will be to understand when two different perimeter codes (words) for $\mathscr{P}$ give rise to homeomorphic (i.e. topologically equivalent) surfaces. For this purpose, we shall define a surface combinatorially, purely in terms of a word and the operations on it that preserve topological type. To interpret these operations, we shall rely on visual intuition to understand $\mathscr{M}$ as the quotient space $\widehat{\mathscr{P}}$.

Classification theorem (v2). Any connected surface without boundary arises from a polygon with an even number of edges, identified in pairs, and is uniquely specified by exactly one of the following words:
(i) aa^{-1}, or $a_1b_1a_1^{-1}b_1^{-1}\cdots a_gb_ga_g^{-1}b_g^{-1}$ with $g \geqslant 1$,
(ii) $c_1c_1\cdots c_hc_h$ with $h \geqslant 1$.

Don't forget that polygon means 'filled polygon': an open interior together with boundary edges and vertices.

The two cases coincide with those on p68. In particular, the code $a_ib_ia^{-1}{}_ib^{-1}{}_i$ signals the presence of an attached torus or handle, and c_jc_j the presence of a crosscap.

6.3.2 Sewing the projective plane and Klein bottle

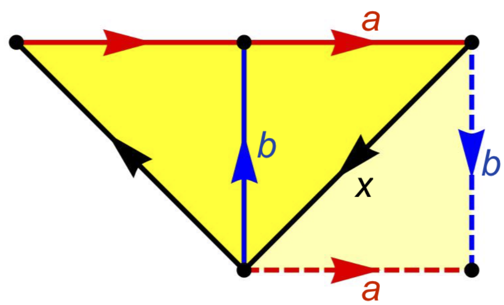
Consider the square model of the projective plane P on p66, with perimeter code $abab$. If we replace one of the a's with c, we obtain the code $abcb$ that represents a surface with boundary. It is clear that $abcb$ is the code for describing a Möbius band M in an abstract way. The boundary consists of ac^{-1} which (in view of the matching vertices) is homeomorphic to a circle.

The simplest way of visualizing M in $\mathbb{R}^3$ is to regard the square as a piece of paper, and attach the two edges labelled b by twisting the paper by 180°. The boundary is then an *unknot* (whilst had the twist been through 540°, it would have been a trefoil, corresponding to the cloth model on p59). Forgetting the band, we can flatten and deform the unknot into the boundary of a closed disk D, itself represented by a 2-gon with code ac^{-1}. In notation that will be developed in §7.1, we can write

$$ac^{-1} + abcb \;\sim\; ac^{-1} + cbab \;\sim\; abab,$$

which all goes to show that P is obtained by uniting D to M. One can try to visualize an immersion of P in $\mathbb{R}^3$ by sewing the disk onto the twisted band, though in practice this requires some intersection of the disk with the band (so, a self-intersection of P).

We saw on p66 that the cylinder model of the Klein bottle K leads to a square with word $aba^{-1}b$, but this is not in normal form, so is not recognizable from the theorem. We can convert the word $aba^{-1}b$ into normal form with a 'cut and paste' operation, namely we remove the triangle below the diagonal and re-position it so as to match up the two b edges.



First we split the diagram into triangles axb, $a^{-1}x^{-1}b$. These can be combined by 'flipping' the second over to match the b edges. In symbols,

$$axb + (a^{-1}x^{-1}b)^{-1} \sim axxa \sim aaxx,$$

where $\sim$ indicates that the associated surfaces are homeomorphic. The new word is $c_1c_1c_2c_2$ in the notation of the theorem, confirming (we shall see) that K is a sphere with two crosscaps.

7 Symbolic operations

7.1 Cutting, flipping and pasting

The usual model of the Klein bottle K shows that it is the quotient of a closed square in $\mathbb{R}^2$ by the equivalence relation defined by the perimeter code $aba^{-1}b$. The explanation on p72 shows (by dividing that square into two triangles and rearranging) that K can also be obtained from a square with code $aaxx$. We write $aba^{-1}b \sim aaxx$ to indicate that the two quotients are homeomorphic.

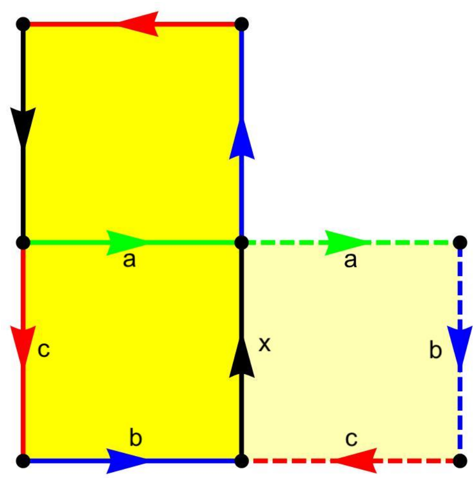
A variant of the proof (focussing on the triangle with the sides a, b and adding the other diagonal) goes like this symbolically:

$$\begin{aligned} \boxed{ab}a^{-1}b \sim abx^{-1} + xa^{-1}b &\sim xb^{-1}a^{-1} + xa^{-1}b \\ &\sim a^{-1}xb^{-1} + bxa^{-1} \sim a^{-1}xxa^{-1} \sim a^{-1}a^{-1}xx. \end{aligned}$$

The symbol '$+$' merely unites codes arising from a disjoint union of polygons.

In this section, we shall start with a 6-gon and perimeter code $aabcb^{-1}c^{-1}$, and show that we get the same (i.e. a homeomorphic) surface if instead we start with the code $aabbcc$, or the normal form $(c_1c_1)(c_2c_2)(c_3c_3)$ in the notation of the classification theorem v2:

Proposition. $(aa)(bcb^{-1}c^{-1}) \sim (c_1c_1)(c_2c_2)(c_3c_3)$.



We start the proof of the proposition with the perimeter code $a\boxed{abc}b^{-1}c^{-1}$ of the lower horizontal rectangle. The black cut (closing the boxed edges a, b, c) gives the union of two squares, and the 'sum' of two words

$$abcx \ + \ x^{-1}b^{-1}c^{-1}a,$$

the first labelling the right-hand square.

By inverting the first summand (corresponding to flipping the square), and starting the second with a, this can be combined into

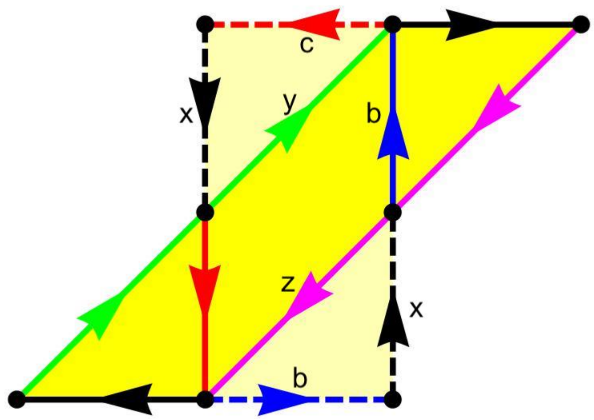
$$(abcx)^{-1}(ax^{-1}b^{-1}c^{-1}) \sim x^{-1}c^{-1}b^{-1}x^{-1}b^{-1}c^{-1},$$

a code that now represents the vertical rectangle.

Recall that $A \sim B$ means that the two polygons whose pairs of edges are identified by means of the words A, B are homeomorphic surfaces.

7.1.1 Substitution

We have reached the vertical rectangle, with code $x^{-1}c^{-1}b^{-1}x^{-1}b^{-1}c^{-1}$. The first operation is to separate the triangle with two sides $x^{-1}c^{-1}$ top left, flip it over and reattach:



$$\begin{aligned}
&\boxed{x^{-1}c^{-1}}b^{-1}x^{-1}b^{-1}c^{-1}\\
\sim\ & x^{-1}c^{-1}y^{-1} + yb^{-1}x^{-1}b^{-1}c^{-1}\\
\sim\ & xyc + c^{-1}yb^{-1}x^{-1}b^{-1}\\
\sim\ & yb^{-1}x^{-1}b^{-1}xy\\
\sim\ & \boxed{x^{-1}b^{-1}}xyyb^{-1}.
\end{aligned}$$

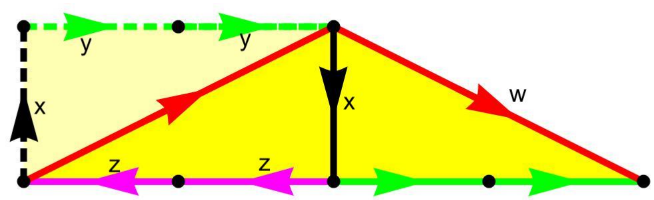
To do this, we have effectively substituted $y = x^{-1}c^{-1}$ (so $c^{-1} = xy$). Now we move the triangle bottom right, which amounts to substituting $z = x^{-1}b^{-1}$ (so $b^{-1} = xz$):

$$\sim\ x^{-1}b^{-1}z^{-1} + zxyyb^{-1}\ \sim\ xzb + b^{-1}zxyy\ \sim\ xzzxyy.$$

This is the code for the parallelogram.

We now represent the parallelogram by a smaller rectangle with the same perimeter code $xyyxzz$. The diagonal cut splits this into two triangles, but we can skip the cutting and pasting steps by substituting $w = xyy$ (so $x = wy^{-1}y^{-1}$):

$$\boxed{xyy}xzz\ \sim\ w(wy^{-1}y^{-1})zz\ \sim\ wwy^{-1}y^{-1}zz.$$



Setting $c_1 = w$, $c_2 = y^{-1}$, $c_3 = z$ completes the proof. $\square$

Both sides in the Proposition are juxtapositions of codes corresponding to P and T. Soon, we shall interpret this as a homeomorphism of connected sums

$$P\#T = P\#P\#P = 3P.$$

7.2 Combinatorial models

Let $\mathscr{P}$ be a polygon modelling a surface $\mathscr{M} = \widehat{\mathscr{P}}$ without boundary. The essence of this model is a perimeter code of directed edges, telling us how to identify points of $\partial\mathscr{P}$. Our aim is to transform the code into one of the normal forms

(i) aa^{-1}, or $a_1b_1a_1^{-1}b_1^{-1}\cdots a_gb_ga_g^{-1}b_g^{-1}$,

(ii) $c_1c_1\cdots c_hc_h$.

This will be achieved by a sequence of cutting and pasting operations of the sort we have already seen that lead to homeomorphic surfaces. During this process, we could split a single polygon into one or more smaller ones, each with its own code. However, we would like to carry out these operations without having to draw pictures.

Definition. A **combinatorial model** of a surface is defined by a collection of **letters** $a_1, a_2, \ldots$ (representing edges) and one or more **words** $W_1, W_2, \ldots$ (representing complete boundaries) involving these letters or their inverses (normally at least 3 per word), so each letter appears **exactly twice overall** (if $\mathscr{M}$ is to have no boundary).

We call this data $\langle a_1, a_2, \ldots \mid W_1, W_2, \ldots\rangle$ a presentation of the surface.

7.2.1 Valid operations

Starting with the presentation $\langle a_1, a_2, \ldots \mid W_1, W_2, \ldots\rangle$, in which each a_i appears exactly twice on the right, we can recover a surface $\mathscr{M}$ by constructing a regular polygon of unit side for each word, to make it easy to identify the edges in pairs.

Proposition. The following operations on words then result in homeomorphic surfaces:

name	word		word(s)	
rotate	aB	$\sim$	Ba	
reflect	A	$\sim$	A^{-1}	
cut or paste	AB	$\sim$	$Ax^{-1} + xB$	i.e. $\{AB\} \sim \{Ax^{-1},\, xB\}$
insert or fold	A	$\sim$	Axx^{-1}	
relabelling	A	$\sim$	r	(see (b) below)

Moreover:

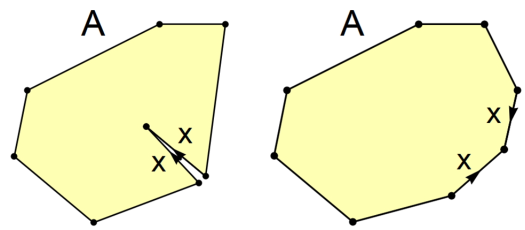
(a) Words have at least 3 letters, except that aa and aa^{-1} are words.

(b) $A, B, \ldots$ stand for 2 or more letters (excluding x), but are not in general words. If A occurs twice (perhaps with $^{-1}$), A can be replaced with a new single letter r.

(c) We shall explain after a sketch proof of the proposition that the operations can be understood group theoretically, so inverses are computed in that context. Thus $(a^{-1})^{-1} = a$, and if $A = abc$ then $A^{-1} = c^{-1}b^{-1}a^{-1}$.

The first two operations result in homeomorphisms because we merely rotate or reflect the regular polygons. Proving that cut and paste does not affect the quotient follows from the pictorial interpretations earlier.

The two images on the right explain the insert (left to right) and folding (right to left) operations. Here, A is the left boundary consisting of 6 letters (when the polygon is fully closed up).



The edges are then widened to accommodate insertion of xx^{-1}, and this process leads to a homeomorphism of the quotients of the two polygons provided identification of the edges of A is adjusted accordingly. $\square$

7.2.2 Notation from group theory

As we discovered in §7.1.1, all these operations are consistent with simple substitutions in group theory. One can push this further, and interpret the letters as generators of a (typically) infinite group, and each boundary code W as a relation $W = e$ (where e is the identity, and the word describing the sphere).

Next, the proof of the proposition on p74 is streamlined by replacing all the cutting, rearranging and pasting by substitutions.

At each step overleaf, on alternate lines, a group of boxed letters is replaced by a new letter, which is also used to replace the second occurrence of the red letter. The symbol $\sim$ means both sides define the same relation (when each is set equal to the identity), which again encodes homeomorphic surfaces:

$$
\begin{array}{lll}
aabcb^{-1}c^{-1} & \sim \boxed{abc}\,b^{-1}c^{-1}a & (abc = x^{-1} \;\Rightarrow\; a = x^{-1}c^{-1}b^{-1}) \\
 & \sim x^{-1}\,b^{-1}c^{-1}\,x^{-1}c^{-1}b^{-1} & \\
 & \sim \boxed{x^{-1}c^{-1}}\,b^{-1}x^{-1}b^{-1}c^{-1} & (x^{-1}c^{-1} = y \;\Rightarrow\; c^{-1} = xy) \\
 & \sim y\,b^{-1}x^{-1}b^{-1}\,xy & \\
 & \sim \boxed{x^{-1}b^{-1}}\,xyyb^{-1} & (x^{-1}b^{-1} = z \;\Rightarrow\; b^{-1} = xz) \\
 & \sim z\,xyy\,xz & \\
 & \sim \boxed{xyy}\,xzz & (xyy = w \;\Rightarrow\; x = wy^{-1}y^{-1}) \\
 & \sim w\,wy^{-1}y^{-1}\,zz & \\
 & \sim c_1c_1\;c_2c_2\;c_3c_3. &
\end{array}
$$

Exercise. Prove that $abca^{-1}b^{-1}c^{-1} \sim xyx^{-1}y^{-1}$ starting with $\boxed{ab}$.

7.2.3 Counting vertices

If the surface $\mathscr{M}$ is the quotient of a $2n$-gon $\mathscr{P}$ with a boundary code, then

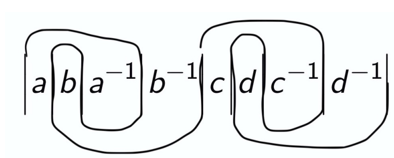
$$\chi = V - n + 1$$

where V is the number of equivalence classes of vertices (so V is the number of distinct points of $\mathscr{M} = \widehat{\mathscr{P}}$ arising from vertices of $\mathscr{P}$).

The integer V can be quickly evaluated by carrying out the identifications around $\partial\mathscr{P}$.

Key observation. If the word describing $\partial\mathscr{P}$ is in normal form, and not aa^{-1}, then $V = 1$. Therefore, $\chi = 2 - 2g$ in case (i), and $\chi = 2 - h$ in case (ii).

Let's verify this for $W = aba^{-1}b^{-1}cdc^{-1}d^{-1}$. Squiggles (here, connected) indicate that the vertices (points between letters) are equivalent:



Exercise. When W is written out linearly as above, its initial 'vertex' will always *(whether or not the perimeter code is in normal form) be equivalent to its final 'vertex', without having to rely on the fact that these two coalesce on $\mathscr{P}$.*

7.3 Proof of the classification theorem v2

We know that a connected compact surface $\mathscr{M}$ can be regarded as the quotient $\widehat{\mathscr{P}}$ of a polygon $\mathscr{P}$, defined by a word W that identifies points on the boundary $\partial\mathscr{P}$. We shall apply operations to find a sequence of words

$$W \sim W_1 \sim W_2 \sim W_3 \sim W_4$$

satisfying various conditions, with W_4 in one of the forms (i) or (ii) in the theorem.

Set $\mathbb{A}_g = (a_1 b_1 a_1^{-1} b_1^{-1}) \cdots (a_g b_g a_g^{-1} b_g^{-1})$ for $g \geqslant 1$, and $\mathbb{A}_0 = aa^{-1}$.
Set $\mathbb{C}_h = (c_1 c_1) \cdots (c_h c_h)$ for $h \geqslant 1$, and $\mathbb{C}_0 = \varnothing$.

The proof is then divided into the following steps:

Step 1. $W \sim W_1$ where all the vertices defined by W_1 map to a unique point of $\mathscr{M}$.

Step 2. $W_1 \sim W_2$ where $W_2 = \mathbb{C}_h B$ for $h \geqslant 0$, and B has no repeated letters like $b \cdots b$ without an inverse. For this, it is enough to observe that $aCaD \sim xxC^{-1}D$.

Step 3. $W_2 \sim W_3$, where $W_3 = \mathbb{C}_h \mathbb{A}_g$, assuming that $B \neq \varnothing$.

Step 4. We can assume that one of h, g is zero. This follows immediately from the proposition on p74, which tells us that $\mathbb{C}_1 \mathbb{A}_1 \sim \mathbb{C}_3$.

Implementing steps 1 and 2

For steps 1,2 and 3, we shall resort to the cutting and pasting notation, though we invite the reader to streamline the arguments with group-theoretic substitutions.

We have to modify W so that there is only 'one' vertex, i.e. $V = 1$. If this is not the case, $\partial\mathscr{P}$ has two inequivalent vertices p, q. We can assume they are adjacent and set $a = \vec{pq}$.

Suppose first that a appears as $a \cdots a$, so we can write $W = aBaC$, with $B \neq \varnothing$ (otherwise $p = q$). Then

$$W \sim aBx^{-1} + xaC \ \sim \ xB^{-1}a^{-1} + aCx \ \sim \ xB^{-1}Cx \ \sim \ xxB^{-1}C.$$

In the first cut, the ends of x are both p, so the pasting has eliminated one occurrence of q. By induction, we can get down to one vertex.

We have incidentally succeeded in converting two separated occurrences of a to a single xx, which is what is needed for Step 2.

If instead, a appears as $a \cdots a^{-1}$ then $W = aBa^{-1}C$ with $B \neq \varnothing$ (otherwise is a sphere, or aa^{-1} can be folded). A trickier argument (illustrated by q4 on Sheet 7) allows us to eliminate one occurrence of q.

Implementing step 3

Suppose that $W \sim \mathbb{C}_h B$ ($\mathbb{C}B$ for shorthand). with $B \neq \varnothing$. We need to convert B to normal form. Since $\mathscr{P}$ has a 'unique vertex', we must be able to spot a configuration $a \cdots b \cdots a^{-1} \cdots b^{-1}$ for some a, b (otherwise there would be two inequivalent vertices, one before a^{-1} and one after, there being no edges to link them). So

$$B = a\boxed{B_1 b B_2}a^{-1}C_1 b^{-1} C_2$$

for some a, b, B_i, C_j, and

$$\begin{aligned} W \sim B_1 b B_2 y^{-1} + y a^{-1} C_1 b^{-1} C_2 \mathbb{C} a &\sim B_2 y^{-1} B_1 b + b^{-1} C_2 \mathbb{C} a y a^{-1} C_1 \\ &\sim B_2 y^{-1}\boxed{B_1 C_2 \mathbb{C} a}y a^{-1} C_1 \\ &\sim B_1 C_2 \mathbb{C} a z^{-1} + z y a^{-1} C_1 B_2 y^{-1} \\ &\sim z^{-1} B_1 C_2 \mathbb{C} a + a^{-1} C_1 B_2 y^{-1} z y \\ &\sim z^{-1} B_1 C_2 \mathbb{C} C_1 B_2 y^{-1} z y \\ &\sim \mathbb{C} C_1 B_2 (y^{-1} z y z^{-1}) B_1 C_2. \end{aligned}$$

We can extract another term $a_2 b_2 a_2^{-1} b_2^{-1}$ provided $C_1 B_2$, $B_1 C_2$ are not both empty. $\square$

8 Surfaces with boundary

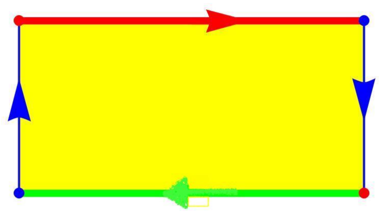
We shall continue to model surfaces combinatorially, by means of one or more words. But to allow boundaries, we shall allow each letter (or edge) to appear **twice or once**.

8.1 Cuffs

As mentioned on p72, the word $abcb$ describes a Möbius band M, whose boundary $d = ac^{-1}$ is homeomorphic to a circle. That's because the start of a and end of c^{-1} get mapped to the same point of M, as do the end of a and start of c^{-1}. If we set $ab = x^{-1}$ (equivalent to cutting $abx + x^{-1}cb$ and then pasting) then

$$\boxed{ab}cb \sim abx + x^{-1}cb \sim xab + b^{-1}c^{-1}x \sim xac^{-1}x \sim xxd.$$

We can now 'sandwich' d, using an insert and two more (cut and paste) substitutions:

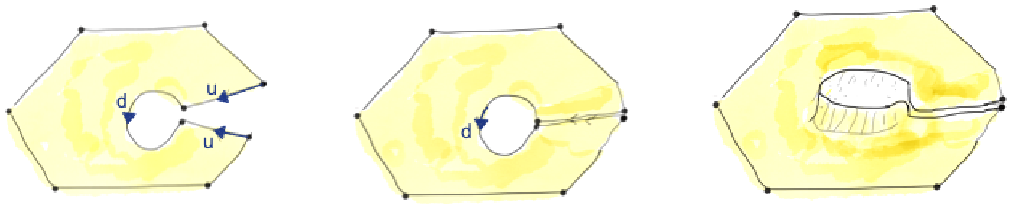


$$\begin{aligned} xxd &\sim xxdu^{-1}u \\ &\sim \boxed{ux}xdu^{-1} \quad (y = ux \Rightarrow x = u^{-1}y) \\ &\sim \boxed{yu^{-1}}ydu^{-1} \quad (z = yu^{-1} \Rightarrow y = zu) \\ &\sim zz(udu^{-1}). \end{aligned}$$

The trio udu^{-1} within a word (in which d, u do not appear elsewhere) is called a **cuff**. The letter u helps to distinguish different boundary components, and the presence of a cuff indicates that one disk has been removed from the surface:

Proposition. If a word W represents a surface $\mathscr{M} = \widehat{\mathscr{P}}$ (with or without existing boundary), then $Wudu^{-1}$ represents a surface that is homeomorphic to $\mathscr{M}$ minus an open disk (assuming that u, d do not appear in W).

Proof. A polygon with boundary $Wudu^{-1}$ is shown left. Folding the u edges defines the required homeomorphism between the quotients.



8.1.1 Naked edges

A problem is that two separate single edges inserted into the boundary of a polygon do not always produce two components. Here they unite into a single curve:

$$\begin{aligned} abda^{-1}eb^{-1} &\sim \boxed{abd}\,a^{-1}eb^{-1} \quad (abd = x \;\Rightarrow\; b^{-1} = dx^{-1}a) \\ &\sim x\,a^{-1}e\,dx^{-1}a \\ &\sim x^{-1}axa^{-1}\,ed. \end{aligned}$$

Since all vertices of $x^{-1}axa^{-1} \sim \mathbb{A}_1$ are equivalent, ed is homeomorphic to a circle, so the quotient surface is a torus with only one disk removed. To obtain two circles, one must 'insulate' the edges with cuffs; having done this, the derivation above yields

$$ab(udu^{-1})a^{-1}(vev^{-1})b^{-1} \sim (x^{-1}axa^{-1})(vev^{-1})(udu^{-1}).$$

This word represents an orientable surface with $\chi = 3 - 6 + 1 = -2$, and each cuff contributes -1 (1 extra vertex and 2 edges).

The number r of cuffs is the number of boundary components (necessarily connected compact 1-manifolds, each *homeomorphic* to a circle). If the surface is orientable and embedded in $\mathbb{R}^3$, then the boundary is a **link** and each component is a **knot**. The latter will not in general be *ambient isotopic* to a circle (think Seifert surface).

8.2 Final classification

We shall now generalize the classification theorem to connected compact surfaces (2-manifolds) **with boundary**. Such a surface can be defined by asserting that every point is contained in a subset that is homeomorphic to a **closed** disk in $\mathbb{R}^2$, as this allows boundary points.

This matches the definition of a triangulation on p68, and we can always regard such a surface $\mathscr{M}$ with boundary as a quotient $\widehat{\mathscr{P}}$ of a single polygon whose own boundary $\partial\mathscr{P}$ is defined by a word W in which each letter appears **twice or once**.

If the boundary is empty, our previous theorem asserts that we can convert W into one of the normal forms:

$\mathbb{A}_g = (a_1 b_1 a_1^{-1} b_1^{-1}) \cdots (a_g b_g a_g^{-1} b_g)$ for $g \geqslant 1$, with $\mathbb{A}_0 = aa^{-1}$.

$\mathbb{C}_h = (c_1 c_1) \cdots (c_h c_h)$ for $h \geqslant 1$.

To allow for a boundary, it turns out that we merely need to add one or more cuffs with groups of letters:

$\mathbb{D}_r = (u_1 d_1 u_1^{-1}) \cdots (u_r d_r u_r^{-1})$ for $r \geqslant 1$, with $\mathbb{D}_0 = \varnothing$.

The exact choice of letters is immaterial provided that they do not clash.

8.2.1 Topological type

Classification theorem (v3). Any connected compact surface $\mathscr{M}$ with boundary (possibly empty) arises from a polygon and a single word of exactly one of the following types:

(i) $\mathbb{A}_0$, or $\mathbb{D}_r$ with $r \geqslant 1$, or $\mathbb{A}_g\mathbb{D}_r$ with $g \geqslant 1$ and $r \geqslant 0$.

(ii) $\mathbb{C}_h\mathbb{D}_r$ with $h \geqslant 1$ and $r \geqslant 0$.

The integer r represents the number of boundary components. Since the removal of a disk (homeomorphic to the interior of a triangle) affects only F, it reduces χ by 1. This can also be understood by tracing vertices and counting edges in the normal forms above:

(i) $\chi(\mathbb{A}_g\mathbb{D}_r) = 2 - 2g - r$

(ii) $\chi(\mathbb{C}_h\mathbb{D}_r) = 2 - h - r$.

Both χ and r are topological invariants, meaning that homeomorphic surfaces have the same values. In case (i), g is again called the **genus** of the orientable surface. We have now justified the preview given on p61:

Corollary. The topological type (homeomorphism class) of a connected compact surface is entirely determined by its orientability (Y/N), together with χ and r.

8.2.2 A trickier example

We previously showed that

$$abda^{-1}eb^{-1} \sim xyx^{-1}y^{-1}ed \sim (xyx^{-1}y^{-1})(ufu^{-1}) \sim \mathbb{A}_1\mathbb{D}_1.$$

The left-hand word encodes the insertion of distinct edges in the two bottom corners of the square defining the torus T. Let's see what happens when replace T by P.

Example. The boundary code $W = abdaeb$ determines a non-orientable surface with boundary. Since $\chi = 2 - 4 + 1 = -1$, it could be $\mathbb{C}_1\mathbb{D}_2$ or $\mathbb{C}_2\mathbb{D}_1$.

$$\begin{array}{rcll} W & \sim & \boxed{abd}\,aeb & \qquad (x = abd \;\Rightarrow\; a = xd^{-1}b^{-1}) \\ & \sim & xxd^{-1}(b^{-1}eb). & \end{array}$$

We can see pictorially that the naked edge d cannot be combined with the edge e to get a common boundary, so $r = 2$ and the normal form is $\mathbb{C}_1\mathbb{D}_2$.

Cuffs occur naturally in an attempt to convert a word into normal form. But one may also be left with one or more naked letters, like d^{-1} above. One can convert them into cuffs using the trick on p81, which shoes how a single letter u can be 'passed through' $xx \sim \mathbb{C}_1$. The next lemma establishes the analogous result when $\mathbb{C}_1$ is replaced by $\mathbb{A}_1$.

8.2.3 Commutation relations

Lemma A. Represent $\mathbb{A}_1$ by $aba^{-1}b^{-1}$. Let $W = u\mathbb{A}_1E$ be a word, where E is any group of letters not involving u, a, b. Then $W \sim \mathbb{A}_1uE$, using the operations of p76.

Proof. This is accomplished by means of four (cut and paste) substitutions, each of which moves u into a different position within the group $\mathbb{A}_1$ (in the order 1,4,3,2,5). Note that E merely acts as a buffer in the proof:

$$\begin{array}{rcll} u\mathbb{A}_1E & \sim & \boxed{ua}\,ba^{-1}b^{-1}E & \qquad (x = ua \;\Rightarrow\; a^{-1} = x^{-1}u) \\ & \sim & \boxed{xb}\,x^{-1}ub^{-1}E & \qquad (y = xb \;\Rightarrow\; b^{-1} = y^{-1}x) \\ & \sim & \boxed{yx^{-1}}\,uy^{-1}xE & \qquad (z = yx^{-1} \;\Rightarrow\; x = z^{-1}y) \\ & \sim & \boxed{zuy^{-1}}\,z^{-1}yE & \qquad (w = zuy^{-1} \;\Rightarrow\; y = w^{-1}zu) \\ & \sim & (wz^{-1}w^{-1}z)uE & \sim\; \mathbb{A}_1uE. \qquad \square \end{array}$$

Together with the easier result ('Lemma C') that $u\mathbb{C}_1E \sim \mathbb{C}_1uE$, Lemma A is an ingredient in completing the proof of theorem v3 on p83. For, if we are left with a word of the form $W = \mathbb{A}_gd$, we can deduce that $W \sim \mathbb{A}_gdu^{-1}u \sim u\mathbb{A}_gdu^{-1} \sim \mathbb{A}_g\mathbb{D}_1$.

8.3 Connected sum of surfaces

We know that any connected compact surface **without boundary** is a sphere with either $g \geqslant 0$ handles or $h \geqslant 1$ crosscaps. This can be made precise using the concept of connected sum. The latter is easiest to define combinatorially in the first instance.

Definition. Let $\mathscr{M}_1, \mathscr{M}_2$ be two surfaces without boundary described as quotients of polygons by words W_1, W_2, without letters in common. Their **connected sum** is the surface (denoted $\mathscr{M}_1 \# \mathscr{M}_2$) defined by W_1W_2, i.e. the operation of **juxtaposition**.

To regard W_1W_2 as the boundary of a new polygon, we need to 'cut' the boundary of $\mathscr{P}_1$ where W_1 starts/ends and insert W_2 (also cut from $\mathscr{P}_2$). For the definition to make sense, we should be able to insert W_2 between *any* two edges of W_1. Indeed, the surface defined by W_1W_2 must only depend on those defined by $\mathscr{M}_1$ and $\mathscr{M}_2$:

Lemma J. Suppose that W_1, W_2 have even length and no unpaired letters, and W_1', W_2' similarly. If $W_i \sim W_i'$ for $i = 1, 2$ then $W_1W_2 \sim W_1'W_2'$.

Recall that '$W_i \sim W_i'$' means that one can be ontained from the other by a sequence of operations from p76, and that the quotients $\widehat{\mathscr{P}_i}$ and $\widehat{\mathscr{P}_i'}$ are then homeomorphic.

8.3.1 Rearranging independent words

Lemma J provides a short cut to dreaming up a sequence of valid operations on words to produce homeomorphic surfaces. We shall illustrate it with examples in place of a proof. Note that W_1, W_2 must be 'independent' (meaning, no letters in common).

Examples. (i) Take $W_1 = W_1'$ ($= aba^{-1}b^{-1}$, for example), $W_2 = uccu^{-1}$ and $W_2' = cc$. We know that $W_2 \sim W_2'$ by rotation, so the theorem implies

$$W_1uccu^{-1} \sim W_1cc.$$

We can deduce this by a (cut and paste) substitution, inverting what we did on p81:

$$\begin{array}{rcl} W_1W_2 & = & W_1\boxed{uc}cu^{-1}, \quad x = uc,\ c = u^{-1}x \\ & \sim & W_1xu^{-1}xu^{-1} \ \sim\ W_1W_2'. \end{array}$$

(ii) Lemma J implies that

$$W_1uxyx^{-1}y^{-1}u^{-1} \ \sim\ W_1xyx^{-1}y^{-1},$$

which is also a consequence of Lemma A. One might encounter the left-hand word during conversion to normal form.

8.3.2 Geometrical interpretation

Given a word $W = W_1W_2$ not containing the letter x, we have

$$W \;\sim\; W_1x^{-1} + xW_2.$$

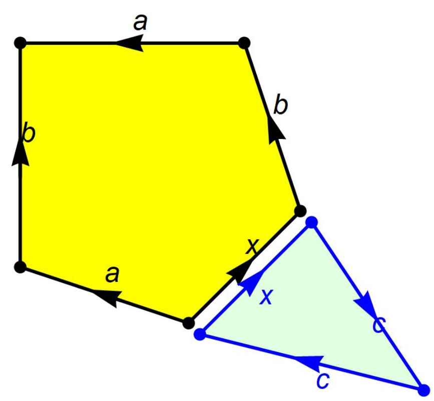
Each summand represents (taken alone) a surface minus a disk. The valid pasting operation identifies their boundary circles x, so they 'stick together' as a single connected surface. Lemma J is a way of asserting that it does not matter from which parts of the respective surfaces the disks are removed.

Examples. (i) One can interpret the word W_1xx^{-1} as the connected sum $\mathscr{M}_1\#S$ of $\mathscr{M}_1$ with a sphere, which is of course homeomorphic to $\mathscr{M}_1$.

(ii) Take $W_1 = aba^{-1}b^{-1}$ and $W_2 = cc$. A disk is removed from T and one from P, allowing the boundary circles to be glued to obtain $T\#P$. We now know (from p72 and p74) that

$$T\#P = P\#T = P\#P\#P = K\#P.$$

Here the equals sign denotes 'same homeomorphism class'. See below.



For homeomorphism classes of surfaces, the connected sum operation $\#$ is associative (since$(W_1W_2)W_3 = W_1(W_2W_3)$). It is commutative (since $W_1W_2 \sim W_2W_1$). Note that xx^{-1} acts as the identity (since $Wxx^{-1} \sim W$, and we are attaching a sphere).

Provided that W_1 has no single letters (so $\mathscr{M}_1$ has no boundary), it is true that $W_1x^{-1} \sim W_1ux^{-1}u^{-1}$, since adding the cuff does not affect the topology. This can be regarded as a extension of Lemma J. Similarly for $xW_2 \sim vxv^{-1}W_2$, hence

$$W_1W_2 \;\sim\; W_1ux^{-1}u^{-1} + vxv^{-1}W_2 \;\sim\; u^{-1}W_1uv^{-1}W_1v \;\sim\; W_1wW_2w^{-1},$$

where $w = uv^{-1}$. The last word represents a surface in which (if in space) $\mathscr{M}_1, \mathscr{M}_2$ are joined by a tube. The image shows two ways of visualizing $T\#T$:

8.3.3 Euler characteristicof a connected sum

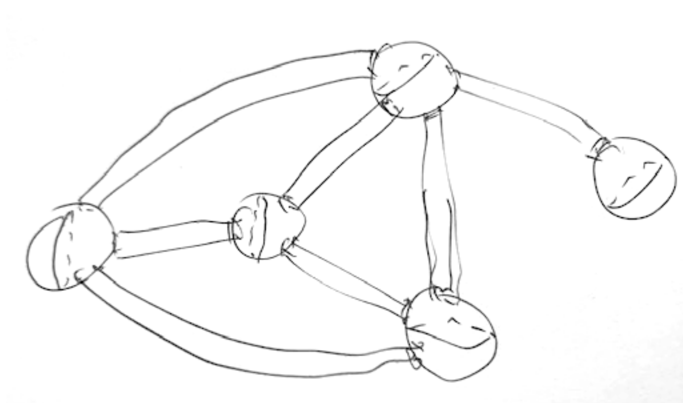
Proposition. $\chi(\mathscr{M}_1 \# \mathscr{M}_2) = \chi(\mathscr{M}_1) + \chi(\mathscr{M}_2) - 2.$

Proof. Compute χ by counting E, V, F arising from W_1W_2, or by joining $\mathscr{M}_1, \mathscr{M}_2$ across a triangle whose interior is removed. See q5 on Sheet 8. $\square$

Example. Take a connected graph G in space with v vertices and e edges. Place a sphere at each vertex, and attach a tube between spheres for each edge, so as to form an oriented surface $\mathscr{M}$ of genus g. Unless the graph has no cycles (so is a tree), $\mathscr{M}$ cannot be a connected sum of spheres, since we will be adding a handle from a surface to itself at some point. Nevertheless, we can apply the proposition to deduce that

$$2 - 2g = \chi = 2v - 2e \quad \Rightarrow \quad g = 1 - v + e.$$

If G is a planar graph, and is drawn in the plane with $f - 1$ internal regions, then it is clear that $\mathscr{M}$ is a sphere with $f - 1$ handles, which is consistent with Euler's formula $v - e + f = 2$. Here $g = 3$:



9 Paths and loops

9.1 Paths in topological spaces

From now on, X will denote a topological space. Here are some examples:

- Euclidean space $\mathbb{R}^n$, including the real line, the plane, and ordinary space ($n = 1, 2, 3$). The Argand plane $\mathbb{C}$, which (unlike $\mathbb{R}^2$) is 'pre-oriented'.
- The n-dimensional sphere $S^n = \{\mathbf{v} \in \mathbb{R}^{n+1} : \|\mathbf{v}\| = 1\}$, which is compact. Basic topology asserts that S^n minus a point is homeomorphic to $\mathbb{R}^n$.
- The n-dimensional torus T^n, which is homeomorphic to the **product** $S^1 \times \cdots \times S^1$ of n circles, and to the **quotient** (group and top space) $\mathbb{R}^n/\mathbb{Z}^n$.
  The 3-torus T^3 is obtained from a solid cube by identifying opssoite sides.
- Any graph in $\mathbb{R}^2$ or $\mathbb{R}^3$, including the figure-eight $\infty \approx q(\partial\mathscr{P})$ arising from the boundary code $aba^{-1}b^{-1}$ of the square $\mathscr{P}$ representing T.
- Any surface, in particular an orientable $\mathscr{M}_{Y,g,r}$, or a non-orientable $\mathscr{M}_{N,h,r}$.
- More general quotients of polygons like 'xxx' that are not locally Euclidean.
- The **complement** of a knot (this 3-manifold is an open subset of $\mathbb{R}^3$ or S^3).

9.1.1 Joining paths

Let X be a topological space, $x_0, x_1 \in X$. Let $\mathbb{I} = [0, 1]$ (to save space).

Definition. A **path** from x_0 to x_1 is a continuous map $\alpha\colon \mathbb{I} \to X$ such that $\alpha(0) = x_0$ and $\alpha(1) = x_1$. There is no requirement that α be injective. If $x_0 = x_1$ it is called a **loop** based at x_0. Once x_0 is fixed, $s \mapsto x_0$ $(\forall s)$ defines the **constant loop** ε at x_0.

Assume that X is **path-connected**, i.e. any two points of X can be joined by a path. This is implies that X is **connected**; these two properties are equivalent for surfaces.

If α is a path from x_0 to x_1, and β is a path from x_1 to x_2, then $\alpha\beta$ will denote the **concatenated path**, defined by traversing α then β both at double speed:

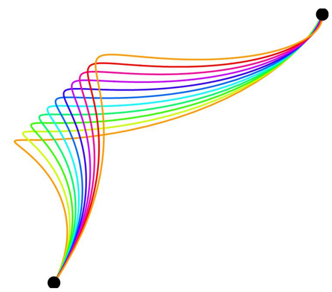
$$s \mapsto \begin{cases} \alpha(2s) & \text{if } s \in [0, \frac{1}{2}] \\ \beta(2s-1) & \text{if } s \in [\frac{1}{2}, 1]. \end{cases}$$

This is continuous: its image has no 'gap'.

We can define $\alpha\beta\gamma$ as $(\alpha\beta)\gamma$, or else divide into thirds (it won't really matter, see p91). We also define a 'backwards path' α^{-1} by $\alpha^{-1}(s) = \alpha(1-s)$.

9.1.2 Homotopy of paths

Definition. Fix $x_0, x_1 \in X$. Two paths α, β from x_0 to x_1 are **path-homotopic** if there exists a continuous mapping $H\colon \mathbb{I}^2 \to X$ such that



$$H(s,t) = \begin{cases} \alpha(s) & \text{if } t = 0 \\ \beta(s) & \text{if } t = 1, \end{cases} \quad \text{and} \quad \begin{cases} x_0 & \text{if } s = 0 \\ x_1 & \text{if } s = 1. \end{cases}$$

In this module, we shall write $\alpha \approxeq \beta$ or $\alpha \underset{H}{\approxeq} \beta$.

The map H represents a deformation of the paths fixing their endpoints, and it helps to write $H(s,t) = H_t(s)$ so $H_0 = \alpha$ and $H_1 = \beta$.

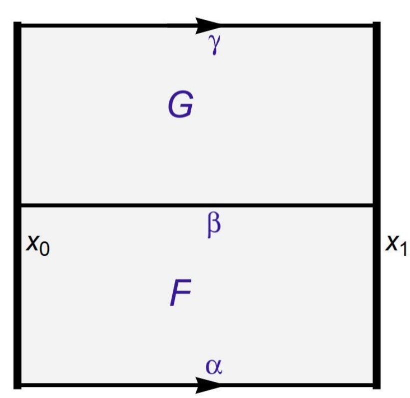
Ordinary homotopy ($\alpha \simeq \beta$) would only insist on the left-hand equations, and then *any* path starting from x_0 would be homotopic to ε, by setting $H(s,t) = \alpha\big((1-t)s\big)$.

Lemma 0. $\approxeq$ is an equivalence relation on the set of all paths from x_0 to x_1.

The proof is next. In future, we shall denote such an equivalence class by $[\alpha]$.

9.1.3 Patching homotopies

Proof of Lemma 0.
To see $\alpha \approxeq \alpha$, take $F(s,t) = \alpha(s)\ \forall t$.
To see that $\alpha \underset{F}{\approxeq} \beta$ implies $\beta \approxeq \alpha$, take



$$G(s,t) = F(s, 1-t).$$

To see that $\alpha \underset{F}{\approxeq} \beta$ and $\beta \underset{G}{\approxeq} \gamma$ implies $\alpha \approxeq \gamma$, take

$$H(s,t) = \begin{cases} F(s, 2t) & \text{if } t \in [0, \frac{1}{2}] \\ G(s, 2t-1) & \text{if } t \in [\frac{1}{2}, 1], \end{cases}$$

so this time speed is doubled vertically.

To see that H is continuous, it suffices to show that if C is a closed subset of X then then $H^{-1}(C)$ is a closed subset of $\mathbb{I}$. Let $A = \mathbb{I} \times [0, \frac{1}{2}]$ and $B = \mathbb{I} \times [\frac{1}{2}, 1]$, noting that the squashed versions $\tilde{F}, \tilde{G}$ agree on $A \cap B$ (where $t = \frac{1}{2}$). Then

$$H^{-1}(C) = (\tilde{F}^{-1}(C) \cap A) \cup (\tilde{G}^{-1}(C) \cap B)$$

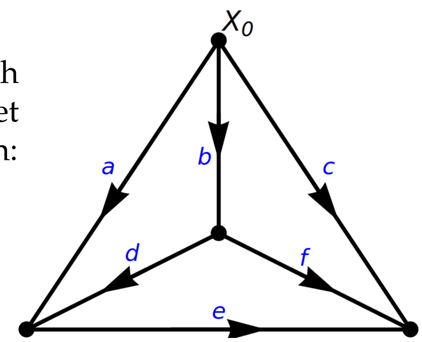
is the union of two closed subsets of $\mathbb{I}^2$. □

9.1.4 Loops in a graph

Any graph can be regarded as a subset of $\mathbb{R}^3$, so that edges only meet at vertices.

The diagram shows that K_4 (the complete tetrahedral graph with 4 vertices) is planar, since it can be embedded as a subset of $\mathbb{R}^2$. Each face defines a loop based at x_0 by concatenation:

$$\alpha = bda^{-1}, \qquad \beta = aec^{-1}$$
$$\gamma = cf^{-1}b^{-1}, \qquad \delta = bdef^{-1}b^{-1}.$$



Note that δ resembles insertion of a cuff at the vertex x_0 of a polygon. Moreover, $\alpha a = bda^{-1}a \approxeq bd$. This path-homotopy will be proved in §9.2.1, but matches our treatment of perimeter codes. Using similar equivalences,

$$\delta \approxeq (\alpha a)(a^{-1}\beta c)(c^{-1}\gamma b)b^{-1} \approxeq \alpha\beta\gamma,$$

We shall show that this path-homotopy of loops can be interpreted as a relation in the so-called fundamental group. The latter is generated by the path-homotopy classes of loops based at a fixed point of a topological space (like x_0 in K_4).

9.2 The fundamental group

Fix a 'basepoint' $x_0 \in X$. Recall that a **loop** at x_0 is a path from x_0 to x_0. For example, the constant loop ε.

Definition. $\boldsymbol{\pi_1}(X, x_0) = \{[\alpha] : \alpha \text{ is a loop based at } x_0\}$. The product of two elements of $\boldsymbol{\pi_1}(X, x_0)$ is given by $[\alpha][\beta] = [\alpha\beta]$.

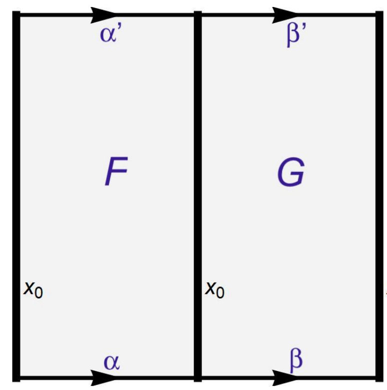
We shall show that this multiplication is well defined (Lemma 1), and makes $\boldsymbol{\pi_1}(X, x_0)$ into a group with identity element $e = [\varepsilon]$, and inverses $[\alpha]^{-1} = [\alpha^{-1}]$ (Lemmas 2,3,4).

Lemma 1. If $\alpha \underset{F}{\approxeq} \alpha'$ and $\beta \underset{G}{\approxeq} \beta'$ then $\alpha\beta \approxeq \alpha'\beta'$.

Proof. Take

$$H(s,t) = \begin{cases} F(2s,t) & \text{if } s \in [0,\frac{1}{2}] \\ G(2s-1,t) & \text{if } s \in [\frac{1}{2},1], \end{cases}$$

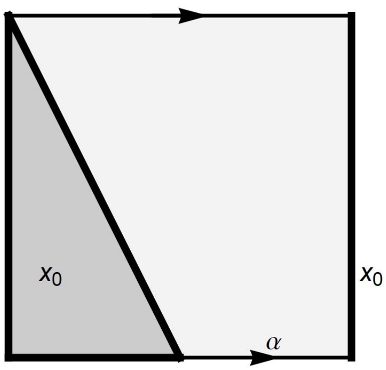
noting that $F(1,t) = x_0 = G(1,t)$. The map H is therefore continuous, as in the proof of Lemma 0. It gives the required path-homotopy $\alpha\beta \approxeq \alpha'\beta'$. □



9.2.1 Group axioms

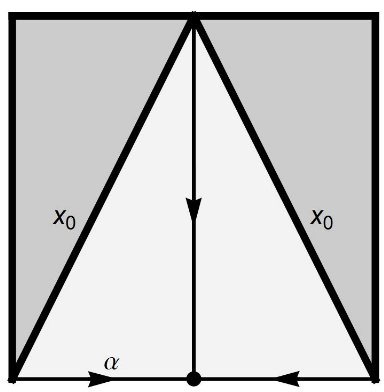
Lemma 2. $\varepsilon\alpha \approxeq \alpha \approxeq \alpha\varepsilon$.

Proof of the first path-homotopy:



Lemma 3. $\alpha\alpha^{-1} \approxeq \varepsilon \approxeq \alpha^{-1}\alpha$.

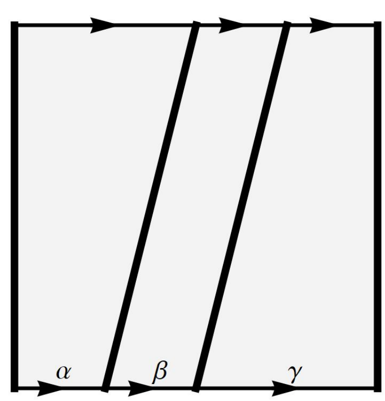
Proof of the first path-homotopy:



Lemma 4. $(\alpha\beta)\gamma \underset{H}{\approxeq} \alpha(\beta\gamma)$.

Proof.

$$H(s,t) = \begin{cases} \alpha(\frac{4s}{1+t}) & \text{if } s \leqslant \frac{1}{4}(1+t) \\ \beta(4s-1-t) & \text{if } \frac{1}{4}(1+t) \leqslant s \leqslant \frac{1}{4}(2+t) \\ \gamma(\frac{4s-2-t}{2-t}) & \text{if } \frac{1}{4}(2+t) \leqslant s. \end{cases}$$



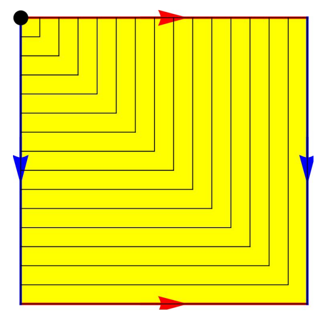
9.2.2 The Clifford torus

We are used to thinking of the torus as embedded as the surface of a doughnut in $\mathbb{R}^3$, but it is more naturally the subset

$$\{(\cos(2\pi s),\ \sin(2\pi s),\ \cos(2\pi t),\ \sin(2\pi t)) : (s,t) \in \mathbb{I}^2\} \subset S^3 \subset \mathbb{R}^4$$

(a realization attributed to W.K.Clifford, who arrived at KCL from Exeter in 1860).

As an abstract topological space, we know that $T = \widehat{\mathbb{I}^2}$ is the quotient of the unit square by means of the word $aba^{-1}b^{-1}$. With this description, a and b define loops in T based at the unique vertex $x_0 \in T$.



The picture describes a homotopy $aba^{-1}b^{-1} \approxeq \varepsilon$ (there's an animation of this homotopy on a torus of revolution on Keats). Thus $[a][b][a]^{-1}[b]^{-1} = e$, and

$$[a][b] = [b][a]$$

in $\boldsymbol{\pi_1}(T, x_0)$. The latter is an **abelian group** isomorphic to $\mathbb{Z}^2 = \mathbb{Z} \times \mathbb{Z}$, and T is a quotient **group**

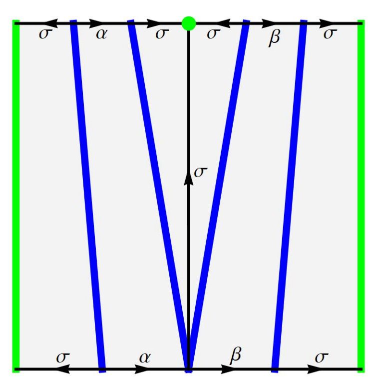
$$\mathbb{R}^2/\mathbb{Z}^2 \cong (\mathbb{R}/\mathbb{Z}) \times (\mathbb{R}/\mathbb{Z}) \cong S^1 \times S^1.$$

9.3 Properties of $\boldsymbol{\pi_1}$

Suppose that X is a path-connected topological space. Let $x_0, x_1 \in X$, and choose a path σ from x_0 to x_1.

Define a map $\phi_\sigma : \boldsymbol{\pi_1}(X, x_0) \to \boldsymbol{\pi_1}(X, x_1)$ by

$$\phi_\sigma[\alpha] = [\sigma^{-1}\alpha\sigma].$$



It is well defined:

Lemma 5. $\alpha \approxeq \alpha'$ implies that $\sigma^{-1}\alpha\sigma \approxeq \sigma^{-1}\alpha'\sigma$.

The map is also a group homomorphism:

Lemma 6. $\sigma^{-1}(\alpha\beta)\sigma \approxeq (\sigma^{-1}\alpha\sigma)(\sigma^{-1}\beta\sigma)$.

Proof. See diagram right →

Lemma 7. ϕ_σ is bijective.

Proof omitted.

It follows that $\boldsymbol{\pi_1}(X, x_0) \cong \boldsymbol{\pi_1}(X, x_1)$ as abstract groups, so we can speak of 'the' fundamental group $\boldsymbol{\pi_1}(X)$.

9.3.1 Functoriality

Suppose that X, Y are topological spaces, and $f\colon X \to Y$ is continuous.

Definition. The **induced mapping** on fundamental groups is

$$f_*\colon \boldsymbol{\pi_1}(X, x_0) \longrightarrow \boldsymbol{\pi_1}(Y, y_0), \quad \text{where } y_0 = f(x_0),$$

given by $f_*[\alpha] = [f \circ \alpha]$.

f_* is well defined, because we can compose a homotopy $H_t\colon \alpha \approxeq \alpha'$ by appling f to it.

Lemma 8. f_* is a group homomorphism.

Proof. If α, β are loops in X based at x_0 then

$$f_*([\alpha][\beta]) = f_*([\alpha\beta]) = [f \circ (\alpha\beta)] = [(f \circ \alpha)(f \circ \beta)] = (f_*[\alpha])(f_*[\beta]).$$

It is equally obvious that $1_* = 1$ and $(g \circ f)_* = g_* \circ f_*$.

If f is a homeomorphism, we can take $g = f^{-1}$ to deduce that f_* is an isomorphism of groups:

Corollary. If X is homeomorphic to Y (often written $X \approx Y$) then $\boldsymbol{\pi_1}(X) \cong \boldsymbol{\pi_1}(Y)$.

9.3.2 Simple spaces

Definition. A topological space X is **simply-connected** if it is path-connected and $\boldsymbol{\pi_1}(X) = \{e\}$ is the trivial group.

An obvious example is $\mathbb{R}^n$. Take x_0 to be the origin. Given a loop α based at x_0, we can define $H(s,t) = \alpha(s)t$ in order to 'shrink' α to a point. So $\alpha \approxeq \varepsilon$ and $[\alpha] = e$.

In fact, $\mathbb{R}^n$ has the much stronger property of being **contractible**, meaning that the whole space is homotopic to a point, i.e. there exists $x_0 \in X$ and a map $H\colon X \times I \to X$ such that

$$H(x,t) = \begin{cases} x & \text{if } t = 0 \\ x_0 & \text{if } t = 1. \end{cases}$$

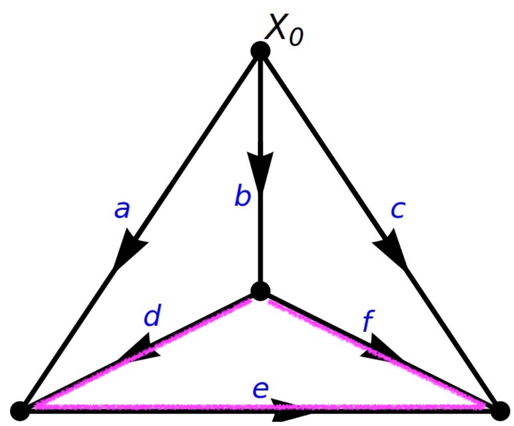
More examples: a star-shaped domain in $\mathbb{C}$, a tree (a connected graph with no cycles).

There are MANY spaces that are simply-connected but not contractible, such as S^n for $n \geqslant 2$. For simplicity, consider $S = S^2$; the open set $S^2 \backslash \{n\}$ (where $n = (0,0,1)$) is homeomorphic to $\mathbb{R}^2$ via stereographic projection, and any loop α (based at say s) can be deformed so as to avoid n, hence shrunk to a point.

9.3.3 Fundamental group of a graph

Fix the basepoint x_0 in K_4 as on p90. Recall the loops

$$\alpha = bda^{-1}, \qquad \beta = aec^{-1}$$
$$\gamma = cf^{-1}b^{-1}, \qquad \delta = bdef^{-1}b^{-1}.$$



Since pairs like cc^{-1} are path-homotopic to constant loops, we recover path-homtopy

$$\delta \ \approxeq\ (\alpha a)(a^{-1}\beta c)(c^{-1}\gamma b)b^{-1} \ \approxeq\ \alpha\beta\gamma.$$

This gives rise to a relation

$$[\delta] = [\alpha][\beta][\gamma]$$

in $G = \boldsymbol{\pi_1}(K_4, x_0)$. In fact, G is generated by any three of these loops, for example $[\alpha], [\beta], [\gamma]$, with *no* relations between them, and is the *free group* F_3.

The fundamental group of any graph is known to be isomorphic to the free group F_n on n generators, where $n = E - V + 1$ is the number of edges remaining when a spanning tree is removed (each such edge defining a loop). But first we must prove that

$$\boldsymbol{\pi_1}(S^1) \cong F_1 \quad (\mathbb{Z} \text{ in additive notation}),$$

the special case $V = E = n = 1$.

10 Covering maps and spaces

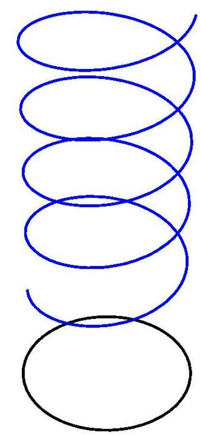
Given a topological space X with a non-trivial fundamental group, one can construct another space Z that resembles X is a precise sense, but has a simpler fundamental group.

10.1 Definitions and examples

10.1.1 Definition.

Let Z, X be topological spaces. Then

$$p\colon\ Z \longrightarrow X$$



is a **covering map** if each point $x \in X$ lies in some open set U for which $p^{-1}U$ is a disjoint union of open subsets V_i of Z such that $p|_{V_i}$ maps V_i homeomorphically onto U.

The V_i are 'sheets' of the covering. It follows that a subset W of X is open iff $p^{-1}(W)$ is open, so p is continuous and X is the **quotient** of Z by the equivalence relation

$$z \equiv z' \Leftrightarrow p(z) = p(z').$$

Examples.

(i) $p\colon \mathbb{R} \to S^1$ defined by $p(s) = e^{2\pi i s} = (\cos 2\pi s, \sin 2\pi s)$ $\nearrow$

(ii) $p\colon S^1 \to S^1$ with $p(z) = z^k$ where $z \in \mathbb{C}$, $|z| = 1$, and $k \in \mathbb{Z}$.

(iii) $p\colon S^n \to \widehat{S^n}$, where $\widehat{S^n}$ is the space of equivalence classes $\{\mathbf{v}, -\mathbf{v}\}$, so p is $2:1$. The quotient (denoted $\mathbb{RP}^n$) is called **real projective space**, and parametrizes straight lines through the origin in $\mathbb{R}^{n+1}$. $\mathbb{RP}^2$ is homeomorphic to the surface P with $\chi = 1$.

10.1.2 Lifting maps uniquely

Convention. From now on 'map' means 'continuous mapping'.

Definition. Given a covering map $p\colon Z \to X$ and a map $f\colon Y \to X$, a **lift** of f is any map $\widetilde{f}\colon Y \to Z$ such that

$$f = p \circ \widetilde{f}, \quad \text{so} \quad p(\widetilde{f}(y)) = f(y)\ \ \forall y \in Y.$$

Proposition. If Y is connected, any two lifts $\widetilde{f}, \widehat{f}$ that agree at one point are equal.

Proof. Suppose $\widetilde{f}(y_0) = \widehat{f}(y_0)$, and consider the subset

$$A = \{y \in Y : \widetilde{f}(y) = \widehat{f}(y)\} \ni y_0.$$

Let $y \in A$; then $\widetilde{f}(y) \in V_i$ for some i. Since $p\colon V_i \to U$ is a bijection, $\widetilde{f}, \widehat{f}$ must agree on the entire open set

$$\widetilde{f}^{-1}(V_i) = \widehat{f}^{-1}(V_i) = f^{-1}(U).$$

It follows that A is an **open** subset of Y. A similar argument (with $\widetilde{f}(y), \widehat{f}(y)$ in disjoint sheets) shows that the complement $Y \backslash A$ is also **open**. Since Y is connected, $A = Y$. $\square$

$$\begin{tikzcd} & Z \arrow[d, "p"] \\ Y \arrow[r] & X \end{tikzcd}$$

We shall apply this proposition to the case in which $Y = \mathbb{I}$ and $f = \alpha$ is a path or loop.

10.1.3 Lifting loops and homotopies

Theorem. Let $Y = \mathbb{I}$. Fix $x_0 \in X$ and $z \in p^{-1}(x_0)$. Let $\alpha, \beta\colon \mathbb{I} \to X$ be equivalent loops based at x_0, meaning $\alpha \approxeq \beta$.

(i) α has a unique lift $\widetilde{\alpha}_z\colon \mathbb{I} \to Z$ (a path, occasionally a loop) such that $\widetilde{\alpha}(0) = z$.

(ii) If $F\colon \mathbb{I}^2 \to X$ is the path-homotopy realizing $\alpha \approxeq \beta$, it has a unique lift $\widetilde{F}\colon \mathbb{I}^2 \to Z$ such that $\widetilde{F}_0 = \widetilde{\alpha}_z$ (the bottom edge of $\mathbb{I}^2$).

Proof. (i) One can partition $\mathbb{I}$ into closed intervals $[a_j, a_{j+1}]$ such that $\alpha([a_j, a_{j+1}])$ lies in an open set U covered by sheets. $\widetilde{\alpha}$ is defined on each in succession by bijectivity.

(ii) This follows from a similar argument by dividing the square $\mathbb{I}^2$ into rectangles R_{jk}, each mapped into a U, then working upwards from $\widetilde{F}_0$ to define $\widetilde{F}$ on each column. $\square$

$\widetilde{F}$ must be constant on the vertical lines $s = 0, 1$, so $\widetilde{F}(0, t) = z$ and $\widetilde{F}_1 = \widetilde{\beta}_z$. Thus,

$$\widetilde{\beta}_z(1) = \widetilde{F}(1, 1) = \widetilde{F}(1, 0) = \widetilde{\alpha}_z(1).$$

Corollary. The endpoint $\widetilde{\alpha}_z(1)$ of Z depends only on the class $[\alpha]$ of α in $\boldsymbol{\pi_1}(X, x_0)$, so fixing z determines a map $\boldsymbol{\pi_1}(X, x_0) \to p^{-1}(x_0)$.

10.2 Groups acting on sets

Let G be a group (finite or infinite, with identity element e), and Ω any set.

Definition. A **right action** of G on Ω is a mapping

$$\Omega \times G \longrightarrow \Omega, \quad \text{written} \quad (z, g) \mapsto z \cdot g,$$

such that (i) $z \cdot e = z$, and (ii) $z \cdot (gh) = (z \cdot g) \cdot h$ for all $z \in \Omega$ and $g, h \in G$.

If Ω is finite of size n, this simply means that there is a group homomorphism $G \to S_n$.

Examples. Let $\Omega = \mathbb{R}^3 = \{(x_1, x_2, x_3)\}$, $G = SO(3) = \{A \in \mathbb{R}^{3,3} : A^\top A = I,\ \det A = 1\}$. G acts on Ω on the right. It also acts on S^2 on the right. Let C be a cube centred at the origin. The sub*group* of G mapping C onto C is known to be isomorphic to S_4; it contains 2-cycles, 3-cycles, 4-cycles (how do each of these rotate the cube?).

Lemma. Fix $z \in \Omega$. Then $G_z = \{g \in G : z \cdot g = z\}$ is a subgroup of G, called the **stabilizer** of z, and $g \mapsto z \cdot g$ identifies the set $\{G_z g : g \in G\}$ of **right cosets** with Ω. Note that G acts on this set of cosets on the right, so we can dispense with Ω!

10.2.1 Action of $\boldsymbol{\pi_1}$ on a fibre

Let $p\colon Z \to X$ be a covering map, Fix $x_0 \in X$. Set

$$\begin{array}{rcl} G & = & \boldsymbol{\pi_1}(X, x_0) \\ \Omega & = & p^{-1}(x_0), \text{ the `fibre' over } x_0. \end{array}$$

Theorem. (i) Setting $z \cdot [\alpha] = \widetilde{\alpha}_z(1)$ defines a right action of G on Ω.
(ii) If $z \in \Omega$, the induced homomorphism $p_*\colon \boldsymbol{\pi_1}(Z, z) \to \boldsymbol{\pi_1}(X, x_0)$ is **injective**.
(iii) The stabilizer $\{[\alpha] : z \cdot [\alpha] = z\}$ is precisely the subgroup $p_*(\boldsymbol{\pi_1}(Z, z))$ of G.

Proofs. (i) It is obvious that $z \cdot [\varepsilon] = z$. To prove that $z \cdot ([\alpha][\beta]) = (z \cdot [\alpha]) \cdot [\beta]$, observe that $(\widetilde{\alpha\beta})_z = \widetilde{\alpha}_z \widetilde{\beta}_{z'}$ where $z' = \widetilde{\alpha}_z(1)$ (since $\widetilde{\alpha}_z \widetilde{\beta}_{z'}$ *is* a lift of $\alpha\beta$ from z).

(ii) If γ is a loop at z such that $p \circ \gamma \underset{F}{\approxeq} \varepsilon$, then the unique lift $\widetilde{F}$ gives $\gamma \approxeq \varepsilon_z$.

(iii) If $\widetilde{\alpha}_z(1) = z$, then $\widetilde{\alpha}_z$ is a loop and

$$[\alpha] = [p \circ \widetilde{\alpha}_z] = p_*[\widetilde{\alpha}_z].$$

Conversely, if $[\alpha] = p_*[\gamma]$ with γ a loop based at z, then $\gamma = \widetilde{\alpha}_z$ and $z \cdot [\gamma] = z$. $\square$

10.2.2 Fundamental group of the circle

Let $p\colon Z \to X$ be a covering map, fix $x_0 \in X$ and $z \in p^{-1}(x_0)$. Set $G = \boldsymbol{\pi_1}(X, x_0)$ and $\Omega = p^{-1}(x_0)$ as before.

Lemma. The map $G \to \Omega$ defined by $[\alpha] \mapsto z \cdot [\alpha]$ is
(i) surjective if Z is path-connected,
(ii) bijective if Z is simply-connected (meaning path-connected *and* $\boldsymbol{\pi_1}(Z) = \{e\}$).

Proof of surjectivity. Given $z' \in \Omega$, choose a path σ from z to z' in Z. Then $\alpha = p \circ \sigma$ is a loop based at x_0. By uniqueness, $\widetilde{\alpha}_z = \sigma$, so $\widetilde{\alpha}_z(1) = z'$.
Injectivity. If $z \cdot [\alpha] = z \cdot [\beta]$ then $z \cdot g = z$ where $g = [\alpha][\beta]^{-1}$. But the stabilizer of z is trivial by part (iii) of the previous Theorem, so $[\alpha] = [\beta]$. $\square$

We can apply the lemma to $X = S^1 \subset \mathbb{C}$, $x_0 = 1$, $Z = \mathbb{R}$, $p(z) = e^{2\pi i s}$. Then $\Omega = \mathbb{Z}$. The path $s \mapsto sn$ in $\mathbb{R}$ projects to the loop $\alpha_n \colon s \mapsto \exp(2\pi i s n)$ in S^1, and $0 \cdot [\alpha_n] = n$.

Corollary. $\boldsymbol{\pi_1}(S^1, 1)$ is isomorphic to $(\mathbb{Z}, +)$, i.e. the infinite cyclic group F_1.

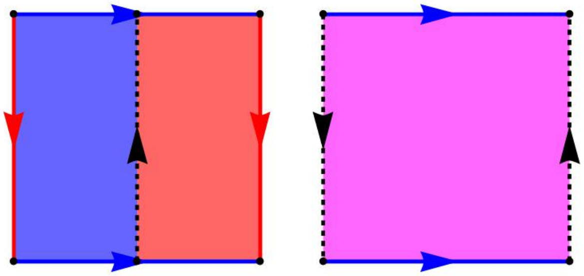
10.3 Double coverings

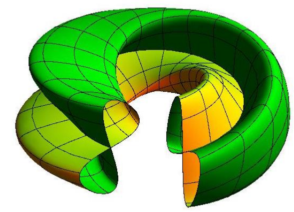
On p95, we defined $\mathbb{RP}^n$ to be the quotient of S^n obtained by identifying each point $\mathbf{v} = (x_0, \dots x_n)$ with $-\mathbf{v} = (-x_0, \dots, -x_n)$, its antipodal point.

Recall (see also p14) that $\mathbb{RP}^2$ is homeomorphic to the surface P originally defined from a square. The square model showed that P is also the quotient of a circular disk D in which opposite points on the boundary ∂D are identified. Now deform D into the southern hemisphere of S^2 and retain ∂D as its equator. Then any equivalence class $\{\mathbf{v}, -\mathbf{v}\}$ (defining a point of $\mathbb{RP}^2$) has a unique representative in P.

Let's explain why the group $SO(3)$ of rotations fixing the origin is homeomorphic to $\mathbb{RP}^3$. Any non-identity rotation is determined by an axis (i.e. a unit vector $\mathbf{v}$), and an angle $\theta \in [0, \pi]$. Map this to the point $(\theta/\pi)\mathbf{v}$ in $\mathbb{R}^3$. The only ambiguity is that an angle π around $\mathbf{v}$ is the same rotation as π around $-\mathbf{v}$. So we are identifying antipodal points of the boundary $S^2 = \partial B$ of the solid unit ball B. Since B is just a 3-dimensional analogue of D above, the resulting space is $\mathbb{RP}^3$ by the same argument.

10.3.1 Torus to Klein bottle in squares





Another immersion of K in $\mathbb{R}^3$, with twisted 8's

Represent the torus T by the square $\mathbb{I}^2$ with boundary $aba^{-1}b^{-1}$, giving $q_T \colon \mathbb{I}^2 \to T$. Represent the Klein bottle K by $\mathbb{I}^2$ with boundary $cd^{-1}c^{-1}d^{-1}$, giving $q_K \colon \mathbb{I}^2 \to K$.

Lemma. The mapping $p\colon T \to K$ defined by

$$p(q_T(s,t)) = \begin{cases} q_K(2s,t) & \text{if } s \leqslant \frac{1}{2} \\ q_K(2s-1,1-t) & \text{if } s \geqslant \frac{1}{2}. \end{cases}$$

is a 2:1 covering map.

If we regard a, c as loops in T, K respectively, then $p \circ a = cc$, whilst $p \circ b = d$.

We can expreess the mapping $T \to K$ in letters and words. Fix $z = q_T(0,1)$ as a basepoint $z \in T$, and $x_0 = q_K(0,1) \in K$. The interior of $q_K(\mathbb{I}^2)$ allows us to define a homotopy $\boxed{cd^{-1}c^{-1}d^{-1} \approxeq \varepsilon}$ with basepoint x_0.

Abusing notation to identify loops with classes in $\boldsymbol{\pi_1}(K, x_0)$, write $cd^{-1}c^{-1}d^{-1} = e$.

$$\begin{aligned} &\Rightarrow \quad cd^{-1} = dc \quad \text{and} \quad dc^{-1} = c^{-1}d^{-1} \\ &\Rightarrow \quad c^2d^{-1} = cdc = d^{-1}c^2. \end{aligned}$$

If we now set $a = c^2$ and $b = d^{-1}$ (justified because $p_* \colon \boldsymbol{\pi_1}(T, x_0) \to \boldsymbol{\pi_1}(K, x_0)$ is injective) we recover the relation $ab = ba$ or $aba^{-1}b^{-1} = e$ in $\boldsymbol{\pi_1}(T, z)$.

We shall see that $\boldsymbol{\pi_1}(K, x_0)$ is the group generated by c and b subject only to the relation $cbc^{-1}b = e$, induced by the boxed homotopy that shrinks the loop to a point. It follows that $\boldsymbol{\pi_1}(T, z)$ is isomorphic (via p_*) to the abelian subgroup generated by $a = c^2$ and b. Each element of $\boldsymbol{\pi_1}(T, z)$ can be written uniquely as $a^m b^n$ for some $m, n \in \mathbb{Z}$. Converting to additive notation,

$$\boldsymbol{\pi_1}(T, x_0) \cong \mathbb{Z} \oplus \mathbb{Z} \cong \mathbb{Z}^2.$$

10.3.2 Summary

Topologically, the 2-torus $T = T^2$ is the product $S^1 \times S^1$, and there is a 'combined' covering map $\mathbb{R}^2 \to T$ defined by

$$(s, t) \longmapsto (e^{2\pi i s},\ e^{2\pi i t}).$$

A theorem about $\boldsymbol{\pi_1}$ of a product of spaces confirms the isomorphism $\boldsymbol{\pi_1}(T) \cong \mathbb{Z} \times \mathbb{Z}$.

Covering spaces	Fundamental groups

$$\begin{tikzcd}
\mathbb{R}^2 \arrow[d] & & \{e\} \arrow[d] & \\
T \arrow[d, "2{:}1"'] & & \boldsymbol{\pi_1}(T, z) \arrow[d, "\text{index } 2"'] & = \langle c^2, b : c^2 b = b c^2 \rangle \\
K & & \boldsymbol{\pi_1}(K, x_0) & = \langle c, b : cb = b^{-1} c \rangle
\end{tikzcd}$$

11 Geometric groups

11.1 Free groups

The free group on one letter a is the infinite cyclic group

$$F_1 = \langle a\rangle = \{a^n : n \in \mathbb{Z}\} \;\cong\; (\mathbb{Z}, +),$$

generated by a (including a^{-1}). The element a is not subject to any relation other than the axiom $aa^{-1} = e = a^{-1}a$, and the rule $a^m a^n = a^{m+n}$ (for $m, n \in \mathbb{Z}$) that comes about from the meaning assigned to a^m for $m > 0$, $m = 0$ and $m < 0$.

Now let G, H be groups, written multiplicatively. The **free product** $G * H$ is a group whose elements consist of finite words made up by combining elements from either group (like $g_1h_1g_2h_2\cdots$) in which identity elements are suppressed (or retained as a combined identity if there are no other letters). The same idea allows us to construct a free group on several letters.

Example. The free group on two letters is

$$F_2 = F_1 * F_1 = \{a^{m_1}b^{n_1}a^{m_2}b^{n_2}\cdots a^{m_k}b^{n_k}\cdots : m_i, n_j \in \mathbb{Z},\ k \geqslant 0\}.$$

The usual product $F_1 \times F_1$ (or $\mathbb{Z}^2 = \mathbb{Z}\oplus\mathbb{Z}$ in additive notation) is **not** free; its generators a, b are subject to the relation $ab = ba$ or $aba^{-1}b^{-1} = e$, making it abelian.

11.1.1 Group presentations

An element $aba^{-1}b^{-1}$ of a group called the **commutator** of a, b and is sometimes denoted $[a, b]$ (though we shall use this notation sparingly). We can write

$$F_1 \times F_1 = \left\langle\, a, b \mid aba^{-1}b^{-1} = e \,\right\rangle, \quad \text{or more succinctly } \left\langle\, a, b \mid aba^{-1}b^{-1} \,\right\rangle,$$

in which a, b are the **generators**, $aba^{-1}b^{-1} = e$ is a **relation** and $aba^{-1}b^{-1}$ is a **relator**. The latter spawns other relators by taking its inverse and conjugates:

$$[b, a] = [a, b]^{-1}, \quad [a, b^{-1}] = b^{-1}[b, a]b, \quad \text{etc.}$$

Example. The dihedral group D_3 of symmetries of $\triangle$ is also the symmetric group S_3 of permutations of $\{1, 2, 3\}$. It is generated (for example) by a 3-cycle $a = (123)$ and 2-cycle $b = (12)$, subject to the equation $bab^{-1} = a^2$ or $baba = e$. Thus,

$$S_3 \cong D_3 = \left\langle a, b \mid a^3,\ b^2,\ abab\right\rangle.$$

Any group G can be characterized by a number of generators (letters) $a_1, \ldots, a_n$ and a number of relators (words) $R_1, \ldots, R_m$, giving rise to a so-called 'presentation'

$$G = \left\langle a_1, \ldots, a_n \mid R_1, \ldots, R_m\right\rangle.$$

11.1.2 Normal subgroups

The group $G = \langle a_1, \ldots, a_n \mid R_1, \ldots, R_m \rangle$ is then the quotient of a free group F by a homomorphism whose kernel N is the smallest **normal** subgroup containing the R_i. (Recall that a subgroup is normal if it contains all its conjugates, and that $G \cong F/N$.) Any element of N can be simplified to the identity, using the relations.

Theorems [Nielsen-Schreier, proofs use covering spaces of graphs].
(1) Any subgroup of a free group is free, but in general the R_i will not generate N.
(2) If F_n/N is finite of size i then $N \cong F_{i(n-1)+1}$.

Back to the examples. (i) For S_3, take $j = 3$ and $R_1 = a^3$, $R_2 = b^2$, $R_3 = abab$. Then $F_2/N \cong S_3$, so N is a subgroup of index $i = 6$ in F_2, and $N \cong F_7$. A graph-theoretic algorithm implies that

$$N = \left\langle a^3\,, b^2,\ abab,\ b^{-1}a^3b,\ ab^{-1}ab,\ a^{-1}ba^{-1}b,\ b^{-1}aba \right\rangle .$$

(ii) The kernel of the homomorphism $F_2 \to F_1 \times F_1$ is the normal subgroup generated by *all* the commutators $xyx^{-1}y^{-1}$ with $x, y \in F_2$, and $F_1 \times F_1 \cong F_2/[F_2, F_2]$ is the **abelianization** of F_2.

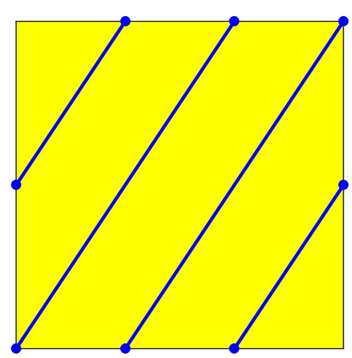
11.1.3 Simplifying relations

Let K be a (left or right) trefoil knot in $\mathbb{R}^3$. The so-called Wirtinger presentation of $G = \boldsymbol{\pi_1}(\mathbb{R}^3 \backslash K)$ is given by

$$G = \left\langle \alpha, \beta, \gamma \mid \alpha^{-1}\beta\alpha\gamma^{-1},\ \beta^{-1}\gamma\beta\alpha^{-1},\ \gamma^{-1}\alpha\gamma\beta^{-1} \right\rangle ,$$

with one relation for each crossing. This is explained on p108, but we are only concerned with the group theory here.

In G, we have $\alpha^{-1}\beta\alpha = \gamma = \beta\alpha\beta^{-1}$, which implies the third equation $\gamma^{-1}\alpha\gamma = \beta$. So we can dispense with γ, and observe that



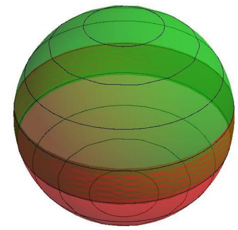
$$\alpha^{-1}\beta\alpha = \beta\alpha\beta^{-1} \ \Rightarrow\ \alpha\beta\alpha = \beta\alpha\beta.$$

Set $a = \alpha\beta$ and $b = \alpha\beta\alpha$. Then $a^3 = b^2$. Since $\alpha = a^{-1}b$ and $\beta = b^{-1}a^2$,

$$G = \left\langle a, b \mid a^3 b^{-2} \right\rangle .$$

This presentation reflects the realization of a trefoil knot wrapping around a torus.

11.2 The van Kampen theorem



We shall state versions of this result without proof.

Recall that $\boldsymbol{\pi_1}(X, x_0)$ denotes the fundamental group of a topological space X, consisting of equivalence classes of loops based at x_0 (under $\approxeq$). If X is path-connected, we saw on p92 that it is independent of x_0 (up to isomorphism).

The vK theorem will allow us to compute $\boldsymbol{\pi_1}$ of a space X which is the union of two **open** subsets U, V with **path-connected** intersection $U \cap V$ (the sphere S^2 being a case in point). The inclusions i, j, I, J give rise to induced homomorphisms i_*, j_*, I_*, J_* by composition $i_* \colon [\alpha] \mapsto [i \circ \alpha]$ etc:

$$\begin{tikzcd} & U \arrow[dr, "I"] & \\ x_0 \in U \cap V \arrow[ur, "i"] \arrow[dr, "j"'] & & X \\ & V \arrow[ur, "J"'] & \end{tikzcd} \qquad \begin{tikzcd} & \boldsymbol{\pi_1}(U) \arrow[dr, "I_*"] & \\ \boldsymbol{\pi_1}(U \cap V) \arrow[ur, "i_*"] \arrow[dr, "j_*"'] & & \boldsymbol{\pi_1}(X) \\ & \boldsymbol{\pi_1}(V) \arrow[ur, "J_*"'] & \end{tikzcd}$$

We have chosen a basepoint $x_0 \in U \cap V$, and $\boldsymbol{\pi_1}(A)$ denotes $\boldsymbol{\pi_1}(A, x_0)$.

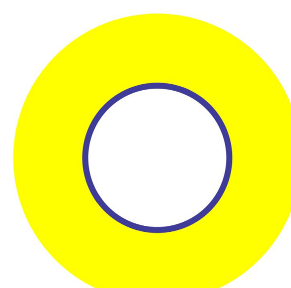
11.2.1 Homotopy of maps and spaces

This notion was introduced in q5 on Sheet 9. Let X, Y be path-connected spaces.

Definitions. (1) Two **maps** $f_0, f_1 \colon X \to Y$ are *homotopic* (written $f_0 \simeq f_1$) if there exists a map $H \colon X \times \mathbb{I} \to Y$ such that $H_0 = f_0$ and $H_1 = f_1$.

(2) The **spaces** X, Y are themselves called *homotopic* (written $X \simeq Y$) if there exist maps $f \colon X \to Y$ and $g \colon Y \to X$ such that $g \circ f \simeq 1_X$ and $f \circ g \simeq 1_Y$.

(3) If $X \subset Y$ then X is called a (strong) **deformation retract** of Y if f is the inclusion, $g \circ f = 1_X$, and $f \circ g \underset{H}{\simeq} 1_Y$ via a homotopy H such that $H_t(x) = x \;\; \forall x \in X$.



Example. S^1 is a deformation retract of $Y = \mathbb{C}^* = \mathbb{C} \backslash \{0\}$. To see this, take $g(z) = z/|z|$ and $H(z, t) = (1-t)z/|z| + tz$.

Proposition. If $X \simeq Y$ then $\boldsymbol{\pi_1}(X) \cong \boldsymbol{\pi_1}(Y)$.

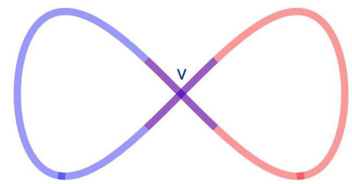
This is easy to prove in the case of a deformation retract, by composing loops with g. The proposition generalizes the corollary on p93.

11.2.2 First version

Theorem vK1. If $U \cap V$ is simply-connected, then $\boldsymbol{\pi_1}(X) \cong \boldsymbol{\pi_1}(U) * \boldsymbol{\pi_1}(V)$.

Example. In the description of a torus T as the quotient of a square $\mathscr{P}$, its boundary $\partial\mathscr{P}$ maps to the 'wedge' $S^1 \vee S^1$ of two circles, homeomorphic to the figure-eight shown. Call this space X. (See q4 on Sheet 8 for similar examples.)

The circles can be enlarged into open sets U, V with $U, V \simeq S^1$, and $U \cap V$ a 'cross'. This cross is homotopic to a point: the arcs can be continuously shortened until the figure has been collapsed to the central vertex v. This point is a deformation retract of the cross; one says that $U \cap V$ **retracts** to v.



Then vK1 and the previous proposition imply that

$$\boldsymbol{\pi_1}(S^1 \vee S^1) \cong \boldsymbol{\pi_1}(S^1) * \boldsymbol{\pi_1}(S^1) \cong F_1 * F_1 = F_2.$$

If we regard the figure-eight X as a graph, then $\{v\}$ is itself a minimal spanning tree (MST). It's a general result that the fundamental group of a graph is isomorphic to F_n where $n = E - (V-1)$ is the number of edges not in a MST (see also p94).

11.2.3 Second version

If $K = \boldsymbol{\pi_1}(U \cap V)$ is non-trivial, it will modify the free product $\boldsymbol{\pi_1}(U) * \boldsymbol{\pi_1}(V)$. The idea is that classes of loops in $U \cap V$ (which i_* and j_* map to $\boldsymbol{\pi_1}(U)$ and $\boldsymbol{\pi_1}(V)$) cannot count *twice* in computing $\boldsymbol{\pi_1}(U \cap V)$.

Theorem 2 [van Kampen, Seifert]. With U, V still open and $U \cap V$ path-connected, the fundamental group of $X = U \cup V$ is isomorphic to the **amalgamated product**

$$\boldsymbol{\pi_1}(U) *_K \boldsymbol{\pi_1}(V) = (\boldsymbol{\pi_1}(U) * \boldsymbol{\pi_1}(V))/N,$$

where N is the smallest normal subgroup containing all elements $i_*(k)^{-1} j_*(k)$, $k \in K$.

Example. Take $X = T$. Let o be the centre of the square $\mathscr{P}$, $\hat{o}$ its image in $T = \widehat{\mathscr{P}}$, and $U = T \backslash \{\hat{o}\}$. Then U retracts to the boundary $\widehat{\partial\mathscr{P}}$ of the unique face of T, and

$$U \simeq \widehat{\partial\mathscr{P}} \quad \Rightarrow \quad \boldsymbol{\pi_1}(U) \cong \boldsymbol{\pi_1}(\widehat{\partial\mathscr{P}}) \cong F_2.$$

Now take V to a small open disk (or square) containing $\hat{o}$. Then $U \cap V$ is homotopic to $\partial\mathscr{P}$ and a circle, and i_* maps the generator of $\boldsymbol{\pi_1}(U \cap V)$ to the class $aba^{-1}b^{-1}$ in

$\pi_1(U)$, whilst $\pi_1(V) = \{e\}$ is trivial. Therefore N is the smallest normal subgroup of $\pi_1(U) * \pi_1(V) \cong F_2$ containing $aba^{-1}b^{-1}$, and

$$\pi_1(T) \cong \left\langle a, b \mid aba^{-1}b^{-1} \right\rangle.$$

11.2.4 The fundamental groups of spheres

Theory in §10 led to a rigorous proof that $\pi_1(S^1) \cong \mathbb{Z}$. By contrast,

Theorem. S^n is simply connected (i.e. $\pi_1(S^n) \cong \{e\}$) if $n \geqslant 2$.

One proof exploits the fact that S^n minus a point is homeomorphic to $\mathbb{R}^n$ (see q2 on Sheet 9 for the case $n = 2$). Any loop in S^n can then be regarded as a loop in $\mathbb{R}^n$ based at the origin, and $H(s,t) = (1-t)\alpha(s)$ deforms it to the constant loop.

Proof using vK2. Fix $n \geqslant 2$. Then S^n is the union of 3/4-spheres

$$\begin{aligned} U &= \{(x_1, \ldots, x_{n+1} : \textstyle\sum x_i^2 = 1,\ x_{n+1} \geqslant -\frac{1}{2}\} \\ V &= \{(x_1, \ldots, x_{n+1} : \textstyle\sum x_i^2 = 1,\ x_{n+1} \leqslant \frac{1}{2}\}, \end{aligned}$$

each homeomorphic to a disk, and simply connected. The 'equator' $x_{n+1} = 0$ can be identified with S^{n-1}, and is path-connected for $n \geqslant 2$. By Theorem vK2, $\pi_1(S^n)$ is a quotient of the free group $F = \pi_1(U) * \pi_1(V) = \{e\}$, and therefore trivial. □

If we accept that S^2 is simply connected, we can deduce by induction that S^n is for $n \geqslant 3$, using only vK1. For the equator S^{n-1} above is a deformation retract of $U \cap V$. So now all of $U, V, U \cap V$ are simply-connected for $n \geqslant 3$.

11.2.5 The fundamental group of a surface of genus 3

The description that follows works for any genus, and an easy modification works for non-orientable surfaces without boundary We have chosen $g = 3$ merely as an aid tounderstanding.

Let $\mathscr{M}$ be an orientable surface of genus 3 without boundary. It can be constructed as the quotient of a 12-sided polygon $\mathscr{P}$, whose boundary code has the normal form

$$\mathbb{A}_3 = a_1 b_1 a_1^{-1} b_1^{-1} a_2 b_2 a_2^{-1} b_2^{-1} a_3 b_3 a_3^{-1} b_3^{-1} = [a_1, b_1][a_2, b_2][a_3, b_3].$$

Because all the vertices map to a unique point $x_0 \in \mathscr{M}$, each of the 12 edges of $\mathscr{P}$ (suitably parametrized) maps to a *loop* in $\mathscr{M}$, and an element in $\pi_1(\mathscr{M}, x_0)$.

To determine $\boldsymbol{\pi_1}(\mathscr{M}, x_0)$, we repeat the argument with $\mathscr{M}$ in place of T on p104, but in more detail motivated by the artistic image on the next page of $\mathscr{P}$, which q maps onto $\mathscr{M}$. The proof divides into the folowing steps:

(i) Take U to be $\mathscr{M}$ minus the image of a small disk (the crimson dot) around the centre of the 12-gon $\mathscr{P}$, and take V to be the image of a larger disk (dark blue union crison) centred at $\hat{o}$. Then U can be deformed to the image of the boundary of $\mathscr{P}$ (represented by the outer cyan ring), and is homotopic to a wedge of 6 circles, whereas $U \cap V$ is an annulus homotopic to S^1:

$$U \simeq \bigvee\nolimits_6 S^1, \qquad V \simeq \{\hat{o}\}, \qquad U \cap V \simeq S^1.$$

(ii) It follows that $\boldsymbol{\pi_1}(U)$ is a free group on 6 letters $a_1, a_2, a_3, b_1, b_2, b_3$, which we shall now (by abuse of notation) regard as path-homotopy classes (so a_1 really stands for $[a_1]$ etc). On the other hand, $\boldsymbol{\pi_1}(V)$ is trivial, and $\boldsymbol{\pi_1}(U \cap V) \cong \mathbb{Z}$.

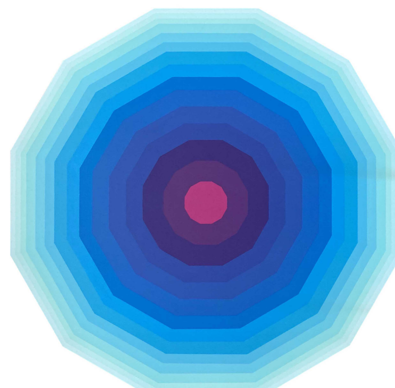
(iii) The homomorpism $i_*\colon \boldsymbol{\pi_1}(U \cap V) \to \boldsymbol{\pi_1}(U)$ maps the (clockwise) generator of $K = \boldsymbol{\pi_1}(U \cap V)$ to $\mathbb{A}_3$, interpretated as a product of 12 elements of $\boldsymbol{\pi_1}(\mathscr{M}, x_0)$.

(iv) Theorem vK2 implies that $\boldsymbol{\pi_1}(\mathscr{M})$ is the quotient of the free product

$$\boldsymbol{\pi_1}(U) * \boldsymbol{\pi_1}(V) \cong \boldsymbol{\pi_1}(U)$$

by the normal subgroup N generated by $\mathbb{A}_3$. Thus

$$\begin{aligned}
\boldsymbol{\pi_1}(\mathscr{M}) &\cong \boldsymbol{\pi_1}(U) *_K \boldsymbol{\pi_1}(V) \\
&\cong F_6 *_K \{e\} \\
&\cong F_6/N \\
&\cong \langle a_1, \cdots, b_3 \mid [a_1, b_1][a_2, b_2][a_3, b_3] \rangle .
\end{aligned}$$



11.2.6 Conclusion

Recall our notation for normal forms for the boundary codes. Each letter defines a loop in $\mathscr{M}$ (as all vertices map to the same point, call it x_0), and we use the same letter to denote an element of $\boldsymbol{\pi_1}(\mathscr{M}, x_0)$. In this way, the code becomes a relator:

$$\begin{aligned}
\mathbb{A}_g &= (a_1 b_1 a_1^{-1} b_1^{-1}) \cdots (a_g b_g a_g^{-1} b_g), && g \geqslant 1 \\
\mathbb{C}_h &= (c_1 c_1) \cdots (c_h c_h), && h \geqslant 1 \\
\mathbb{D}_r &= (u_1 d_1 u_1^{-1}) \cdots (u_r d_r u_r^{-1}), && r \geqslant 1.
\end{aligned}$$

If $\mathscr{M}$ is an orientable surface of genus g and no boundary, then

$$\boldsymbol{\pi_1}(\mathscr{M}) \cong \left\langle a_1, \ldots, a_g, b_1, \ldots b_g \mid \mathbb{A}_g \right\rangle .$$

It is abelian iff $g = 0$ or $g = 1$, and $\mathscr{M}$ is the sphere $S = S^2$ or torus $T = T^2$.

If $\mathscr{N}$ is a non-orientable surface with $\chi = 2 - h$ and no boundary, then

$$\boldsymbol{\pi_1}(\mathscr{M}) \cong \left\langle c_1, \ldots, c_h \mid c_1^2 \cdots c_h^2 \right\rangle .$$

It is abelian (and finite) iff $h = 1$ and $\mathscr{N}$ is the projective plane P.

If $\mathscr{M}'$ is an orientable surface of genus g with (e.g.) $r = 1$ boundary component,

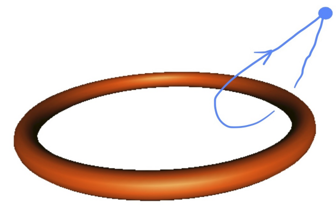
$$\boldsymbol{\pi_1}(\mathscr{M}) \cong \left\langle a_1, \ldots, a_g, b_1, \ldots b_g, d \mid \mathbb{A}_g d \right\rangle \cong \left\langle a_1, \ldots, a_g, b_1, \ldots b_g \right\rangle \cong F_{2g}.$$

11.2.7 Exhale I by Zarah Hussain

The image on p106 used to apply van Kampen's theorem is reproduced by courtesy of the artist. It was on display at the Royal Academy's summer exhibition in 2022, and could have been commissioned for the module, since its colour gradations capture the homotopies!

11.3 Knot complements

Let K be a knot in space. The aim of this final section is to describe the fundamental group $\boldsymbol{\pi_1}(\mathbb{R}^3 \backslash K)$ of the **knot complement**, which is a 3-manifold.



Let K be an unknot represented by $S^1 \subset \mathbb{R}^2 \subset \mathbb{R}^3$, and fix a basepoint x_0. An obvious element of $\boldsymbol{\pi_1}(\mathbb{R}^3 \backslash S^1, x_0)$ is $[\alpha_1]$ where α_1 wraps once around S^1. It is intuitively clear that

$$\boldsymbol{\pi_1}(\mathbb{R}^3 \backslash S^1, x_0) = \left\{[\alpha_1]^n : n \in \mathbb{Z}\right\} \cong F_1.$$

One can make this rigorous by first using Theorem vK1 to show that

$$\boldsymbol{\pi_1}(\mathbb{R}^3 \backslash S^1) \cong \boldsymbol{\pi_1}(S^3 \backslash S^1),$$

as outlined in q8 on Sheet 11. Then, $S^3 \backslash S^1$ can be retracted onto another circle (think Hopf link, and Clifford torus on p92). So $\boldsymbol{\pi_1}(S^3 \backslash S^1) \cong \boldsymbol{\pi_1}(S^1)$.

The following result that we cite is deeper:

Theorem. If K is a knot and $\boldsymbol{\pi_1}(\mathbb{R}^3 \backslash K) \cong F_1$ then K is ambient isotopic to an unknot.

11.3.1 Wirtinger relation

This is illustrated below, and relates arcs, loops, and crossings.

Motivated by the case of the unknot, Let K be a knot and 'flatten' it to resemble a diagram D, *above* which we place our basepoint x_0. Choose an orientation for D.

If D has c crossings, then it has c arcs (connected strands). Place loops $\alpha_1, \ldots, \alpha_c$ based at x_0 around each (think puppet strings), satisfying the right-hand rule.

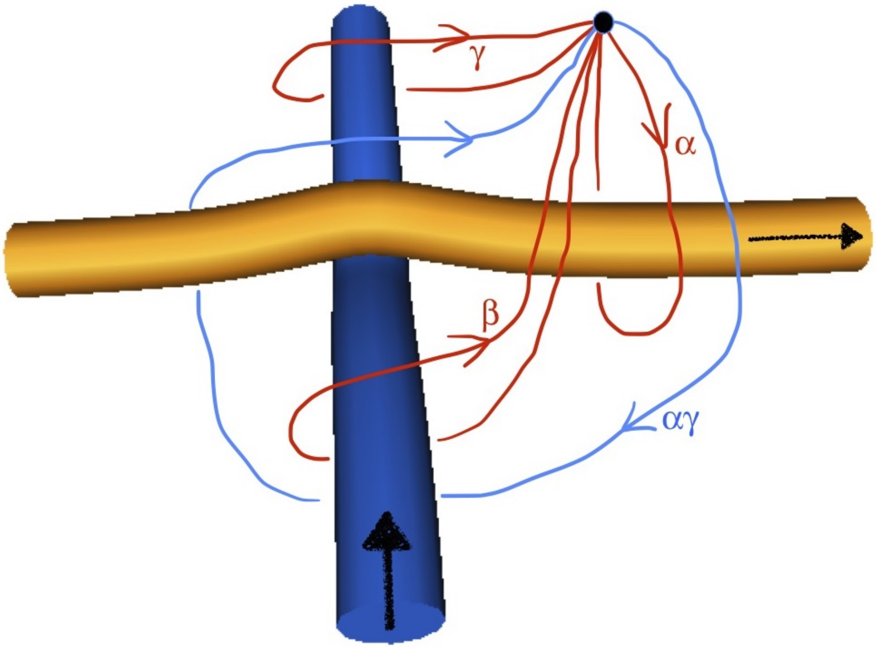
Suppose that at a particular crossing with positive writhe sign, $\alpha = \alpha_i$ overpasses $\beta = \alpha_j$ leading to $\gamma = \alpha_k$. Then the concatenations below show that

$$\alpha^{-1}\beta\alpha \approxeq \gamma \quad \text{or} \quad [\alpha_i]^{-1}[\alpha_j][\alpha_i] = [\alpha_k].$$

At a negative writhe crossing,

$$[\alpha_i][\alpha_j][\alpha_i]^{-1} = [\alpha_k].$$

The fundamental group $\boldsymbol{\pi_1}(\mathbb{R}^3 \backslash K)$ is then characterized by such 'conjugation' equations, one for each crossing. However, it can be shown that one relation is always redundant (as on p14), in an analogy with the colouring relations.



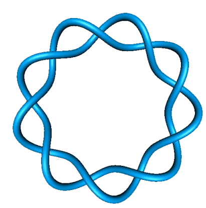
Joining loops in a way reminiscent of Reidemeister moves shows that

$$(\alpha\gamma)\alpha^{-1} \approxeq \beta.$$

11.3.2 Final remarks

Theorem vK2 can be used to prove that the **knot group** $\boldsymbol{\pi_1}(\mathbb{R}^3 \backslash K)$ is generated by the loops $\alpha_1, \ldots \alpha_c$ associated to the arcs of a knot diagram, subject to any $c-1$ of the crossing relations

$$[\alpha_i][\alpha_j][\alpha_i]^{-1} = [\alpha_k]$$



For a torus knot of type (p, q) (with p, q positive coprime integers), we have

$$\boldsymbol{\pi_1}(\mathbb{R}^3 \backslash K) \cong \left\langle a, b \mid a^p b^{-q} \right\rangle.$$

The nature of $\boldsymbol{\pi_1}(\mathbb{R}^3 \backslash K)$ permits one to construct finite covering spaces of $\mathbb{R}^3 \backslash K$ (for example, one based on the group S_3 if K is a trefoil so $(p, q) = (3, 2)$).

These ideas underlie work of William Thurston (1946–2012), which revolutionized the study of 3-manifolds.

If $K = 3_1$ then $S^3 \backslash K$ is $\mathsf{SL}(2, \mathbb{R})/\mathsf{SL}(2, \mathbb{Z})$, giving a link with the Lorenz attractor, mentioned in §0.3.

If $K = 4_1$ then $S^3 \backslash K$ can be obtained from two solid tetrahedra by gluing pairs of faces in a prescribed way (and removing the unique vertex in the quotient).