Introduction to Surfaces

Instructor: Nikhil Sahoo

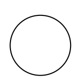
21 June 2019

1 Low-Dimensional Objects

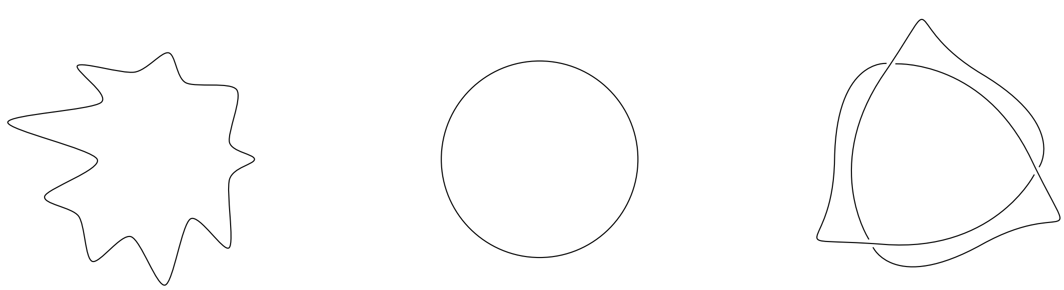
The smallest objects we may consider (the atoms of all other spaces) are just points. These are called "zero-dimensional," reflecting that there is no room for movement. Next (in dimension one) we have curves.

Definition. A curve is a space that locally looks like the real line $\mathbb{R}$. We also allow curves to have endpoints, at which it locally looks like the half-line $[0, \infty)$. This somewhat informal definition is best with examples.

Examples. We categorize these examples by whether they are bounded and whether they have endpoints.

	Bounded	Unbounded
Endpoints	$[0, 1]$	$[0, \infty)$
No Endpoints		$\mathbb{R}$

To classify all (connected) curves, we must first say what it means for two curves to be the same. We are interested in the intrinsic properties of these spaces, as opposed to how they sit in a larger space. Intuitively, this means we can continuously deform the curve, pass it through itself, or cut and paste (keeping careful track of the cuts we have made). For example, the following pictures all represent the same curve (a circle).



Although the last curve is knotted, we may pass it through itself to turn it into one of the other circles. Alternatively, we can cut the knot (making a line segment), untie it, then paste the endpoints back together.

Proposition. The only connected curves are the four described in the above table.

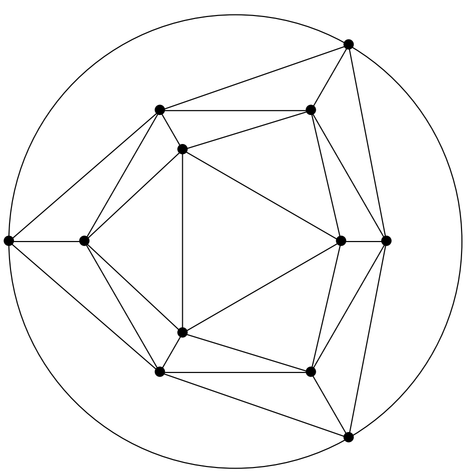
Proof. Pick a starting point on the curve. When we say the curve is connected, we mean that every point can be accessed by moving along the curve from this (or any) starting point. Since a curve locally looks like a line, there are only two directions to move from the starting point. We pick a direction and start moving that way. If we loop back to the starting point, then the curve is a circle. Otherwise, we must consider both directions from the starting point. In each direction, one of two things may happen: either we can keep going without bound (drawn as an arrowhead $\rightarrow$) or we get stopped by an endpoint (drawn as a bullet $\bullet$). The three possible combinations of unbounded/endpoint yield the three curves $[0,1]$, $[0,\infty)$, and $\mathbb{R}$. $\square$

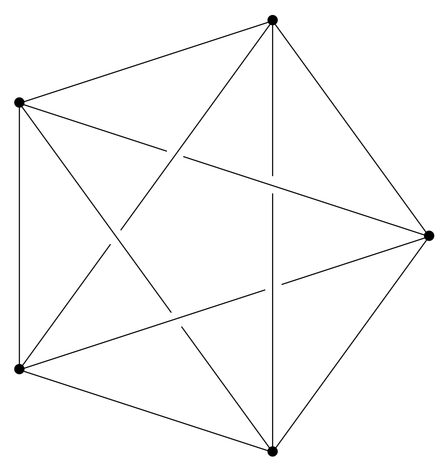
These are the only *connected* curves, but they are not the only curves. For example, ten lines and a circle form a perfectly acceptable curve, but it is impossible to "move" from one line to another without "jumping." In general, any curve is just a union of one or more connected curves.

Exercise 1. In terms of this classification, figure out which curve describes each type of conic section.

Our ultimate goal is to understand the classification of certain two-dimensional spaces. Before we dive in, there is one more useful type of space that we will consider, which has both zero- and one-dimensional parts.

Definition. A graph $G = (V, E)$ is a space built as follows. The set V consists of points, called vertices. The set E consists of line segments, called edges, connecting pairs of vertices (i.e. its endpoints are vertices). We require that no two edges intersect, except possibly at their endpoints.





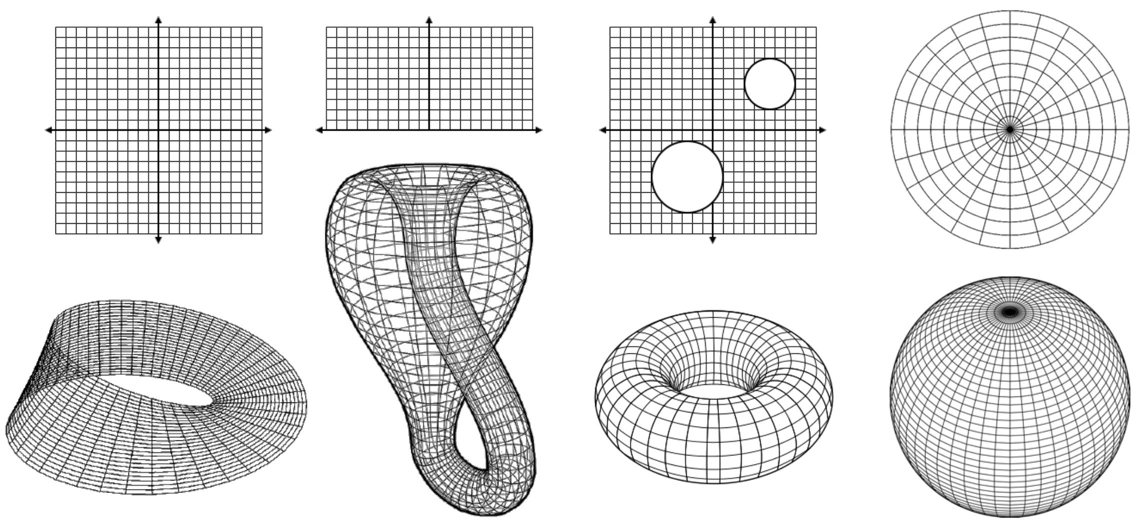
These two diagrams illustrate graphs. The right-hand graph cannot be drawn in the plane without crossings, but the space that this drawing represents does not have any intersections between the different edges.

Exercise 2: Explicitly show that every curve can be drawn as a graph. Is every graph a curve?

2 What is a Surface?

Definition. Just as a curve locally looks like $\mathbb{R}$, we define a surface to locally look like the plane $\mathbb{R}^2$. Analogous to endpoints, we allow the surface to possess a boundary, on which it locally looks like the half-plane $\mathbb{H} = \{(x, y) \in \mathbb{R}^2 : y \geq 0\}$. As with the curves above, this definition will be clarified by examples.

Examples. Illustrated below are several examples of surfaces, with and without boundary.



These images are from https://commons.wikimedia.org/wiki/File:SurfacesWithAndWithoutBoundary.svg and [6].

The first two surfaces are just the plane $\mathbb{R}^2$ and the half-plane $\mathbb{H}$. These are the surfaces after which all others are modeled, but they "go off to infinity" in a way that can be inconvenient (as does the plane with two disks removed). On the right, we have the two simplest surfaces: the disk $\mathbb{D}$ and the sphere $\mathbb{S}$. You may already be familiar with the Möbius band, pictured in the bottom left. There is also a surface that resembles a donut, which is called the torus and denoted $\mathbb{T}$. Lastly, there is a sort of "twisted version" of the torus, which is called the Klein bottle and denoted $\mathbb{K}$. Just as not all graphs can be drawn in the plane without some self-intersections, not all surfaces can be pictured in three-dimensional space without self-intersections. The Klein bottle $\mathbb{K}$ has this property. However, we shall soon see another way to represent such surfaces.

Proposition. For any surface S with a boundary, the boundary ∂S forms a curve (without endpoints).

Proof. At any point p in ∂S, the surface locally looks like the half-plane $\mathbb{H}$. But the boundary of $\mathbb{H}$ is just the real line $\mathbb{R}$, so ∂S looks like $\mathbb{R}$ near p. Thus ∂S locally looks like $\mathbb{R}$, so it is a curve without endpoints. □

Exercise 3. Which of the above eight surfaces have boundaries? For each that does, what is its boundary?

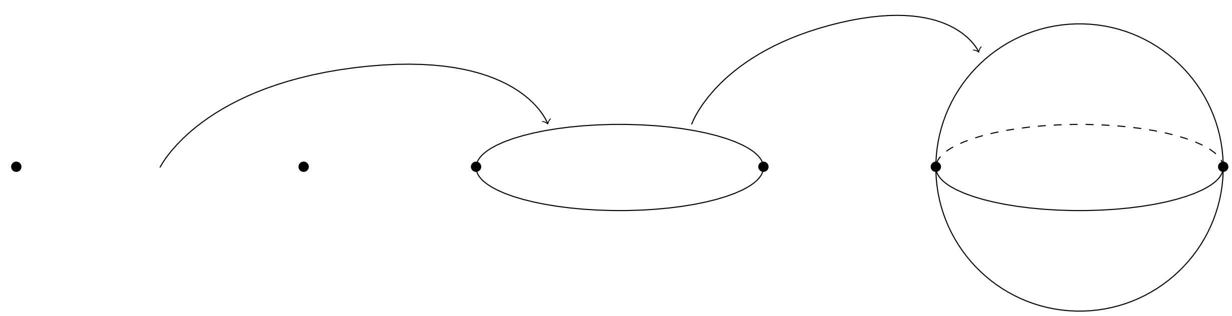
To achieve our goal of classifying surfaces, we first must say what it means for two surfaces to be the same. As with curves, we will allow surfaces to bend, stretch, and pass through themselves. But our main technique will be cutting and pasting surfaces along circles and line segments, as described in the next two sections.

3 Cell Complexes

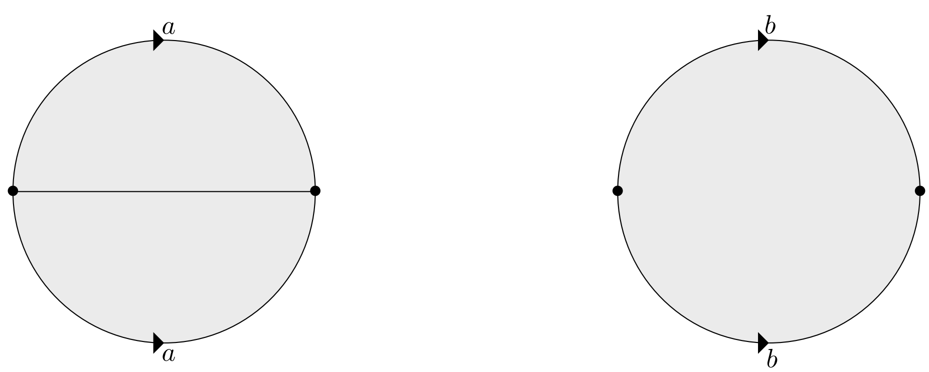
We will now build/describe surfaces as graphs with "faces." Recall that the boundary of a disk is a circle.

Definition. A cell complex $C = (V, E, F)$ is a space built as follows. We first consider its skeleton, which is a graph $G = (V, E)$. The set F consists of disks, called faces, which are glued to the skeleton G along their boundaries. We always imagine that no two faces intersect, except perhaps where their boundaries meet G.

Examples. We begin with a step-by-step example: starting with vertices, adding edges, then adding faces.



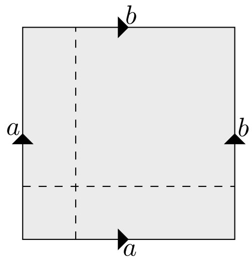
Here, we have begun with two vertices. We then connect these vertices by two edges, forming a circle. Lastly, we create a sphere by gluing two disks to the circle, one for each hemisphere. While this description is valid, it is getting inconvenient to draw dashed lines representing objects that are "behind" others in 3D space. Instead, we can draw all the faces within a planar graph, and label edges that we would like to glue together.

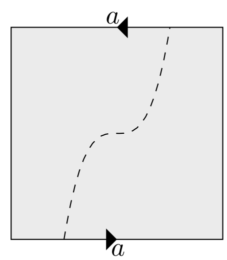


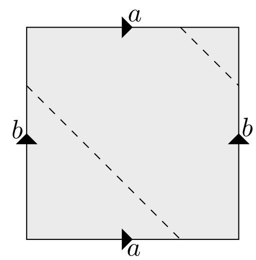
Both of these figures represent a sphere. On the left, we have the same cell complex as illustrated above, consisting of two vertices, two edges, and two faces. On the right, we have a different cell complex for the same surface, eliminating the redundancy of the middle edge. It has two vertices, one edge, and one face.

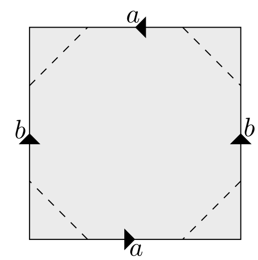
Note that each of the labelled edges has an arrow. The gluing of edges must be done so that the directions of the arrows match. This is no trouble for the sphere, but we will soon see some "twisted gluings" that are much harder to visualize in three dimensions. For these surfaces, the planar representation will be essential.

More Examples. The four planar diagrams below each represent a surface, formed by gluing edges with the same labels. From left to right, these are the sphere, the Möbius band, the torus, and the Klein bottle.



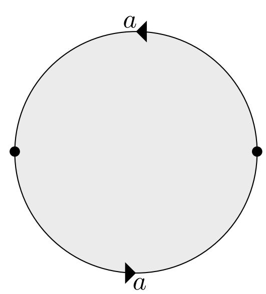


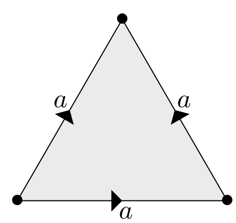




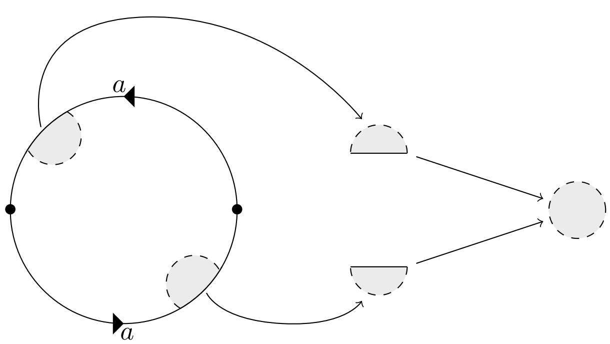
It is helpful to view these planar representations from the perspective of a point moving around in the square. The labelled edges act like portals: the point goes in one and comes out the other. The arrows show which parts of the labelled edges correspond, by giving each edge a direction. For practice, we have drawn a dashed loop on each surface. Follow these loops around, paying attention to how they piece together at each portal.

In addition to representing surfaces that we already know, cell complexes are useful in describing new surfaces. However, not every cell complex is a surface. We give two examples below, only one of which is a surface.





The cell complex on the left is a surface, called the projective plane and denoted $\mathbb{P}$. To prove it is a surface, we must ensure that it locally resembles the plane $\mathbb{R}^2$. This is fairly obvious in the interior of the face, so it suffices to consider the edge. Each point p on the edge is drawn twice. Taking a small region around p gives two half-disks, which are glued together along their common edge. This gluing results in a disk around p.



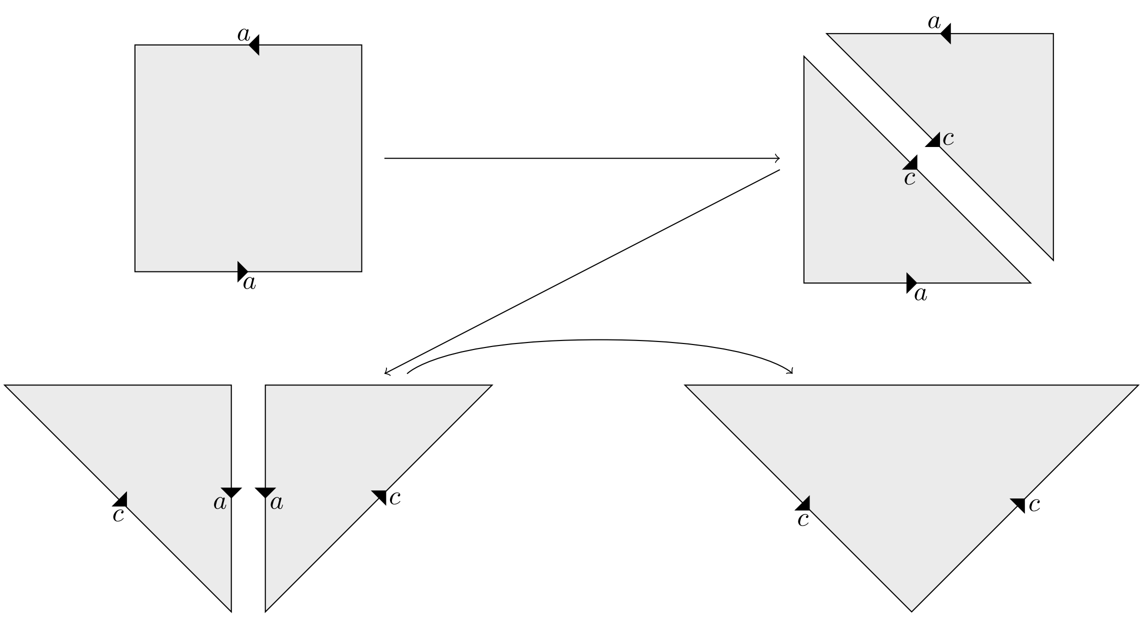
Exercise 4. Give an argument as to why the cell complex on the right (called a dunce cap) is not a surface.

4 The Hauptvermutung

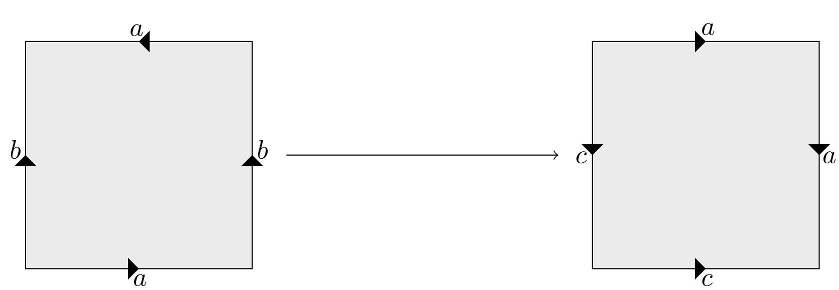
It turns out that every surface can be represented as a cell complex. The proof of this is beyond our scope, so we will restrict our view to cell complexes (with the implicit understanding that this does not matter).

Definition. A cellular decomposition of a surface is a representation of that surface as a cell complex with a connected skeleton, such that there are only finitely many vertices, edges, and faces. If a surface admits a cellular decomposition, we will say that it is compact. A surface is closed if it is compact and boundary-less.

Any compact surface has several cellular decompositions, which we can move between by cutting and pasting. For example, each of the following is a cellular decomposition of the Möbius band. We first make a cut, labelling to keep track of where this cut occurred. Next, we rearrange the two pieces and then glue them.



Exercise 5. Use a similar argument to go between the following two representations of the Klein bottle.

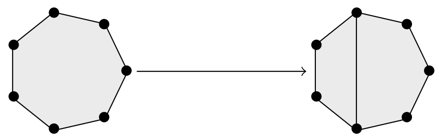


Thus far, we have not actually proven that some surfaces are different from each other. Our goal now is to show that cutting and pasting can be used to go between any two cellular decompositions. This theorem, called the Hauptvermuting (or main conjecture), will help us define invariants to tell different surfaces apart.

Definition. To formalize the manipulations used above, we sort them into three categories.

1. (Vertex Creation) We may split any edge into two edges, by adding a vertex .

2. (Edge Creation) We may split a face in two, by drawing an edge between two vertices on the boundary.

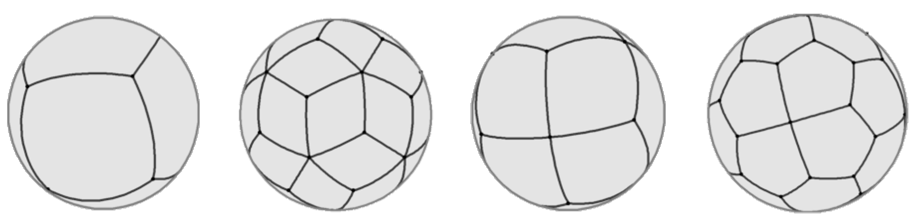


3. (Vertex-Edge Creation) We may create a vertex inside a face, with an edge to it from the boundary.

If A and B are cellular decompositions of a surface, such that B comes from applying these operations to A, then we will say that B is a refinement of A. These operations can also work in reverse, but we will not need to formalize this, because going from B to A via "deletion" is the same as going from A to B via "creation."

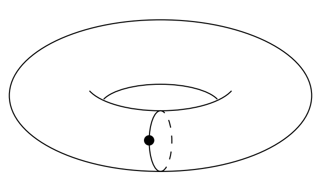
Theorem (Hauptvermutung) [5]. Let A and B be cellular decompositions of some closed surface S. Then S possesses a cellular decomposition C, which is a refinement of both A and B.

Proof. To prove this result, we will have to appeal to intuition and visualizition, with the understanding that these arguments can be made formal in a more advanced setting. Given a closed surface, we may imagine cellular decompositions as connected graphs drawn on the surface, satisfying the condition that each region cut out by the graph must be a disk. Below, we illustrate some examples of these graphs on the sphere $\mathbb{S}$.



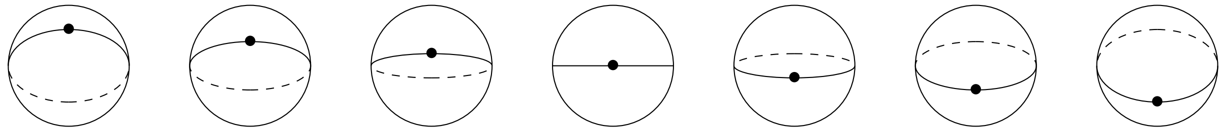
These images are the work of Tom Ruen and are taken from https://en.wikipedia.org/wiki/Spherical_polyhedron.

Each of these graphs possesses a lot of symmetry, so the visible and hidden sides of the sphere look the same. Some graphs may cut out regions that are not disks. In this case, the graph does not represent a cellular decomposition, because the faces of a cell complex must be disks. We draw one such example on a torus.

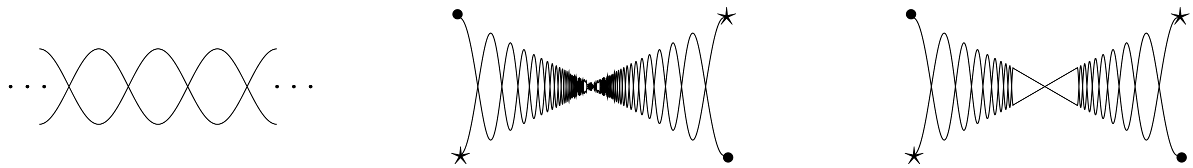


This graph has just one vertex and one edge, but the region it cuts out is a cylinder, rather than a disk.

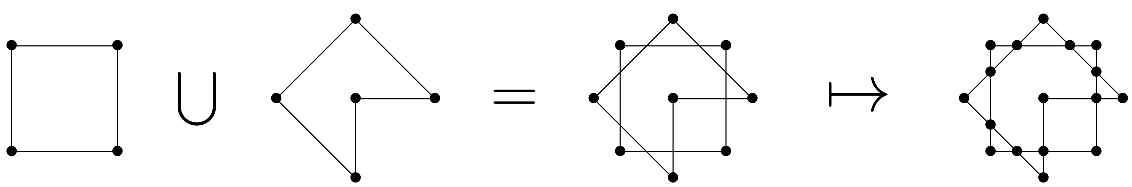
This description of the cellular decomposition based on a graph is unchanged by continuous deformations of the graph along the surface. For example, each of the following illustrates the same cellular decomposition.



The middle drawing lacks three-dimensionality, but is included to show the continuity of the deformation. Now let the cellular decompositions A and B correspond to graphs G and H, respectively, drawn on the surface S. We must make sure that these graphs intersect each other a finite, nonzero number of times. Deforming G and H so that they intersect at least once is easy, but it is more difficult to make sure that the number of intersections is finite. If we can ensure that each edge of G intersects each edge of H finitely many times, then we are done, because there are only finitely many edges. This part of the proof is beyond the scope of this lecture, but I will still try to give an intuitive explanation. Consider the intersections below:



On the left, the points of intersection "go off to infinity." But line segments are bounded, so the edges of G and H cannot go off to infinity. Rather, the existence of infinitely many intersections forces these points to "bunch up," as shown in the middle. Since this "bunching" occurs in a small region, we may rectify it without changing the larger picture, as shown on the right. By "simutaneously untangling" at every point where the intersections "bunch up," we may ensure that G and H a finite, nonzero number of times. Next, we form a new graph K on the surface S. Start with the graph $G \cup H$, which consists of all the vertices and edges belonging to G and H. This graph is not properly drawn on the surface, because it intersects itself. But we may replace each point of intersection by a new vertex, subdividing edges as necessary. For example:



This new graph K is properly drawn on the surface S, because we have dealt with all the points of intersection. Note that K is connected, because G and H intersect at least once; and K is finite, because G and H intersect finitely many times. Now note that K can be built up from either G or H by refinement. Subdividing edges corresponds to vertex creation. Adding new edges corresponds to edge or vertex-edge creation. We have required that the regions cut out by G are disks. This is not changed by refinement, so the regions cut out by K are disks as well. Thus K defines a cellular decomposition C of S, which refines both A and B. $\square$

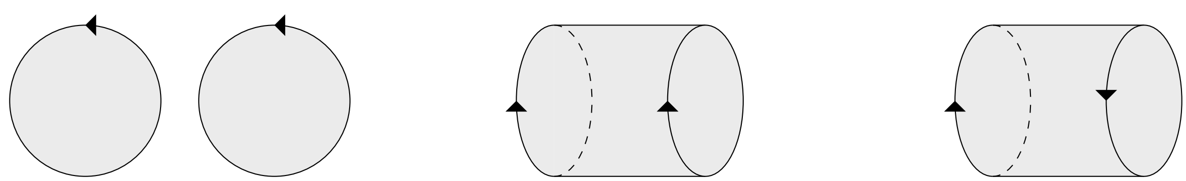
The importance of the Hauptvermutung will arise in defining properties of a surface, which is easiest to do in terms of a cellular decomposition. But it may not be clear that we have defined a property of the surface, rather than something depending on the chosen cellular decomposition. This will be done as follows.

Corollary. Suppose that we define a property/invariant F via cellular decompositions of a closed surface and that this property/invariant is unchanged by refinement of the cellular decomposition. Then F is independent of the choice of cellular decomposition. Therefore, F is a property/invariant of the closed surface itself.

Proof. Suppose that A and B are cellular decompositions of the a closed surface. By the Hauptvermutung, they possess a common refinement C. Since F is unchanged by refinement, we have $F(A) = F(C)$ and $F(B) = F(C)$. This implies that $F(A) = F(B)$, so F does not depend on the choice of decomposition. □

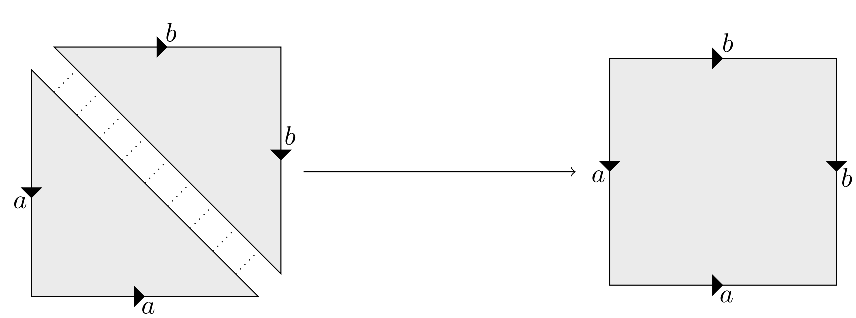
5 Building New Surfaces

We will now discuss some fundamental ways to build new surfaces. These operations will allow us to give a simple enumeration of all closed surfaces, phrased only in terms of those surfaces we have already introduced. Suppose that we have two surfaces, each with a circle as part of its boundary. These two surfaces may be glued along these circles to create a new surface. A surface may also be glued to itself, if its boundary contains two circles. As with edges above, we use arrows to specify how two circles are glued. For example:



On the left, two disks are glued together to form a sphere. You can think of them like the North and South hemispheres of the Earth, meeting in a circle at the equator. In the middle, the ends of a cylinder are glued together to form a torus. On the right is a Klein bottle, because the two circles have opposite orientations.

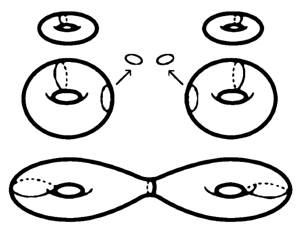
Example. The boundary of a Möbius band is a circle. Using the results of Exercise 5 and the example preceding it, we can see that two Möbius band form a Klein bottle when glued along their boundary circles.



Exercise 6. Show that gluing a disk and a Möbius band along their boundary circles forms $\mathbb{P}$.

We now come to some definitions that only make sense for connected surfaces. This presents no issue, for two reasons: we have required compact surfaces to be connected (since the skeleton graph must be connected) and any surface is just a union of connected surfaces. As such, we will now only consider connected surfaces.

Definition. Given two surfaces S and T, we may form a new surface $S \# T$ called their connected sum. This surface is constructed as follows. Let S' be S with a small disk removed (if S has boundary, we require that this disk not touch the boundary). We define T' in the same way. Since the boundary of the disk is a circle, the surfaces S' and T' both have circles on the boundaries. We form $S \# T$ by gluing S' and T' along these circles. As with much discussed above, this concept is best illuminated by a picture, as follows:



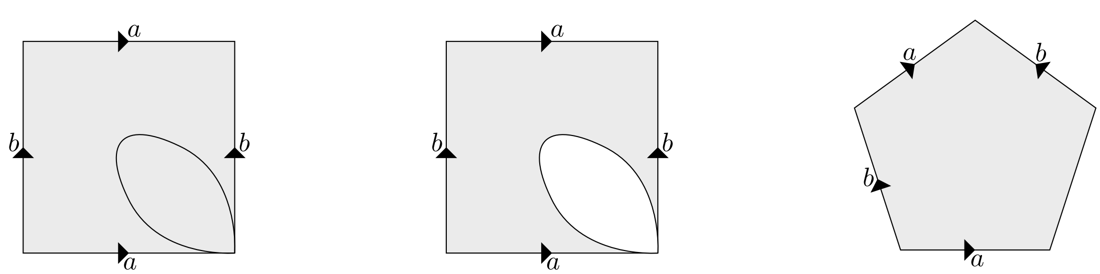
This image is taken from https://www.learner.org/courses/mathilluminated/units/4/textbook/07.php.

Here, we begin with two copies of the torus $\mathbb{T}$. After removing a disk from each, we glue them together, resulting in a two-holed surface called the double torus and denoted $\mathbb{T} \# \mathbb{T}$. Both surfaces are connected, so a disk in one spot can be "slid" to any other spot. Thus the connected sum does not depend on how this disk is chosen. One more complication arises in ensuring that the connected sum does not depend upon arbitrary choices: there are two choices of orientation for the boundary circle. We can't prove it now, but we will take for granted that this orientation does not matter, with justification offered in the next section.

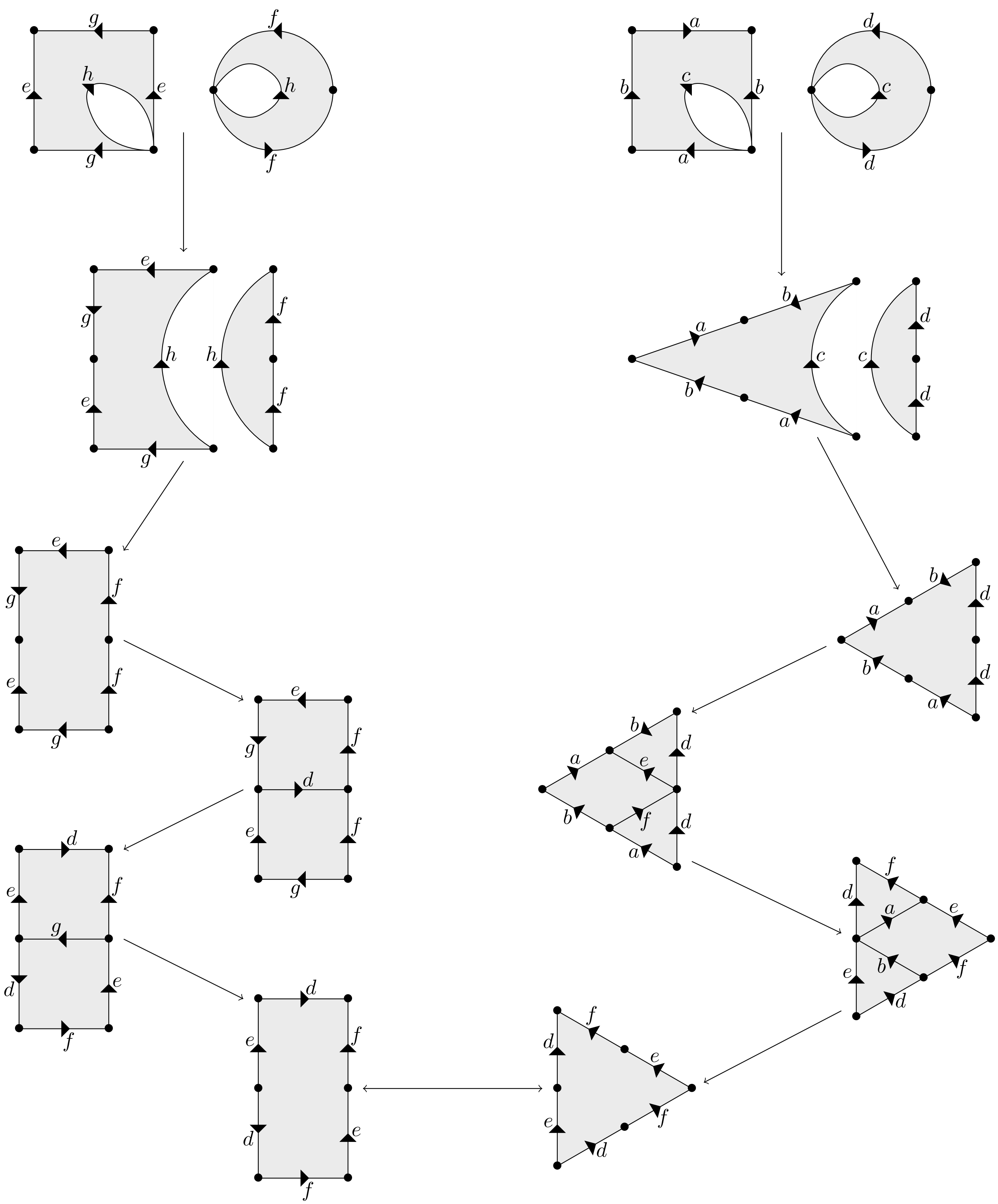
Exercise 7. Use the results of Exercise 6 and the example preceding it to show that $\mathbb{P} \# \mathbb{P} = \mathbb{K}$.

As we have seen, it is easiest to work with surfaces in terms of their cellular decompositions. To calculate a connected sum in terms of a cellular decomposition, we find a face on the surface that does not touch the boundary of the surface (in the case of closed surfaces, this is trivially satisfied). Removing this face is equivalent to cutting out a disk, whose corresponding boundary circle is given by the edges surrounding it.

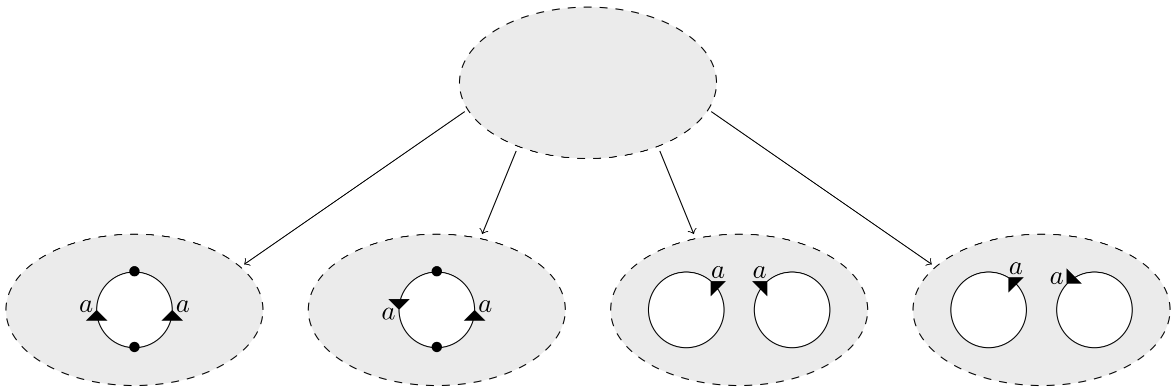
Example. Here we illustrate cutting a disk out of a torus. Keeping track of the edge gluing, you can also see that all the vertices are glued together. Thus we may pull apart this loop, because the gluing of the corresponding vertices is still required by the other edges. This polygonal representation will be useful below.



We now come to a fundamental result. By considering cellular decompositions, we show that $\mathbb{T} \# \mathbb{P} = \mathbb{K} \# \mathbb{P}$. A more elegant approach is outlined in the homework, which should be attempted at the end of this section. We will start with $\mathbb{T} \# \mathbb{P}$ on the left and $\mathbb{K} \# \mathbb{P}$ on the right, moving towards a common representation.



There are yet four more ways that we can alter a surface, by locally modifying a small disk on the surface. These four modifications are illustrated below. They are called the cap, cross-cap, handle, and cross-handle.



As with connected sums, the choice of disk doesn't matter, because it may be slid around a connected surface.

Exercise 8. Apply each of these operations to the sphere. Show that the surfaces formed, from left to right, are a sphere, projective plane, torus, and Klein bottle. Note that adding a cap doesn't change the surface.

Proposition. We now apply these operations to a surface S. Adding a cross-cap to S forms the surface $S \# \mathbb{P}$. Adding a handle to S forms the surface $S \# \mathbb{T}$. Adding a cross-handle to S forms the surface $S \# \mathbb{K}$.

Proof. In Exercise 8, we applied each of these operations to a disk on a sphere. Without loss of generality, we may assume that this disk is the northern hemisphere. Call the resulting surface R. In forming the connected sum $S \# R$, we first remove a disk from R and a disk from S. For R, we may remove the southern hemisphere, which was unchanged in our modifications and thus is still a disk. Now we glue the modified northern hemisphere to S, in place of the removed disk. In all, what have we changed about S? We removed a disk and replaced it by the modified northern hemisphere, which is just a disk altered by one of the above operations. It follows that forming $S \# R$ is the same as performing the corresponding operation on S. $\square$

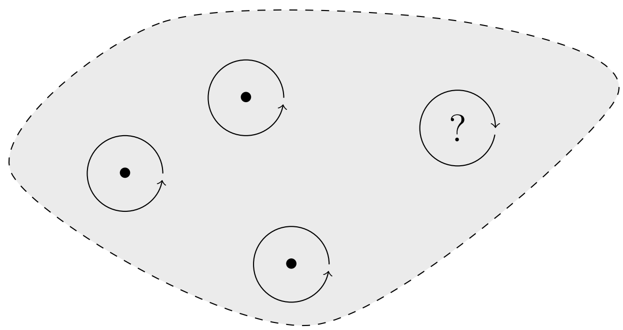
Exercise 9. For any surface R, show that $R = R \# \mathbb{S}$ (recall that $\mathbb{S}$ denotes the sphere).

Note. As mentioned above, you now have the tools necessary to attempt Homework 1. There, you will find an exercise greatly simplifying the proof that $\mathbb{T} \# \mathbb{P} = \mathbb{K} \# \mathbb{P}$, as well as an extension of our main result.

6 Orientability

In the previous section, we began to see an importance in the direction chosen for a circle. A more careful consideration of this phenomenon will not only help with gluing constructions, it will also provide the first property to tell surfaces apart. Locally, an orientation of a surface is just a choice of direction for circles: clockwise or counterclockwise. Depending on how the surface twists, this may or may not extend to a global choice of direction. In what follows, we will build intuition, while successively making this more rigorous.

Consider an oriented circle, sitting on a surface. We assume that this circle sits in a small, disk-shaped region (for example, it cannot wrap around the hole of a torus). If the circle can be moved around the surface, so that it returns to the same position with the opposite orientation, then we will say that the surface is non-orientable. If this is impossible, then there is a way to globally choose a "correct" direction for circles. In this case, we will say that the surface is orientable. Any orientable surface has two choices of orientation, corresponding to a choice of "clockwise or counterclockwise."



Note. The concept of orientability allows us to revisit the question of connected sums and the orientation of the boundary circle. Clearly, orientation does matter in gluing, since pasting together the ends of a cylinder can form either a torus or a Klein bottle. But when considering a connected sum $R \# S$, the orientation of these circles does not matter. If either R or S is non-orientable, then the boundary circle can be slid around to get the opposite orientation. Therefore, it is only when R and S are both orientable that the orientation of the boundary circle might be an issue. However, we shall prove that the only orientable closed surfaces are of the form $m\mathbb{T} = \mathbb{T}\#\mathbb{T}\#\ldots\#\mathbb{T}$ (henceforth, we shall write $mS = S\#S\#\ldots\#S$, iterated $m \geq 0$ times, with the convention that $0S$ is a sphere). For the sphere, the torus, or any of the iterates $m\mathbb{T}$, the direction of the circle does not matter, because the surface is equivalent to itself with opposite orientation (this is not always true in high dimensions). For any $m \geq 0$, this is because $m\mathbb{T}$ can be drawn with reflective symmetry. Centering the removed disk on the plane of symmetry, we see that a reflection about this plane takes the surface to itself and reverses the orientation of the boundary circle. This symmetry is clearest pictorially.

This image is taken from https://people.math.osu.edu/fiedorowicz.1/math655/classification.html.

Exercise 10. If S is a non-orientable surface, show that $S \# \mathbb{T} = S \# \mathbb{K}$.

We now wish to give a rigorous definition of orientability for a closed surface. To do so, we will define the property in terms of a cellular decomposition, and then argue for invariance by the Hauptvermutung. As we have seen above, it is always possible to pick an orientation on a disk. Rather than considering every circle on a disk simultaneously, it suffices to choose a direction for one circle, such as its boundary circle. The direction of this circle will determine a direction for any other circle, i.e. an orientation for the disk.

Definition. Consider two faces F_1 and F_2 in a cellular decomposition, meeting at an edge E. Give each face an orientation. Consider a correctly oriented circle on F_1. If we move this circle across E to F_2, then it may or may not be correctly oriented on F_2. If the circle is still correctly orienteed, we will say that the orientations of F_1 and F_2 agree at E. Otherwise, we will say that they disagree. This is illustrated below.

Orientations Agree   Orientations Disagree

Notice that, when the two orientations agree, they seem to be going in opposite directions along the edge. In contrast, when the orientations disagree, they seem to be going in the same direction along this edge.

Definition. An orientation of a cellular decomposition is a choice of orientation for each face, such that the faces agree at each edge. If a surface has an orientation, it is said to be orientable. Otherwise, it is said to be non-orientable. If a connected surface is orientable, then a choice of orientation for a single face determines the orientation of every other face. Therefore, every connected surface has either zero or two orientations.

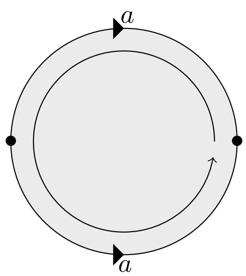
For closed surfaces, we can use the corollary to the Hauptvermutung to prove that the property of orientability does not depend on the choice of cellular decomposition. We must confirm that the property is unchanged under refinement. We will demonstrate this for two of the three cases, leaving vertex creation to the reader.

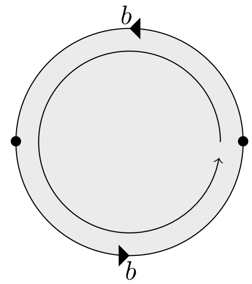
- (Edge Creation) In forming an edge within a face, we do not need to change the orientation of the face. An illustration shows that the face agrees with itself along the created edge. Therefore, edge creation does not change orientability: the surface was orientable before if and only if it is still orientable after.

- (Vertex-Edge Creation) This operation is more complicated, since it adds a new face. But an orientation of this face is equivalent to orienting each of the two resulting faces, with the requirement that they agree along the created edge. Again, this is best understood pictorially. Along any of the other edges, the orientation is unchanged. Thus it agrees with other faces before if and only if it still agrees after.

Exercise 11. Show that the operation of vertex creation does not change the property of orientability.

Example. To get a feel for discerning orientability from a planar diagram, we consider the sphere and the projective plane. In each diagram, there is only one face to orient and the two choices are essentially identical. To see if the surface is orientable, we must check the one edge to see if the face agrees with itself.





When the orientations of two faces agree along an edge, recall that they seem to go in *opposite* directions near the given edge. Orienting a face is the same as picking one of two directions along its boundary circle. Now that the edge has been given a direction as well, we will say that the face and edge agree if the edge points in the direction induced by the orientation of the face. If we give an edge between two faces a direction, then the two faces agree along that edge if and only if the edge agrees with one face and *disagrees* with the other.

With the sphere on the left, the top edge agrees with the face and the bottom edge disagrees (these edges are really the same). Thus the face agrees with itself, so the sphere is orientable. With the projective plane on the right, the face agrees with "both" edges. Thus the face disagrees with itself, so the projective plane is non-orientable. We have now shown that $\mathbb{S} \neq \mathbb{P}$, which is our very first distinction among closed surfaces.

Exercise 12. Using a similar line of reasoning, show that $\mathbb{T}$ is orientable, but $\mathbb{K}$ is not.

Proposition. Let R and S be closed surfaces. Then $R \# S$ is orientable if and only if both R and S are.

Proof. By edge creation, we can create a new face bounded by a single edge. In forming a connected sum, we may therefore assume that the faces deleted each have only one edge. Let R' and S' denote R and S with these faces removed. The boundary circle of R' is an edge D and the boundary circle of S' is an edge E. Suppose that R and S are both oriented. Then we can orient the faces of R and S so that they agree along every edge. Because the faces of R' and S' inherit this agreeability, the orientability of $R \# S$ only depends on agreement along the glued edge $D = E$. But D is on the boundary of R', so it touches one face instead of two. We orient D to agree with this face. Similarly, we orient E to disagree with the face that it touches. In gluing D and E such that these orientations are respected, one of the faces in question agrees with the newly identified edge $D = E$, while the other one disagrees with this edge. Thus the two faces agree along the edge, so $R \# S$ is orientable. Conversely, suppose that $R \# S$ is orientable. Then we can orient the faces of $R \# S$ so that they agree along every edge. Since S' sits inside $R \# S$, it inherits this agreeability. We recover S by gluing a disk to E, adding one new face. Since we can choose the orientation of this face, we just choose it to agree with the other face touching E. Thus S is orientable (and R is analogous). □

Exercise 13. Show that $m\mathbb{T}$ and $n\mathbb{P}$ are distinct surfaces for any integers $m \geq 0$ and $n > 0$.

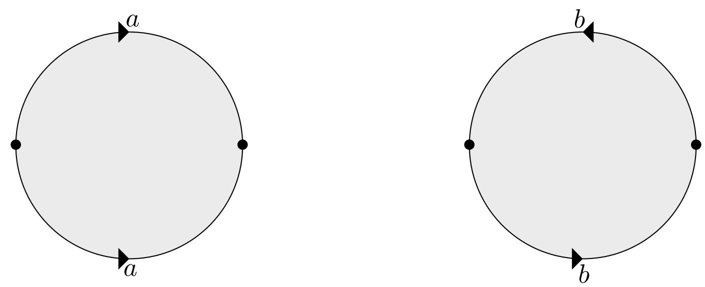
7 Euler Characteristic

We now give a second, more discerning invariant, which will be very helpful in telling surfaces apart.

Definition. Consider a closed surface S, having a cellular decomposition consisting of V vertices, E edges, and F faces. The Euler characteristic of S is $\chi(S) = V - E + F$ (this does not depend on the decomposition).

Exercise 14. Show that the Euler charactereistic is unchanged by the three operations of refinement.

The equation for the Euler characteristic is straightforward, but there is a small difficulty in calculating this number from a planar diagram. When two edges are glued, we must be sure to only count this edge once. Moreover, the gluing of vertices is not even explicitly specified; it must be reasoned from the edge gluings, in order to count the number of vertices. As an example, we consider the sphere and the projective plane.



On the right, we have a cellular decomposition of the projective plane. It clearly has one face and one edge. We can also see that the b arrow points towards "both" of the vertices, so the are really one and the same. Thus $\chi(\mathbb{P}) = 1 - 1 + 1 = 1$. On the left, we have a cellular decomposition of the sphere, also having one face and one edge. But the left vertex has "both" a arrows pointing away from it and the right vertex has "both" arrows pointing towards it. Therefore, the gluing of this edge does not cause the vertices to be glued, so there really are two vertices. Thus $\chi(\mathbb{S}) = 2 - 1 + 1 = 2$. This has given us another proof that $\mathbb{S} \neq \mathbb{P}$.

Exercise 15. Using this technique and the diagrams in Section 3, calculate $\chi(\mathbb{T})$ and $\chi(\mathbb{K})$.

Proposition. For any closed surfaces R and S, we have $\chi(R \# S) = \chi(R) + \chi(S) - 2$.

Proof. Given a cell complex T, we will write V_T, E_T and F_T for the numbers of vertices, edges and faces, respectively. Using edge creation, we can create a new face, whose boundary circle has only one edge and one vertex. In forming a connected sum, we may therefore assume that the faces deleted are of this form. Let R' and S' denote R and S with these faces removed. Then $F_{R'} = F_R - 1$ and $F_{S'} = F_S - 1$, although the numbers of vertices and edges are unchanged. In gluing the boundary circles of R' and S', we identify two edges and two vertices, so we have $V_{R\#S} = V_{R'} + V_{S'} - 1$ and $E_{R\#S} = E_{R'} + E_{S'} - 1$. It follows that

$$\begin{aligned}
\chi(R \# S) &= V_{R\#S} - E_{R\#S} + F_{R\#S} \\
&= V_{R'} + V_{S'} - 1 - (E_{R'} + E_{S'} - 1) + F_{R'} + F_{S'} \\
&= V_R + V_S + E_R + E_S + F_R - 1 + F_S - 1 = \chi(R) + \chi(S) - 2.
\end{aligned}$$

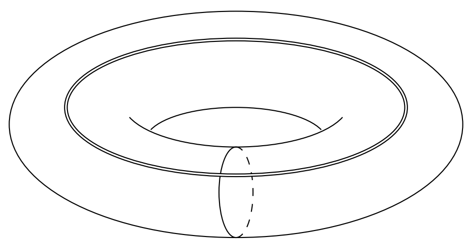
□

Exercise 16. Fix integers $m, n \geq 0$ with $m \neq n$. Show that $m\mathbb{T}$ and $n\mathbb{T}$ are distinct surfaces. Show that $m\mathbb{P}$ and $n\mathbb{P}$ are distinct surfaces. By this and Exercise 13, the surfaces $\mathbb{S}, \mathbb{T}, 2\mathbb{T}, \ldots, \mathbb{P}, 2\mathbb{P}, \ldots$ are all distinct.

Lastly, one might ask: what is the geometric significance of the Euler characteristic? Essentially, it measures the number of cuts one can make on a surface without actually cutting it in two. We cut along a circle, which may be twisted about the larger surface, in contrast to the small circles we considered previously.

Definition. For a closed surface S, its genus $g(S)$ is the largest number of circular cuts that can be made without disconnecting the surface. We cut one circle at a time, so the circles cannot intersect each other.

For example, either one of the following circles can be cut out of a torus, but not both. After cutting out one circle, the other curve has also been cut and is no longer a circle. We have $g(\mathbb{T}) = 1$, because any circle that does not disconnect $\mathbb{T}$ cuts it into a cylinder, which cannot be cut by another circle without disconnecting. The bigger circle is drawn as a double-line to clarify the illustration and exemplify cutting one of the circles.



The sphere has genus $g(\mathbb{S}) = 0$, because any circular cut separates it into two disks. Proofs of such results are beyond our scope, but there are two simple formulas relating genus with Euler characteristic. If S is an orientable closed surface, then $\chi(S) = 2 - 2g(S)$. If S is a non-orientable closed surface, then $\chi(S) = 2 - g(S)$.

Exercise 17. Use these formulas to calculate the genera (plural of genus) of $\mathbb{P}$, $\mathbb{K}$ and $\mathbb{T} \# \mathbb{T}$. In each case, find a maximal number of circular cuts. Can you see why any further cutting will disconnect the surface?

8 Classification

We have now seen that the surfaces $\mathbb{S}, \mathbb{T}, 2\mathbb{T}, \ldots, \mathbb{P}, 2\mathbb{P}, \ldots$ are all different. It remains to show that these are the only closed surfaces. It is equivalent to show that any closed surface may be formed by adding handles and cross-caps to a sphere (recalling that $\mathbb{T} \# \mathbb{P} = \mathbb{P} \# \mathbb{P} \# \mathbb{P}$). We will also consider caps and cross-handles, but these are subsumed by the previous case, because $\mathbb{K} = \mathbb{P} \# \mathbb{P}$ and adding a cap does not alter a surface.

Definition. We will call a surface ordinary if it can be formed from a finite number of spheres, by finitely iterating the following operations: removing disks and adding caps, cross-caps, handles and cross-handles.

Whereas our definition required compact surfaces to be connected, an ordinary surface may be disconnected. For example, any finite number of disks is ordinary, because it can be formed by cutting a disk out of each of several spheres. However, we will demonstrate a close relationship between ordinary and compact surfaces.

Lemma. The connected sum of two connected ordinary surfaces is again ordinary.

Proof. Suppose that N and S are both connected and ordinary. Then they are each formed by performing the above operations on a single sphere. Their connected sum is formed by removing a disk from each and gluing the resulting circles together. Let N' and S' denote these surfaces with a disk removed from each. Since N comes from a modified sphere, the surface N' can be seen as coming from the same modifications applied to a disk. The surfaces S and S' can be related in the same way. But the sphere can be viewed as two disks (the North and South hemispheres) meeting in a circle (the equator). Forming the connected sum $N \# S$ is the same as replacing the North hemisphere by N' and the South hemisphere by S'. Thus $N \# S$ is formed by performing the allowable modifications on each hemisphere, so the final surface is ordinary. □

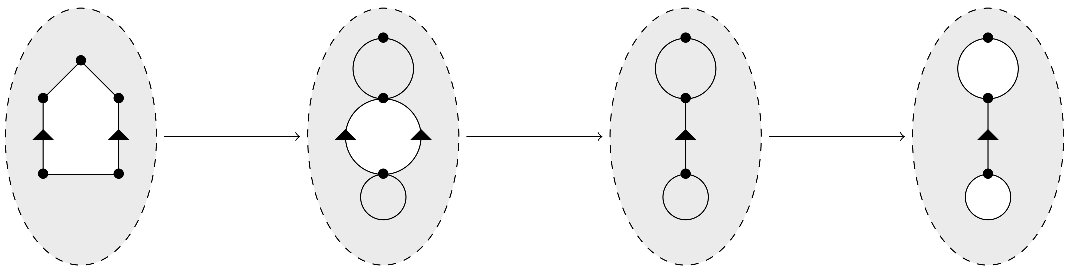
Theorem (Classification) [2]. A surface is ordinary if and only if it is a finite union of compact surfaces.

Proof. We follow Conway's approach, as described in [2]. From previously mentioned cellular decompositions, it is clear that a connected, ordinary surface is compact. Since an ordinary surface has finitely many connected pieces (coming from the finite number of spheres), it can therefore be formed as a union of finitely many compact surfaces. We now wish to show that a finite union of compact surfaces is ordinary. By definition, a compact surface has a cellular decomposition, so a union of compact surfaces can be expressed as a cell complex with finitely many vertices, edges and faces. If we forget all identifications of edges, we may imagine these faces as a finite collection of disks. We can get back the original surface by re-gluing edges, one pair at a time. We noted above that a finite number of disks forms an ordinary surface, so the disjoint collection of faces starts out as ordinary. To show that the final surface is ordinary, it suffices to show that gluing two edges preserves the property of being ordinary. Suppose we are given two edges, D and E, occurring along the boundary of an ordinary surface S. We will prove that the surface stays ordinary after gluing D and E.

(a) Suppose that D and E each form a full circle. By the definition of ordinary surfaces, each boundary circle must have come from deleting a disk from another ordinary surface. If D and E lie on the same connected part of S, then removing two disks and identifying their boundary circles is just the construction of a handle or cross-handle. If D and E lie on a different connected parts of S, then this action corresponds to taking a connected sum. In either case, the resulting surface is again ordinary.

(b) Suppose that D and E form the only two edges along a common boundary circle. Again, this circle must have come from deleting a disk from another ordinary surface. Gluing the two sides of this boundary circle is just the construction of a cap or cross-cap. Thus the resulting surface is still ordinary.

(c) Now suppose that one of the edges forms part—but not all—of a boundary. Then this boundary circle contains an arc of other edges (or two such arcs, cut out by both of the edges D and E). We may get rid of this pesky arc by gluing it to the boundary of a disk, although this forces us to lose some of the original gluing information. By attaching one or two disks in this way, our situation is reduced to either (a) or (b), so we may perform the gluing as in this previous case. The new surface is ordinary, but it has lost some information. But we can now remove the place-holder disk, re-exposing the edges of the pesky arc (and recovering the corresponding gluing information). Since we have now glued the edges D and E, these other edges are no longer an issue. But the surface clearly remains ordinary after deleting any finite number of disks. Thus we have: glued D and E; showed that the surface remains ordinary; and not lost any gluing information from the other edges. Below, we illustrate this process.



We have now proven the desired result! The classification of closed surfaces will follow easily. □

Corollary. Every closed surface occurs exactly once in the list $\mathbb{S}, \mathbb{T}, 2\mathbb{T}, \ldots, \mathbb{P}, 2\mathbb{P}, \ldots$.

Proof. Let S be a closed surface. Then S is ordinary and connected, by the above. A closed surface has no boundary, so S must have been obtained from a single sphere by adding finitely many caps, cross-caps, handles and cross-handles. As noted at the beginning of this section, this implies that S occurs in the list. We proved in previous sections that every surface in the list is disjoint, completing the proof. □

If you wish to see some illuminating illustrations, I strongly recommend perusing [2]. This short article offers an amazing informal exposition (much more of a compact and elegant presentation than I have given here).

References

[1] Adams, C. C. (2004). *The Knot Book.* Providence, RI: The American Mathematical Society.

[2] Francis, G. K., & Weeks J. R. (1999). Conway's ZIP Proof, *The American Mathematical Monthly*, *106*(5).

[3] Harary, F. (1994). *Graph Theory.* Boulder, CO: Westview Press.

[4] Hatcher, A. (2001). *Algebraic Topology.* Cambridge, UK: Cambridge University Press.

[5] Henle, M. (1994). *A Combinatorial Introduction to Topology.* New York, NY: Dover.

[6] Patrascu, A. (2015). The mapping of non-planar diagrams into planar diagrams and the exact solution of QCD. Retrieved from https://www.researchgate.net.

[7] Richards, I. (1963). On the Classification of Noncompact Surfaces, *Transactions of the AMS, 106*(2).

[8] Spivak, M. (1999). *A Comprehensive Introduction to Differential Geometry* (3rd ed., Vol. 1). Houston, TX: Publish or Perish, Inc.