THE CLASSIFICATION OF SURFACES WITH BOUNDARY

THOMAS GEORGE

Abstract. The classification of surfaces theorem was one of the earliest triumphs of algebraic topology. It states that any closed connected surface is homeomorphic to the sphere, the connected sum of n tori, or the connected sum of m projective planes. This paper begins by defining the geometric, topological, and algebraic tools necessary to understanding the theorem, then proceeds through the technical motivation of the theorem and its proof, and closes with its extension to surfaces with boundary. Along the way, a variety of examples and other pertinent information are provided. Some knowledge of algebraic topology will be helpful.

Contents

1. The Goal

The main result of this paper, the classification of surfaces theorem, is as follows:

Theorem 1.1 (Classification of Surfaces)**.** *Let X be a compact connected surface. Then X is homeomorphic to S^2, the n-fold torus T_n, or the m-fold projective plane P_m.*

The proof of the theorem has three essential parts. The first is to show that a compact surface is triangulable; we omit this section of the proof, as noted in Section 5. The next step involves reducing the problem of classifying compact surfaces to that of classifying quotient spaces of polygons, and verifying that this reduction maintains rigor. This portion of the proof is sketched in Section 5. Finally, there remains to classify polygonal quotient spaces; this portion of the proof we give in detail, beginning in Section 2.

As noted in the abstract, the final section of the paper extends the above result to surfaces with boundary.

Date: August 26, 2011.

2. Building Blocks

We begin with an explanation of the geometric terms and tools that will allow us to construct surfaces from 2-dimensional polygons.

Consider a set of points p_i in the plane, all of which lie on a circle. Consider also the set of lines drawn between any two successive points. Then the space enclosed by these lines is called the **polygonal region** P determined by the points p_i. The points p_i are called the **vertices** of P; the line segment joining p_{i-1} and p_i is called an **edge** of P; the union of the edges of P is denoted ∂P, and $P - \partial P$ is denoted $\text{Int}(P)$.

Given a line segment L in $\mathbb{R}^2$, an **orientation** of L is an ordering of its end points. The first, say a, is called the **initial point** and the second, say b, is called the **final point** of the oriented line segment. If L' is another line segment, oriented from c to d, then the **positive linear map** of L onto L' is the homeomorphism h that carries the point $x = (1-s)a + sb$ of L to the point $h(x) = (1-s)c + sd$ of L'.

Let P be a polygonal region in the plane. A **labelling** of the edges of P is a map from the set of edges of P to a set S called the set of **labels**. Given an orientation of each edge of P, and given a labelling of the edges of P, we define an equivalence relation on the points of P as follows.

Each point of $\text{Int}(P)$ is equivalent only to itself. Given any edges of P that have the same label, let h be the positive linear map of one onto the other, and define each point x of the first edge to be equivalent to the point $h(x)$ of the second edge. This relation generates an equivalence relation on P. The quotient space X obtained from this equivalence relation is said to have been obtained by *pasting the edges of P together* according to the given orientations and labelling.

Let P be a polygonal region with successive vertices $p_0, \dots, p_n$, where $p_0 = p_n$. Given orientations and a labelling of the edges of P, let $a_1, \dots, a_m$ be the distinct labels that are assigned to the edges of P. For each k, let a_{i_k} be the label assigned to the edge $p_{k-1}p_k$, and let $\epsilon_k = +1$ or -1 according as the orientation assigned to this edge goes from p_{k-1} to p_k or the reverse. Then the number of edges of P, the orientations of the edges, and the labelling are completely specified by the symbol $w = (a_{i_1})^{\epsilon_1}(a_{i_2})^{\epsilon_2} \cdots (a_{i_n})^{\epsilon_n}$. We call this a **labelling scheme of length** n for the edges of P; it is simply a sequence of labels with exponents $+1$ or -1.

Example 2.1. The 1-fold torus T_1 can be expressed by the labelling scheme of length 4 given by $w = aba^{-1}b^{-1}$.

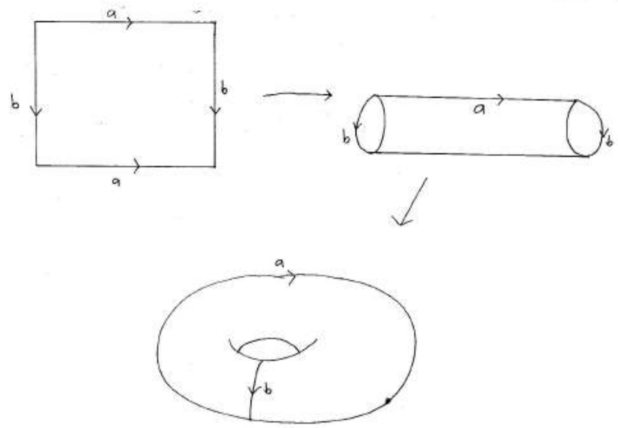
We now verify an important property of the surfaces X which we are constructing from polygonal regions.

Proposition 2.2. *Let X be the space obtained from a finite collection of polygonal regions by pasting edges together according to some labelling scheme. Then X is a compact Hausdorff space.*

Proof. We prove the proposition for one polygonal region; the general case is a simple extension.

Because the quotient map pastes edges together continuously, X is clearly compact. To demonstrate that X is Hausdorff, it suffices to show that the quotient map π is closed, or that for each closed set C of the polygonal region P, the set $\pi^{-1}\pi(C)$ is closed in P. Note that the set $\pi^{-1}\pi(C)$ contains all points of C, and also any other points of P that are pasted to C under π. In order to determine these other

FIGURE 1. The torus T_1 can be realized as the labelling scheme $w = aba^{-1}b^{-1}$



points of P, define for each edge e of P C_e to be the compact subspace $C \cap e$ of P. If e_i is an edge of P that is pasted to e under the pasting homeomorphism f_i, then the set $D_e = \pi^{-1}\pi(C) \cap e$ equals the union of C_e and the spaces $f_i(C_{e_i})$, taken as e_i ranges over all edges of P that are pasted to e. This union is compact, and therefore closed in P. Finally, because $\pi^{-1}\pi(C)$ is the union of C and all the D_e, it is closed in P. □

We continue by defining two of the three structures to which we will demonstrate all surfaces are homeomorphic. The third, the sphere, is simply the set $S^2 = \{(x, y, z) \mid x^2 + y^2 + z^2 = 1\} \subset \mathbb{R}^3$.

Definitions 2.3. Consider the space obtained from a $4n$-sided polygonal region P by means of the labelling scheme $(a_1b_1a_1^{-1}b_1^{-1})(a_2b_2a_2^{-1}b_2^{-1}) \cdots (a_nb_na_n^{-1}b_n^{-1})$. This space is called the **n-fold connected sum of tori**, or simply the **n-fold torus**, and is denoted T_n.

The labelling scheme $abab$ is called the **projective plane**. Let $m > 1$. Consider the space obtained from a $2n$-sided polygonal region P in the plane by means of the labelling scheme $(a_1a_1)(a_2a_2) \cdots (a_ma_m)$. This space is called the **m-fold connected sum of projective planes**, or simply the **m-fold projective plane**, and is denoted P_m.

We now consider constructions that will bring us to a definition of the fundamental group of a space. This tool will prove amply useful in demonstrating whether spaces are homeomorphic.

Definitions 2.4. A **path** in a space X is a continuous map $f : I \to X$ where I is the unit interval $[0, 1]$. A **homotopy** of paths in X is a family
$f_t : I \to X, 0 \leq t \leq 1$, such that

(1) The endpoints $f_t(0) = x_0$ and $f_t(1) = x_1$ are independent of t.
(2) The associated map $F : I \times I \to X$ defined by $F(s, t) = f_t(s)$ is continuous.

Two paths connected by a homotopy of paths are called **homotopic**. We write this as $f_0 \simeq f_1$.

We now demonstrate a basic piece of information about the algebraic structure of homotopies on paths:

Proposition 2.5. *The relation of homotopy on paths with fixed endpoints in any space is an equivalence relation.*

The following definitions provide clarification and tools to prove this proposition:

Definitions 2.6. The equivalence class of a path f under the equivalence relation of homotopy will be denoted $[f]$ and called the **homotopy class** of f.

Given two paths $f, g : I \to X$ such that $f(1) = g(0)$, there is a **composition** or **product path** $f \cdot g$ that traverses first f and then g, and moves twice as fast along either path as it did originally. We define it as:

$$f \cdot g(s) = \begin{cases} f(2s) & 0 \le s \le 1/2 \\ g(2s-1) & 1/2 \le s \le 1 \end{cases}$$

Proof. We must demonstrate reflexivity, symmetry, and transitivity.

Reflexivity is evident: set $f_t = f$. Symmetry follows from the fact that if $f_0 \simeq f_1$ via f_t, then via f_{1-t} we have $f_1 \simeq f_0$.

Transitivity requires us to demonstrate that if $f_0 \simeq f_1$ via f_t, and $g_0 \simeq g_1$ via g_t with $f_1 = g_0$, then $f_0 \simeq g_1$. This follows from the above definition of a composition path, wherein we have $f_0 \simeq f_1$ via h_t, which equals f_{2t} for $0 \le t \le 1/2$, and which equals g_{2t-1} for $1/2 \le t \le 1$. Because $f_1 = g_0$, h_t is well-defined at $t = 1/2$, and furthermore h_t is continuous because it is composed of continuous maps. □

Definitions 2.7. Consider a path $f : I \to X$ such that $f(0) = f(1) = x_0 \in X$. Then we call f a **loop** and call x_0 the loop's **basepoint**. The set of all homotopy classes $[f]$ of loops $f : I \to X$ at the basepoint x_0 is denoted $\pi_1(X, x_0)$, and is called the **fundamental group** of X at the basepoint x_0.

Example 2.8. The fundamental group of the circle S^1 is $\mathbb{Z}$.

Essentially, we are examining two objects here: $\pi_1(S^1)$ and $\mathbb{Z}$. Both, of course, are groups. This example states that there exists an isomorphism between the two groups. To see this, picture the isomorphism as a map, say i, that sets up a correspondence between an integer and the number of times a given loop circles S^1. So, for example, the integer 12 would correspond to the loop in S^1 that traverses the circle 12 times.

It is a basic fact of algebraic topology that the set $\pi_1(X, x_0)$ is a group with respect to the product $[f][g] = [f \cdot g]$; we do not prove it here.

Furthermore, the spaces X we are constructing from the above-defined polygonal regions will all be path-connected, and thus there exists a path f between any two given points x_0 and y_0. Therefore, considering the fundamental group of the space from the basepoint x_0 is equivalent to considering the fundamental group of the space from the basepoint y_0; it requires merely a composition with f, the path between them. This does not change the fundamental group because no extra loop has been traversed, we have merely moved our point of consideration along a well-defined path in the space. For this reason, this paper will from now on consider the fundamental group of a space $\pi_1(X)$, without worrying about the basepoint.

Fundamental groups are incredibly important and useful algebraic tools for analyzing topological structures. We now use them to demonstrate a fundamental portion of the classification theorem: that the spaces under consideration are themselves topologically distinct.

Theorem 2.9. *Let T_n and P_m denote the n-fold connected sum of tori and the m-fold connected sum of projective planes, respectively. Then the surfaces S^2; $T_1, T_2, \ldots$; $P_1, P_2, \ldots$ are topologically distinct.*

This theorem is a consequence of Van Kampen's Theorem, a fundamental tool of algebraic topology which we will not state or address here. Essentially, Van Kampen provides us with the resources to calculate the fundamental groups of a wide variety of spaces, including the n-fold torus and the m-fold projective plane. It is easy to use Van Kampen to see that the n-fold torus and the m-fold projective plane do not have isomorphic fundamental groups. Furthermore, the sphere has trivial fundamental group, because any loop on the sphere can be shrunk over the surface of the sphere to a point; thus our spaces are topologically distinct.

3. Technical Motivation

The following three sections, that is Sections 2, 3, and 4, follow the general form of Munkres' classification of surfaces in his *Topology*.

Our present goal is to prove the classification theorem for polygonal quotient spaces, which is as follows:

Theorem 3.1 (Classification of Surfaces)**.** *Let X be the quotient space obtained from a polygonal region in the plane by pasting its edges together in pairs. Then X is homeomorphic to S^2, the n-fold torus T_n, or the m-fold projective plane P_m.*

In order to do so, we need to be able to 'cut and paste' various polygonal regions in the plane in order to represent a space given by a certain collection of regions and a labelling scheme as a different collection of regions and a different labelling scheme.

First, we formalize our notion of 'cutting apart' a polygonal region. Consider a polygonal region P with vertices $p_0, \ldots, p_n = p_0$, and take k with $1 < k < n-1$. Then we have the polygonal regions Q_1, with vertices $p_0, p_1, \ldots, p_k, p_0$, and Q_2, with vertices $p_0, p_k, \ldots, p_n = p_0$. If we slide the region Q_1 by a translation in $\mathbb{R}^2$, and in doing so produce another region Q_1' that has vertices $q_0, q_1, \ldots, q_k, q_0$, where q_i is the image of p_i under the translation, we have *cut apart* the region P along the line from p_0 to p_k.

Similarly, we can 'paste together' the regions Q_1' and Q_2 to reform the region P with the positive linear map between the edge q_0q_k of Q_1' and p_0p_k of Q_2.

Now we discuss elementary operations on labelling schemes. An **elementary operation** on a labelling scheme is an operation on the scheme that does not affect the resulting quotient space X. These operations will be useful in helping us demonstrate that a wide variety of possible labelling schemes can only result in an easily-enumerated list of potential quotient spaces. These operations are as follows:

(i) *Cut*: The scheme $w_1 = y_0y_1$ can be replaced with the sequence of schemes y_0c^{-1}, cy_1, as long as c does not appear elsewhere in the total scheme and y_0 and y_1 both have length at least two.
(ii) *Paste*: The reverse of the above, the sequence of schemes y_0c^{-1}, cy_1 can be replaced with the scheme $w_1 = y_0y_1$.

(iii) *Relabel*: All instances of a given label can be changed to another label that does not already appear in the scheme. Similarly, the sign of all the exponents of a given label can be changed.
(iv) *Permute*: Any scheme w_i can be replaced by a cyclic permutation of w_i.
(v) *Flip*: The scheme $w_i = (a_{i_1})^{\epsilon_1}(a_{i_2})^{\epsilon_2}\cdots(a_{i_n})^{\epsilon_n}$ can be replaced with its formal inverse $w_i^{-1} = (a_{i_n})^{-\epsilon_n}\cdots(a_{i_1})^{-\epsilon_n}$.
(vi) *Cancel*: The scheme $w_i = y_0aa^{-1}y_1$ can be replaced by the scheme y_0y_1, as long as a does not appear elsewhere in the total scheme and both y_0 and y_1 have length at least two.
(vii) *Uncancel*: The reverse of the above, the scheme y_0y_1 can be replaced with the scheme $y_0aa^{-1}y_1$, where a appears nowhere else in the total scheme.

We define two labelling schemes for collections of polygonal regions to be **equivalent** if one can be obtained from the other by a sequence of elementary scheme operations. Because each elementary operation has as its inverse another such operation, this is an equivalence relation.

We will now embark on demonstrating the above-stated geometric portion of the classification theorem: every space obtained by pasting the edges of a polygonal region together in pairs is homeomorphic either to S^2, the n-fold torus T_n, or the m-fold projective plane P_m. We note here that this does not completely accomplish the task of classifying surfaces, as there remains to show that all surfaces can in fact be obtained by these polygonal quotient spaces; we demonstrate this in Section 5.

Definitions 3.2. Suppose $w_1, \ldots, w_k$ is a labelling scheme for the polygonal regions $P_1, \ldots, P_k$. If each label appears exactly twice in this scheme, we call it a **proper** labelling scheme. Note that if we apply any sequence of elementary operations to a proper scheme, we obtain another proper scheme.

Let w be a proper labelling scheme for a single polygonal region. We say that w is of **torus type** if each label in it appears once with exponent $+1$ and once with exponent -1. Otherwise, we say w is of **projective type**.

The following are a list of lemmas and corollaries necessary to prove the theorem. Because of space constraints, we omit their proofs. All proofs can be constructed from manipulating a polygonal region with the given labelling scheme in the appropriate manner.

Lemma 3.3. *Let w be a proper scheme of the form $w = [y_0]a[y_1]a[y_2]$, where some of the y_i may be empty. Then one has the equivalence $w \sim aa[y_0y_1^{-1}y_2]$, where y_1^{-1} denotes the formal inverse of y_1.*

Corollary 3.4. *If w is scheme of projective type, then w is equivalent to a scheme of the same length having the form $(a_1a_1)(a_2a_2)\cdots(a_ka_k)w_1$, where $k \geq 1$ and w_1 is either empty of of torus type.*

Lemma 3.5. *Let w be a scheme of torus type that does not contain two adjacent terms having the same label. Then w is equivalent to a scheme of the same length as w, and of the form $w_1 = aba^{-1}b^{-1}w_2$, where w_2 is of torus type or is empty.*

Lemma 3.6. *Let w be a proper scheme of the form $w = w_0(cc)(aba^{-1}b^{-1})w_1$. Then w is equivalent to the scheme $w' = w_0(aabbcc)w_1$.*

Remark 3.7. Note that, because of our cyclic permutation operation, the changes that are realized in the two preceding lemmas can be realized in any other portion of a scheme. For example, lemma 3.5 can just as easily be used to obtain a scheme in which the sequence of terms $aba^{-1}b^{-1}$ comes at the end.

4. The Classification

Theorem 4.1 (Classification of Polygonal Quotients)**.** *Let X be the quotient space obtained from a polygonal region in the plane by pasting its edges together in pairs. Then X is homeomorphic to S^2, the n-fold torus T_n, or the m-fold projective plane P_m.*

Proof. Take w to be the labelling scheme by which X is formed from the polygon P. We claim that w is equivalent to one of the following schemes:

(1) $aa^{-1}bb^{-1}$, which produces the 2-sphere,
(2) $abab$, which produces the 1-fold projective plane P_1,
(3) $(a_1a_1)(a_2a_2)\cdots(a_ma_m), m \geq 2$, which produces the m-fold projective plane P_m, or
(4) $(a_1b_1a_1^{-1}b_1^{-1})(a_2b_2a_2^{-1}b_2^{-1})\cdots(a_nb_na_n^{-1}b_n^{-1}), n \geq 1$, which produces the n-fold torus T_n.

We begin by demonstrating that if w is of torus type, it is equivalent to scheme (1) or scheme (4). Note that if w has length four, by the definition of a scheme of torus type, it must either have the form $aa^{-1}bb^{-1}$ or $aba^{-1}b^{-1}$, the first of which is scheme (1) and the second of which is of the type of scheme (4).

If w has length greater than four and is equivalent to no shorter scheme of torus type, then w contains no pair of adjacent terms with the same label. Lemma 3.5 then tells us that w is equivalent to a scheme of the form $aba^{-1}b^{-1}w_3$, where w_3 cannot be empty because w has length greater than four, and so must be of torus type. Further, w_3 cannot contain adjacent terms with the same label because w cannot be reduced. Thus we may apply the lemma to w_3, drawing the conclusion that w is equivalent to a scheme of the form $(aba^{-1}b^{-1})(cdc^{-1}d^{-1})w_4$, where w_4 is either empty or of torus type. If empty, we are done, and if of torus type, we continue inductively to the conclusion that w is of the type of scheme (4).

We now deal with schemes of projective type. In these cases, w will turn out to be equivalent to either scheme (2) or a scheme of the form of scheme (3). Beginning with w of length four, corollary 3.4 implies that w is equivalent to either $aabb$ or to $aab^{-1}b$, the first of which is of type (3). By applying lemma 3.3 with y_0 empty and $y_1 = y_2 = b$, we see that $aab^{-1}b$ can be written in the form $ay_1ay_2 = abab$, which is scheme (2).

Again we induct on the length of w. If w has length greater than four, corollary 3.4 implies that w is equivalent to a scheme of the form $w' = (a_1a_1)\cdots(a_ka_k)w_1$, where $k \geq 1$ and w_1 is empty or of torus type. If w_1 is empty then we are done, and if it has two adjacent terms with the same label then w' is equivalent to a shorter scheme of projective type and the induction hypothesis applies. If neither of these is the case, we apply lemma 3.5 to show that w' is equivalent to a scheme of the form $w'' = (a_1a_1)\cdots(a_ka_k)aba^{-1}b^{-1}w_2$, where w_2 is empty or of torus type.

Then lemma 3.6 tells us that w'' is equivalent to the scheme $(a_1a_1)\cdots(a_ka_k)aabbw_2$. Continuing in this manner, we will reach a scheme of type (3). □

5. Compact Surfaces

We must now demonstrate that all compact surfaces can, in fact, be obtained in the manner we have described: cutting and pasting polygonal regions. We give only an outline of the proof. A more detailed version can be found in Munkres' *Topology*.

We begin with the requisite definitions:

Definitions 5.1. Take X to be a compact Hausdorff space. A **curved triangle** in X is a subspace A of X and a homeomorphism $h : T \to A$, where T is a closed triangular region in the plane. A **triangulation** of X is a collection of curved triangles $A_1, \dots, A_n$ in X whose union is X such that for $i \neq j$, the intersection $A_i \cap A_j$ is either empty, or a vertex of both A_i and A_j, or an edge of both. We also require, if h_i is the homeomorphism associated with A_i, that when $A_i \cap A_j$ is an edge e of both, then the map $h_j^{-1} h_i$ is a linear homeomorphism of the edge $h_i^{-1}(e)$ of T_i with the edge $h_j^{-1}(e)$ of T_j.

It is a basic topological theorem that all compact surfaces have a triangulation; we do not prove it here. To begin, we have the following theorem:

Theorem 5.2. *If X is a compact triangulable surface, then X is homeomorphic to the quotient space obtained from a collection of disjoint triangular regions in the plane by pasting their edges together in pairs.*

Proof. Take $A_1, \dots, A_n$ to be a triangulation of X, with associated homeomorphisms h_i. We assume that the T_i are disjoint, and thus the map $h : E = T_1 \cup \cdots \cup T_n \to X$ is a quotient map. Also, because $h_j^{-1} h_i$ is linear if A_i and A_j intersect in an edge, h pastes the edges of T_i and T_j together via a linear homeomorphism.

The proof of the theorem is twofold: first, we must show that (*Claim 1*) for any edge e of a triangle A_i, there is exactly one other triangle A_j such that $A_i \cap A_j = e$, and second that (*Claim 2*) if the intersection $A_i \cap A_j$ equals a vertex v of each, then there is a sequence $A_i, \dots, A_j$ of triangles with v as a vertex, such that the intersection of each triangle of the sequence with its successor equals an edge of each.

(*Claim 1*) Two demonstrate the first claim, we split it into two steps: first, we must show that an edge e of the triangle A_i is an edge of at least one other triangle A_j, and second that this edge is an edge of at most one other triangle A_j.

The first step is a result of the following proposition, which we leave to the reader:

Proposition 5.3. *If X is a triangular region in the plane and if x is a point one one of the edges of X, then x does not have a neighborhood in X homeomorphic to an open 2-ball.*

This proposition is very similar to proposition 6.3 (a), which we treat and prove in full in the next section. A proof of proposition 5.3 is easily constructed in a similar manner, by examining towards a contradiction the fundamental groups of the neighborhood around the point x in the triangle minus x itself, and the neighborhood around the point $h(x)$ (if our homeomorphism is h) in the open 2-ball, minus $h(x)$ itself. The groups will not turn out isomorphic.

The second step of the first claim is, similarly, a consequence of another proposition which we leave to the reader:

Proposition 5.4. *Let X be the union of k triangles in $\mathbb{R}^3$, each pair of which intersect in the common edge e. Let x be a point on e. If $k \geq 3$, then x does not have a neighborhood in X homeomorphic to an open 2-ball.*

To prove this proposition, some knowledge of algebraic topology is very helpful, as the proposition follows from a demonstration that there is no neighborhood W of x in X such that $W - x$ has abelian fundamental group.

(*Claim 2*) Note that if it were not the case that if the intersection $A_i \cap A_j$ equals a vertex v of each, then there is a sequence $A_i, \dots, A_j$ of triangles with v as a vertex, such that the intersection of each triangle of the sequence with its successor equals an edge of each, a situation would be possible in which the surface X consisted of two regions, each with its own triangulation, with only a vertex in common, and this vertex could be the vertex of multiple triangles in each region.

To prove the claim, it is necessary to demonstrate that because X is a surface, this cannot be the case. This can be done by examining the equivalence classes of triangles connected to each other by other triangles with an edge in common; that is, take two triangles A_i and A_j having v as a vertex to be *equivalent* if there is a sequence of triangles with v as a vertex, beginning with A_i and ending with A_j, such that the intersection of each triangle with its successor is an edge of each. The proof concludes with a demonstration that in the situation mentioned above, the space $W - v$ is nonconnected. □

To conclude, we extend this result to the following theorem:

Theorem 5.5. *If X is a compact connected triangulable surface, then X is homeomorphic to a space obtained from a polygonal region in the plane by pasting the edges together in pairs.*

Proof. We already have, from the previous theorem, a collection $T_1, \dots, T_n$ of triangular regions in the plane which are oriented and labelled such that X is homeomorphic to the quotient space of these regions.

To extend, we must merely paste the edges of triangles with the same label together. This gives us a four-sided polygonal region, and we can continue this process inductively until an n-sided polygonal region is obtained. Because the space X is connected, a situation in which there are multiple polygonal regions with no labels in common is impossible. □

6. Extension to Surfaces with Boundary

This final section is based on a series of exercises from Munkres, which can be found on page 476.

Now that we have built the tools necessary to understanding and proven the classification of surfaces theorem, we extend our result to a classification of surfaces with boundary.

Definition 6.1. Let H^2 be the subspace of $\mathbb{R}^2$ consisting of all points (x_1, x_2) with $x_2 \geq 0$.

Definitions 6.2. A **2-manifold with boundary** (or **surface with boundary**) is a Hausdorff space X with a countable basis such that each point x of X has a neighborhood homeomorphic with an open set of $\mathbb{R}^2$ or H^2.

The **boundary** ∂X of a 2-manifold X with boundary consists of those points x such that x has no neighborhood homeomorphic with an open set of $\mathbb{R}^2$.

Proposition 6.3. *Let X be a 2-manifold with boundary, and take $x \in X$.*

(a) No point of H^2 of the form $(x_1, 0)$ has a neighborhood in H^2 that is homeomorphic to an open set of $\mathbb{R}^2$.
(b) There is a homeomorphism h mapping a neighborhood of x onto an open set of H^2 such that $h(x) \in \mathbb{R} \times 0$ if and only if $x \in \partial X$.
(c) The boundary ∂X is a 1-manifold.

Proof. (a) Suppose, towards a contradiction, that there is a point x of the form $(x_1, 0)$ that has a neighborhood A in H^2 that is homeomorphic to an open set B of $\mathbb{R}^2$. We may assume that A and B are contractible to a point by passing to a smaller neighborhood of x, and therefore both have trivial fundamental group. Let us call our homeomorphism f. Then, should we remove x from A and $f(x)$ from B, the resulting spaces $A - x$ and $B - f(x)$ remain homeomorphic. However, the space $A - x$ has, like the space A, trivial fundamental group, but because B is open, $B - x$ is a space with fundamental group $\mathbb{Z}$. This is a contradiction.

(b) If there exists such a homeomorphism h, $h(x)$ is of the form $(x_1, 0)$, and thus by part (a) has no neighborhood in H^2 homeomorphic to an open set of $\mathbb{R}^2$. Then $x \in \partial X$.

To show the converse, let $x \in \partial X$. Take H^2_+ to be $\{(x, y) \mid y > 0\}$. Let h be a homeomorphism taking a neighborhood of x to $U \subset H^2$. Assume $h(x)$ is not in $\mathbb{R} \times \{0\}$. Then $U' = U \cap H^2_+$ contains $h(x)$ and is an open set in $\mathbb{R}^2$, and so $h^{-1}(U')$ is a neighborhood of x in X that is homeomorphic to a subset of $\mathbb{R}^2$. This contradicts the fact that $x \in \partial X$, so our assumption was false and $h(x) \in \mathbb{R} \times \{0\}$.

(c) Open sets in $\mathbb{R}$ are homeomorphic to a countable union of open intervals. We are working with only one path-connected neighborhood, and so the open subset can be taken to be just one open interval. We want to show that for all $x \in \partial X$, x has a neighborhood, call it A, in ∂X homeomorphic to an open interval, call it B. Then by part (b), there exist h, C such that $h(C)$ is open in H^2 and $h(x) \in \mathbb{R} \times 0$. For all points $y \in A$, $h(y) \in \mathbb{R} \times 0$ as well, because $A \subseteq \partial X$. Thus h is a homeomorphism between A and the interval $B \subset \mathbb{R}$. We want to show the interval to be open. Because h is a homeomorphism, open sets are mapped to open sets, so open A maps to open $h(A)$, and we are done. □

Example 6.4. We show that the closed unit ball B^2 in $\mathbb{R}^2$ is a 2-manifold with boundary.

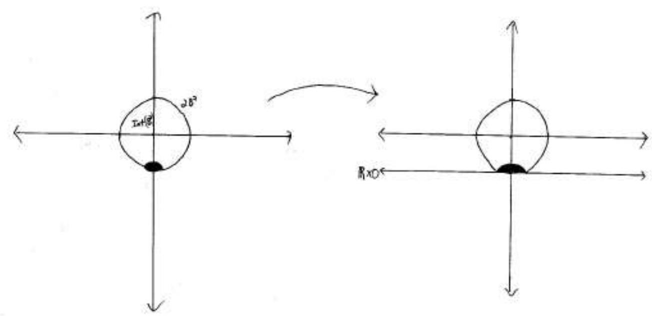
Take $\mathrm{Int}(B^2)$ and $\partial(B^2)$, where $\mathrm{Int}(B^2) = \{(x, y) \mid x^2 + y^2 < 1\}$ is the open unit ball, and $\partial(B^2) = S^1 = \{(x, y) \mid x^2 + y^2 = 1\}$ is the unit circle.

We claim that for all $x \in \mathrm{Int}(B^2)$, x has a neighborhood homeomorphic to an open set of $\mathbb{R}^2$, and for all $x \in \partial(B^2)$, x has a neighborhood homeomorphic to an open set of H^2.

To prove the claim, take $x \in \mathrm{Int}(B^2)$ and an open set $A \subseteq \mathrm{Int}(B^2)$ such that $x \in A$. Then the homeomorphism could be the identity, and we can, for all $x \in \mathrm{Int}(B^2)$, always find an open neighborhood because $\mathrm{Int}(B_2)$ is open. To exhibit the fact that any $x \in \partial(B^2)$ has a neighborhood homeomorphic to an open set of H^2, see figure 2. It is sufficient to consider an open neighborhood of the point $(0, -1) \in B^2$, as any other point on $\partial(B^2)$ has a neighborhood homeomorphic to a neighborhood of this point by a rotation. The homeomorphism pictured essentially 'slides' the curved edge of B^2 down onto $\mathbb{R} \times 0$, and all points above any point on

B^2 are correspondingly moved down by the same amount. That this is a continuous motion of the unit disc with a continuous inverse is clear.

FIGURE 2. A boundary point of B^2 has a neighborhood homeomorphic to an open set of H^2.



Proposition 6.5. *Let X be a 2-manifold; let $U_1, \ldots, U_k$ be a collection of disjoint open sets in X and suppose that for each i, $1 \leq i \leq k$, there is a homeomorphism h_i of the open unit ball with U_i. Let $\epsilon = 1/2$ and let B_ϵ be the open ball of radius ϵ. Then the space $Y = X - \bigcup_{i=1}^{k} h_i(B_\epsilon)$ is a 2-manifold with boundary and ∂Y has k components.*

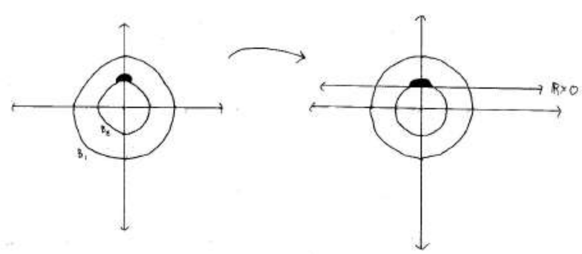
Remark 6.6. We call the space Y in the above proposition **X-with-k-holes**.

Proof. Note that because each h_i is a homeomorphism from the open unit ball to U_i, there are also homeomorphisms from B_ϵ to open subsets of the U_i. Because every point of X has a neighborhood homeomorphic to an open set of $\mathbb{R}^2$, every point of Y that is of distance $a > 0$ from any of the $h_i(B_\epsilon)$ also has a neighborhood homeomorphic to an open set in $\mathbb{R}^2$. For the remaining area of Y, we must show that there are certain points with a neighborhood homeomorphic to an open set in H^2, and these are the boundary points of Y. All other points then must have a neighborhood homeomorphic to an open set in $\mathbb{R}^2$.

Take $a < 1/2$. Then examining the a-neighborhoods around each $h_i(B_\epsilon)$ in Y is equivalent to examining the annulus $B_1 - B_\epsilon$, because the spaces $B_1 - B_\epsilon$ and $h_i(B_1 - B_\epsilon)$ are homeomorphic under the appropriate h_i, and the space $h_i(B_1 - B_\epsilon)$ is equivalent to the a-neighborhood around $h_i(B_\epsilon)$.

Thus our problem is reduced to showing that $B_1 - B_\epsilon$ is a 2-manifold with boundary. This follows from the argument given in example 6.4, but with a small clarification: B_1 is open, and thus every point has a neighborhood homeomorphic to an open set of $\mathbb{R}^2$. Because B_ϵ is also open, $B_1 - B_\epsilon$ contains the curve $\{(x, y) \mid x^2 + y^2 = 1/2\}$. Points on this curve, as illustrated in figure 3, are the points of $B_1 - B_\epsilon$ with no neighborhoods homeomorphic to an open set of $\mathbb{R}^2$, but homeomorphic to an open set of H^2. The homeomorphism pictured, as in figure 2, 'slides' all points above the boundary point below them down the appropriate distance (note that, in figure 3, the point under consideration is $(0, \epsilon = 1/2)$).

FIGURE 3. The boundary points of $B_1 - B_\epsilon$ have neighborhoods homeomorphic to open sets of H^2.



The space ∂Y has k components because all of the U_i are disjoint, and thus each one has a $h_i(B_\epsilon)$ inside of it which is disjoint from every other $h_i(B_\epsilon)$. Because there are k U_i, there must then be k components of the boundary of Y, each homeomorphic to the boundary of $B_1 - B_\epsilon$. □

Theorem 6.7 (Classification of Surfaces with Boundary)**.** *Given a compact connected triangulable 2-manifold Y with boundary such that ∂Y has k components, Y is homeomorphic to X-with-k-holes, where X is S^2 or the n-fold torus T_n or the m-fold projective plane P_m.*

Proof. The strategy of this proof is as follows: we want to (1) 'fill in' the holes of Y by some method, then (2) use the classification of surfaces theorem to show that the resulting space is homeomorphic to the sphere, n-fold torus, or m-fold projective plane, and then (3) 'remove the fillings' of our holes to complete the proof.

(1) The most effective way to 'fill in' the holes is with a homeomorphism h of each hole's boundary with S^1. Then, we include the unit circle into the unit disc with $i : S^1 \to B^2$. Then the map $h^{-1}(i(S^1))$ would 'fill in' our hole as desired. However, in order to show that this works, we must show that each component of ∂Y is homeomorphic to S^1. For this we need:

Lemma 6.8. *The boundary of a compact manifold with boundary is compact.*

Lemma 6.9. *A compact connected 1-manifold is S^1.*

Proof of lemmas. Let X be a compact manifold with boundary. Take a collection of open sets covering the boundary of X, and extend this to an open cover of X. This has a finite subcover, which gives us a finite subcover of the boundary. Thus our original open cover of the boundary has a finite subcover, so the boundary is compact.

Every point of a compact 1-manifold has a neighborhood homeomorphic with an open set of $\mathbb{R}$. But $\mathbb{R}$ is not compact. Thus we are dealing with a space in which every point has a neighborhood homeomorphic to an open set of $\mathbb{R}$, but which is bounded. The two candidates for this space are a circle or an open line segment. But an open line segment is not compact, and thus a compact connected 1-manifold must be S^1. This proves the two lemmas, and shows that our notion of 'filling in holes' is well-defined and legitimate. □

We must now show that the resulting space, after the holes of Y have been 'filled in', is a 2-manifold. That is, we must show that points from the original boundary of Y now have a neighborhood homeomorphic to an open set of $\mathbb{R}^2$. In other words, we must verify that the boundary components of Y and the boundaries of the 'hole-fillers' glue together continuously. Note that there exists, for any boundary point of either $h^{-1}(i(S^1))$ or Y, a neighborhood homeomorphic to an open set of H^2. Furthermore, each point of the boundary of each 'hole-filler' is identified with one point on the boundary of one of the boundary components of Y. Therefore, we have two neighborhoods, one on either side of each point of each boundary component, that are being glued together along the boundaries. This gluing is equivalent to a gluing along $\mathbb{R} \times \{0\}$ of two open sets in H^2, which creates an open set of $\mathbb{R}^2$. Thus, we have that every point formerly on the boundary of Y now has a neighborhood homeomorphic to an open set of $\mathbb{R}^2$, rendering the space Y-with-filled-in-holes a 2-manifold.

(2) Now, we have established that we can 'fill in' the holes of Y in order to create a 2-manifold. It follows trivially from the classification of surfaces theorem that the space Y-with-filled-in-holes is homeomorphic to either S^2, T_n, or P_m.

(3) We must now demonstrate that we can 'remove the fillings' of Y's holes without affecting the space to which it is homeomorphic, except by a corresponding removal of holes, leaving us with a space homeomorphic to X-with-k-holes, where X is either S^2, T_n, or P_m. In order to show this, recall that the original boundary components of Y are homeomorphic to S^1. Then we can remove the interiors of Y's original boundary components with a homeomorphism between these interiors and B^2, as we did above in order to 'fill in' the holes. In this case, we include a component of ∂Y into its interior, and map the result into B^2 via our homeomorphism. Clearly the resulting space is homeomorphic to X-with-k-holes, where X is either S^2, T_n, or P_m. This concludes the proof. □

Acknowledgments. It is a pleasure to thank my mentor, Jared Bass, for his time and efforts in guiding me during the researching and writing process for this paper. Without him it would not have been possible. I would also like to thank all of those at the University of Chicago who taught courses in the REU program this summer, for providing guidance and inspiration.

References

[1] James R. Munkres. *Topology.* Prentice-Hall, Inc. 2000.
[2] Allen Hatcher. *Algebraic Topology.* Cambridge University Press. 2001.