Commentaries for Mathematics 205B

The main reference for this course is the following text:

> **J. R. Munkres.** *Topology* (Second Edition), *Prentice-Hall, Saddle River NJ*, 2000. ISBN: 0–13–181629–2.

This following book will serve as a secondary textbook for the course:

> **A. Hatcher.** *Algebraic Topology* (Third Paperback Printing), *Cambridge University Press, New York NY*, 2002. ISBN: 0–521–79540–0.

This book can be legally downloaded from the Internet at no cost for personal use, and here is the link to the online version:

`www.math.cornell.edu/~hatcher/AT/ATpage.html`

This is the second course in the graduate level topology sequence. The book by Munkres is the default text for the first course, and an online set of lecture notes for the course material is available at the following address:

`http://math.ucr.edu/~res/gentopnotes2005.pdf`

The directory containing this file (`http://math.ucr.edu/~res/math205A`) also contains other files that might be useful for review as needed.

This commentary is a collection of remarks on various points related to the textbook sections that will be covered. In some cases the discussions will be very brief, but in others there will be alternate treatments of certain topics.

Munkres, Section 51

The notion of **homotopy** is central to this course. There are several reasons why it has become so important to topology. First of all, the concept yields effective methods for studying one of the most basic problems in topology; namely, determining whether or not two spaces are homeomorphic. Various considerations during the first two decades of the 20^{th} century showed that it was useful to introduce a weaker relationship called *homotopy equivalence*; in many situations there are good algebraic criteria for showing that certain pairs of spaces cannot be homotopy equivalent, and it follows that the spaces also cannot be homeomorphic. Furthermore, homotopy provides an extremely useful means for sorting the continuous mappings from one given space to another; for many of the most interesting examples of topological spaces, these sets of continuous mappings are uncountable topological spaces. In many contexts it is reasonable to think of two maps as being somehow equivalent if they are sufficiently close to each other in an appropriate sense, and the notion of homotopy may be viewed as one effective method for making this intuitive notion precise, yielding an equivalence relation (See Munkres, Lemma 51.1, p. 324) under which

(i) sufficiently close maps of suitably well-behaved spaces are homotopic,

(ii) the equivalence relation on the uncountable spaces of continuous functions yields countable sets of equivalence classes for such spaces.

The basic definitions of homotopies appear on page 323 of Munkres, and Lemma 51.1 on page 324 shows that this concept defines an equivalence relation on the set of continuous mappings from one space X to another space Y. Following standard usage, we shall denote the set of homotopy classes of maps from one space X to another space Y by $[X, Y]$.

Qualitative results on homotopy classes

In this discussion we shall assume that X is a compact subset of some Euclidean space $\mathbf{R}^m$ and Y is an open subset of some (possibly different) Euclidean space $\mathbf{R}^n$.

PROPOSITION. *Let X and Y be as above, and let $f : X \to Y$ be continuous. Then there is some $\varepsilon > 0$ such that if $g : X \to Y$ is a continuous mapping satisfying $\mathbf{d}(f, g) < \varepsilon$ (with respect to the uniform metric), then f and g are homotopic.*

Proof. We know that the image $f[X]$ is a compact subset of Y. For each $z \in f[X]$ there is some $\varepsilon_z > 0$ such that the open disk W_z of radius ε_z centered at z is contained in Y. Let ε be a Lebesgue number for the open covering of $f[X]$ by the sets W_z. It follows that if $z \in f[X]$ and $\mathbf{d}(z, y) < \varepsilon$, then the entire closed line segment joining z to y is contained in Y. Hence if $g : X \to Y$ is a continuous mapping satisfying $\mathbf{d}(f, g) < \varepsilon$ (with respect to the uniform metric), then the image of the straight line homotopy $H(x, t) \;=\; t\,g(x) \;+\; (1-t)\,f(x)$ is contained in Y. But this means that f and g are homotopic as mappings from X to Y.■

In principle, the preceding result shows that the homotopy relation is the equivalence relation generated by the binary relation $f \sim g$ if and only if for each $x \in X$ the line segment joining $f(x)$ to $g(x)$ lies entirely inside the open set Y. The next result shows that one has only countably many homotopy equivalence classes of mappings for X and Y as above.

PROPOSITION. *Let X and Y be as above. Then the set $[X, Y]$ of homotopy classes of continuous mappings from X to Y is a countable set.*

Proof. We shall use the Stone-Weierstrass Approximation Theorem (Rudin, *Principles of Mathematical Analysis*, Theorem 7.32, pp. 162–164) and the preceding result. More precisely, we shall prove that each continuous mapping f is homotopic to a mapping g whose coordinate functions are all given by polynomials in m variables. Since the set of all such maps is countable, it follows that the collection of all homotopy classes must also be countable.

Let $\mathcal{C}(X)$ denote the space of all continuous real valued functions on X; then the Stone-Weierstrass Theorem implies that the subalgebra $\mathcal{A}$ of all (restrictions of) polynomial functions on X is a dense subset. Given a continuous function $f : X \to Y$, denote its coordinate functions by f_j for $1 \leq j \leq n$.

By the previous result there is some $\varepsilon > 0$ such that $\mathbf{d}(f, g) < \varepsilon$ implies that f and g are homotopic, and in fact by the construction it follows that Y contains all points z such that $\mathbf{d}(\,f(x),\, z) < \varepsilon$ for some $x \in X$. By the observations of the preceding paragraph there are polynomial functions g_j such that

$$\mathbf{d}(f_j, g_j) \quad < \quad \frac{\varepsilon}{\sqrt{n}}$$

for each j, and it follows that the continuous function g with coordinate functions g_j maps X into U. Therefore we know that f is homotopic to a mapping g whose coordinate functions are given by polynomials.

To complete the argument, we need to show that g is homotopic to a mapping h whose coordinate functions are given by polynomials with rational coefficients. Let $\delta > 0$ be the number as in the preceding proposition, so that $\mathbf{d}(g, h) < \delta$ implies that G and h are homotopic and if $y \in \mathbf{R}^n$ satisfies $\mathbf{d}(y,\ f(x)\,) < \delta$ for some x then $y \in Y$.

Let d be the maximum degree of the coordinate functions g_j for g; then each g_j is uniquely expressible as a linear combination of monomials $\sum_\alpha\ b_{\alpha,j}\, x^\alpha$, where x^α runs through all monomials that are products of the fundamental indeterminates $x_1,\ \cdots\ , x_m$ such that $\deg(x^\alpha) \leq d$. Let A be the number of such monomials with degree $\leq d$, let M_α be the maximum of the monomial function x^α on X, and let M be the largest of these maxima M_α (where again the degree is $\leq d$).

If we now choose rational numbers $c_{\alpha,j}$ such that

$$|c_{\alpha,j} - b_{\alpha,j}| \quad < \quad \frac{\delta}{A \cdot M \cdot \sqrt{n}}$$

for all α and j, and we take $h = \sum_\alpha\ c_\alpha\, x^\alpha$, then a standard estimation argument as in 205A or real analysis shows that the rational polynomial functions h_j satisfy

$$\mathbf{d}(h_j, g_j) \quad < \quad \frac{\delta}{\sqrt{n}}$$

which in turn implies that $\mathbf{d}(g, h) < \delta$, so that h maps X into Y and g and h are homotopic as continuous mappings from X to Y.■

A simple variant of the preceding result is often useful. Given a topological space Y and a space U containing Y as a subspace, we shall say that Y is a *retract* of U if there exists a continuous mapping $r : U \to Y$ such that $r|Y$ is the identity. If we let j denote the inclusion of Y in U, the restriction condition can be rewritten as $r \circ j = \mathrm{id}_Y$; in other words, the mapping r is a left inverse to j. As in linear algebra, one-sided inverses to continuous maps are not unique; in topology it is customary to use the term *retraction* to denote a left inverse maps for a retract.

COROLLARY. *Suppose that X is a compact subset of some Euclidean space and Y is a retract of an open subset of some Euclidean space. Then the set of homotopy classes $[X, Y]$ is countable.*

Proof. Let $j : Y \to U$ be the inclusion of Y into the open subset U in some Euclidean space. Since $[X, U]$ is countable, it suffices to show that if f and g are continuous mappings from X to Y such that $j \circ f$ is homotopic to $j \circ g$, then f is homotopic to g. This is less trivial than it may seem; later on we shall see that if $i : Y \to Z$ is an arbitrary inclusion map then it is possible to have $i \circ f \simeq i \circ g$ even when f and g are not homotopic.

Suppose that $j \circ f \simeq j \circ g$, and let H be a homotopy from the first map to the second. Let $r : U \to Y$ be a retraction. Then the composite $r \circ H$ is a homotopy from $r \circ j \circ f$ to $r \circ j \circ g$. Since $r \circ j$ is the identity, the latter mappings are merely f and g respectively, and therefore $r \circ H$ defines a homotopy from f to g. By the comments in the preceding paragraph, this completes the proof.

SOME STANDARD TOPOLOGICAL SPACES. The **standard unit n-disk** D^n is the set of all points $x \in \mathbf{R}^n$ such that $|x| \leq 1$, and the **standard n-sphere** S^n is the set of all points $x \in \mathbf{R}^{n+1}$ such that $|x| = 1$; if $n = 2$, then S^1 is just the unit circle in the plane, and if $n = 1$ then D^1 is just the closed interval $[-1, 1]$.

We shall conclude this discussion with two results on homotopy classes of maps into these standard spaces.

PROPOSITION. *If X is an arbitrary topological space and D^n is as above, then the set of homotopy classes $[X, D^n]$ consists of a single point.*

Proof. It will suffice to show that every mapping from X to D^n is homotopic to the constant map whose value at each point is the zero vector. Let f be an arbitrary continuous mapping. By convexity we know that the closed line segment joining $f(x)$ to the zero vector lies in D^n, and therefore the image of the straight line homotopy $H(x,t) = (1-t)\,f(x)$ lies inside D^n, so that it defines a homotopy from f to the zero map.■

Complement. The same method immediately yields similar conclusions if D^n is replaced by an arbitrary convex subset of some Euclidean space (or Banach space, or real topological vector space).■

Our result for spheres is weaker but still very illuminating.

PROPOSITION. *If X is a compact subset of some Euclidean space and S^n is as above, then the set of homotopy classes $[X, S^n]$ is countable.*

Proof. We only need to check that S^n is a retract of an open subset of $\mathbf{R}^{n+1}$. But if $U = \mathbf{R}^{n+1} - \{\mathbf{0}\}$, then the map $r : U \to S^n$ sending x to $|x|^{-1}\,x$ is a continuous map whose restriction to S^n is the identity.■

Section 54 of Munkres contains a proof that $[S^1, S^1]$ is countably infinite, so in general the cardinality estimate in the proposition is the best possible.

Operations on curves

Given two continuous curves f and g from the unit interval $[0,1] \to X$ for some space X such that $f(1) = g(0)$, a binary operation called the product is defined on page 326 of Munkres. Intuitively speaking, if we are given curves f and g, then this operation describes the curve which behaves like f on the left hand half of the interval and which behaves like g on the right hand half. One can also think of this as "stringing together" the two curves, and for this reason the operation is also known as **concatenation**. Frequently it is also written using "+" rather than "$*$". On page 327, an construction assigning to each curve a *reverse curve* is also described. Intuitively speaking, this corresponds to a reparametrization going backwards from the final point to the initial point, and this construction on a curve f is often denoted by $-f$.

Two simple motivations for the preceding notation can be described using line integrals. Suppose that α and β are piecewise smooth curves into an open subset of some Euclidean space such that $\alpha(1) = \beta(0)$, and suppose that ρ is a continuous function defined on that open subset. Then the line integrals of ρ along α, β, $\alpha + \beta$, and $-\alpha$ can all be defined, and we have

$$\int_{\alpha+\beta} \rho(s)\, ds \quad = \quad \int_{\alpha} \rho(s)\, ds \; + \; \int_{\beta} \rho(s)\, ds \; , \qquad \int_{-\alpha} \rho(s)\, ds \quad = \quad - \int_{\alpha} \rho(s)\, ds \; .$$

POTENTIAL SOURCES OF DIFFICULTIES. Although the plus and minus notation for curves has some advantages, it also has some disadvantages that must be recognized. Ordinarily, when one uses a plus sign to denote a binary operation, the latter is commutative and associative, and moreover expressions like $\alpha + (-\alpha)$ turn out to be trivial. These are **not** the case for our constructions on curves. However, as noted in Munkres, one does have associativity up to homotopy, and expressions of the form $\alpha+(-\alpha)$ are trivial up to homotopy; on the other hand, one does **not** have commutativity up to homotopy, even if $\alpha(0) = \beta(1)$ and $\alpha(1) = \beta(0)$ so that both $\alpha+\beta$ and $\beta+\alpha$ can be constructed. It will take some time for us to describe examples illustrating the lack of commutativity up to homotopy, but eventually we shall do so.

Munkres, Section 52

This is much shorter than the commentary for the previous section.

The higher homotopy groups $\pi_n(X, x_0)$ mentioned in Munkres are defined and discussed in the first few pages of Section 4.1 in Hatcher's book; the material up to Proposition 4.1 on page 342 does not require any background aside from the sections in Munkres covered thus far. As noted on page 340 of Hatcher, the groups $\pi_n(X, x_0)$ are abelian for all $n \geq 2$; in contrast, we shall eventually prove that $pi_1(X, x_0)$ is not necessarily abelian.

Munkres concludes this section by showing that a homomorphism of topological spaces defines an isomorphism of fundamental groups. In fact, the fundamental groups of (X, x) and (Y, y) are isomorphic even if the two pointed spaces are related by the weaker concept of *homotopy equivalence* (see Munkres, Theorem 58.7, pp. 364–365). A homotopy equivalence of pointed spaces is a basepoint preserving map $f : (X, x) \to (Y, y)$ for which there is a *homotopy inverse* $g : (Y, y) \to (X, x)$ such that $g \circ f$ is basepoint preservingly homotopic to the identity on X and $f \circ g$ is basepoint preservingly homotopic to the identity on Y; a homeomorphism is automatically a homotopy equivalence, for we may take g to be f^{-1} and the composites will be homotopic to the identity mappings because they are in fact equal to identity mappings.

Similarly, one can define homotopy equivalences for topological spaces without basepoint, and using Exercise 1 on page 330 of Munkres we shall prove the following result(s).

THEOREM. *If $f : Y \to Z$ is a homotopy equivalence and X is an arbitrary topological space, then f defines an isomorphism from $[X, Y]$ to $[X, Z]$. A similar result holds for basepoint preserving homotopy equivalences of pointed topological spaces.*

The proof of this result will be a simple consequence of the following observation, which turns out to be a special case of some far-reaching general phenomena.

PROPOSITION. *Let X, Y, Z be topological spaces, and let $f : Y \to Z$ be continuous. Then there is a well-defined mapping of homotopy classes $f_* : [X,Y] \to [X,Z]$ such that if $v \in [X,Y]$ is represented by the function h, then $f_*(v)$ is represented by the function $f \circ h$. Furthermore, this construction has the following properties:*

(*i*) *If F is homotopic to f, then $F_* = f_*$.*

(*ii*) *If f is the identity mapping on Y, then f_* is the identity mapping on $[X,Y]$.*

(*iii*) *If $g : Y \to Z$ is another continuous mapping, then $(g \circ f)_* \; = \; g_* \circ f_*$.*

Similar results hold for basepoint preserving mappings of pointed spaces.

Proof. We shall only work the case of ordinary (unpointed spaces). The argument in the pointed case is similar; it requires an analog of the exercise in Munkres for pointed spaces, but it is straightforward to show that such an analog is valid.

Throughout the discussion below, v will denote an element of $[X,Y]$ and the notation $v = [h]$ will indicated that h is a representative for the equivalence class v.

The main point needed to justify the definition of f_* is to show that the construction $f_*(v)$ does not depend upon the choice of function representing v. In other words, if h and h' are homotopic, we need to know that $f \circ h$ is homotopic to $f \circ h'$; but this follows from the exercise in Munkres.

Property (*i*) also follows directly from the exercise in Munkres, and Property (*ii*) merely reflects the identity chain

$$v \quad = \quad [h] \quad = \quad [\mathrm{id}_Y \circ h] \quad = \quad (\mathrm{id}_Y)_*[h] \quad = \quad (\mathrm{id}_Y)_*(v) \ .$$

Finally, Property (*iii*) follows from another simple chain of identities:

$$(g \circ f)_*(v) \quad = \quad [g \circ f \circ h] \quad = \quad g_*([f \circ h]) \quad = \quad g_*\big(\, f_*(v)\big) \quad = \quad g_* \circ f_*(v) \ . \blacksquare$$

Proof of Theorem. Let $v = [h]$ as in the proposition, and likewise let $w = [k] \in [X,Z]$. By Properties $(i) - (iii)$ in the preceding proposition we have

$$g_* \circ f_*(v) \quad = \quad (g \circ f)_*(v) \quad = \quad (\mathbf{id}_Y)_*(v) \quad = \quad v$$

and similarly we have

$$f_* \circ g_*(w) \quad = \quad (f \circ g)_*(w) \quad = \quad (\mathbf{id}_Z)_*(w) \quad = \quad w$$

so that $g_* \circ f_*$ is the identity on $[X,Y]$ and $f_* \circ g_*$ is the identity on $[X,Z]$. But these conditions mean that f_* and g_* are inverse functions to each other, and hence they are both isomorphisms.$\blacksquare$

There are similar **dual results** which show that $[X,Y]$ does not change if we replace X be a space that it is homotopy equivalent to X. We shall merely state the results and leave the details to the reader as exercises.

THEOREM. *If $g : W \to X$ is a homotopy equivalence and Y is an arbitrary topological space, then f defines an isomorphism from $[X,Y]$ to $[W,Y]$. A similar result holds for basepoint preserving homotopy equivalences of pointed topological spaces.*$\blacksquare$

Here is the corresponding dual result which plays the key role in proving the dual theorem.

PROPOSITION. *Let W, X, Y be topological spaces, and let $g : W \to X$ be continuous. Then there is a well-defined mapping of homotopy classes $g^* : [X, Y] \to [W, Y]$ such that if $v \in [X, Y]$ is represented by the function h, then $g^*(v)$ is represented by the function $h \circ g$. Furthermore, this construction has the following properties:*

(*i*) *If G is homotopic to g, then $G^* = g^*$.*

(*ii*) *If g is the identity mapping on X, then g_* is the identity mapping on $[X, Y]$.*

(*iii*) *If $f : V \to W$ is another continuous mapping, then $(f \circ g)^* \;=\; g^* \circ f^*$.*

Similar results hold for basepoint preserving mappings of pointed spaces.∎

Munkres, Section 53

The first paragraph of this section of the text mentions two areas of mathematics in which covering spaces play an important role; there are also other areas outside of topology, including the theory of Lie groups and differential geometry. Analogs of covering spaces also arise in algebraic geometry.

IMPORTANT. Even though the word "covering" appears in the phrases "open covering" and "covering spaces," **there is no direct connection between the usages**; however, in practice this ambiguity usually does not cause any difficulties.

Background material

It is useful to introduce a simple, fundamental construction on topological spaces that does not receive much attention in standard textbooks but nevertheless plays a key role in the subject itself. This is the **disjoint union** construction, which is developed in Section V.2 of the online notes for 205A:

`http://math.ucr.edu/~res/gentopnotes2005.pdf`

Given a collection indexed of topological spaces X_α, the disjoint union $\amalg_\alpha X_\alpha$ is a space which is essentially the union of pairwise disjoint copies of the spaces X_α such that each such subset is both open and closed in $\amalg_\alpha X_\alpha$. If the spaces X_α are all the same space X and Λ is the indexing set, then the disjoint union is just the product of X with Λ, where Λ is taken to have the discrete topology. There are also numerous exercises on disjoint unions in the file

`http://math.ucr.edu/~res/gentopexercises2005.pdf`

and solutions appear in the files `solutions*.pdf`, where $* = 4$ or 5.

It might also be helpful to look at Section V.1 of the previously cited notes. This develops the notion of quotient topology in a manner slightly different from that of Munkres, and the approach in those note will probably be used at various points throughout the present course.

We shall now give some additional examples of covering space projections (= covering maps as defined in Munkres).

THE REAL PROJECTIVE PLANE. This space is denoted by $\mathbf{RP}^2$, and two equivalent constructions of it as a quotient space are described in Unit V and the accompanying exercises for the online 205A notes cited earlier. For our purposes here, it is convenient to think of $\mathbf{RP}^2$ as the quotient of S^2 by the equivalence relation which identifies $\mathbf{x}$ and $\mathbf{y}$ if and only if one of these unit vectors is ± 1 times the other. We claim that the quotient map from S^2 to $\mathbf{RP}^2$ is a covering space projection. It is possible to prove this directly (see Theorem 60.3 on page 372 of Munkres), but it will ultimately be more efficient to prove a general result which will yield larger classes of examples.

Unfortunately, we shall need to introduce some notation. The notion of a group action on a topological space is defined in Exercise 8 on page 199 of Munkres. For our purposes it will suffice to take a group and to view it as topological groups with respect to the discrete topology. If G is such a group and X is a topological space, the group action itself is given by a continuous mapping $\Phi : G \times X \to X$, with $\Phi(g, x)$ usually abbreviated to $g \cdot x$ or gx, such that $1 \cdot x = x$ for all x and $(gh) \cdot x \;=\; g \cdot (h \cdot x)$ for all g, h and x. One can then define an equivalence relation on X by stipulating that $y \sim x$ if and only if $y = g \cdot x$ for some $g \in G$, and the quotient space with respect to this relation is called the *orbit space* of the group action and written X/G. By the cited exercise in Munkres, this space is Hausdorff if X is.

If we are given a group action as above and A is a subset of X, then for a given $g \in G$ it is customary to let $g \cdot A$ (the translate of A by g) be the set $\Phi[\{g\} \times A]$; this is the set of all points expressible as $g \cdot a$ for the fixed g and some $a \in A$.

Definition. We shall say that a group action Φ as above is a **free action** (or G acts freely) if for every $x \in X$ the only solution to the equation $g \cdot x = x$ is the trivial solutions for which $g = 1$. — If $X = S^2$ as above and G is the order two subgroup $\{\pm 1\}$ of the real numbers (with respect to multiplication), then scalar multiplication defines a free action of G on S^2, and the quotient space is just $\mathbf{RP}^2$. Of course, there are also similar examples for which 2 is replaced by an arbitrary positive integer n, and in this case the quotient space $S^n/\{\pm 1\}$ is called *real projective n-space*.

The next result implies that the orbit space projections $S^n \to \mathbf{RP}^n$ are covering space projections.

THEOREM. *Let G be a finite group which acts freely on the Hausdorff topological space X, and let $\pi : X \to X/G$ denote the orbit space projection. Then π is a covering space projection.*

Proof. Let $x \in X$ be arbitrary, and let $g \neq 1$ in G. Then there are open neighborhoods $U_0(g)$ of x and $V_0(g)$ of $g \cdot x$ that are disjoint. If we let $W(g) = U(g) \cap g^{-1} \cdot V(g)$ is another open set containing x, while $g \cdot W(g)$ is an open set containing $g \cdot x$, and we have $W(g) \cap g \cdot W(g) = \emptyset$. Let

$$W \;=\; \bigcap_{h \neq 1} W(h)$$

so that W is an open set containing x.

We claim that if $g_1 \neq g_2$, then $g_1 \cdot W \cap g_2 \cap W \;=\; \emptyset$. If we know this, then it will follow immediately that $\pi[W]$ is an open set in X/G whose inverse image is the open subset of X given by $\cup_g \; g \cdot W$. This and the definition of the quotient topology imply that $\pi[W]$ is an evenly covered open neighborhood of x, and therefore it will follow that π is a covering space projection.

Thus it remains to prove the statement in the first sentence of the preceding paragraph. Note first that it will suffice to prove this in the special case where $g_1 = 1$; assuming we know this, in the general case we then have

$$g_1 \cdot W \cap g_2 \cdot W \quad = \quad g_1 \left(W \cap (g_1^{-1} g_2) \cdot W \right)$$

and the coefficient of g_1 on the right hand side is empty by the special case when $g_1 = 1$ and the fact that $g_1 \neq g_2$ implies $1 \neq g_1^{-1} \cdot g_2$. — But if $g \neq 1$ then we have $W \cap g \cdot W \subset W(g) \cap g \cdot W(g)$, and we know that the latter is empty by construction. Therefore $W \cap g \cdot W = \emptyset$, and as noted before this completes the proof.■

ANOTHER EXAMPLE. Define an action of the finite group $\mathbf{Z}_2$ on the torus $T^2 = S^1 \times S^1$ so that the nontrivial element $T \in \mathbf{Z}_2$ satisfies $T \cdot (z, w) = (-z, \overline{w})$ where S^1 is viewed as the set of unit complex numbers and the bar denotes conjugation. This is a free action because $T(z, w) = (z, w)$ would imply $z = -z$, and we know this is impossible over the complex numbers. In this case the quotient space is the **Klein bottle**.

STILL MORE EXAMPLES. Let $\mathbf{D}$ denote either the complex numbers or the quaternions, let d be the dimension of $\mathbf{D}$ as a real vector space, and let G be a finite subgroup of the group S^{dm-1} of elements of $\mathbf{D}$ with unit length. For example, if $\mathbf{D} = \mathbf{C}$ (the complex numbers), then G can be a cyclic group of arbitrary order, while if $\mathbf{D}$ is the quaternions then one also has some nonabelian possibilities, most notably the quaternion group of order 8 whose elements are given by ± 1, $\pm \mathbf{i}$, $\pm \mathbf{j}$, and $\pm \mathbf{k}$. If $\mathbf{D} = \mathbf{C}$ and $m > 1$, then the quotient spaces $S^{2m-1}/\mathbf{Z}_q$ (for $q > 1$) are the objects known as (simple) **lens spaces** (sometimes the case $q = 2$ is excluded because that quotient is the previously described real projective space); the reason for assuming $m > 1$ is that the corresponding quotient space for S^1 is homeomorphic to S^1. If $\mathbf{D}$ is the quaternions, G is the nonabelian quaternion group of order 8 described above and $m = 1$, then the space S^3/G is called the *3-dimensional quaternionic space form* associated to the group G.

In all examples of this type, for each point y in X/G the inverse image of $\{y\}$ in X consists of $|G|$ points, where $|G|$ is the order of G.

We shall compute the fundamental groups of these examples later in the course.

Composites of covering space projections

Exercise 4 (Munkres, p. 341) shows that under suitable restrictions the composite of two covering space projections is also a covering space projections. However, in general this is not necessarily true, and here is an example: Let X be a connected, locally arcwise connected space, and let $p : E \to X$ be a connected covering map that is nontrivial (not a homeomorphism). Let $Y = X \times X \times X \times \ldots$ be the countably infinite product of X with itself (with the product topology), let E^n denote the product of n copies of E with itself, and for an arbitrary space Y let $Y_n = E^n \times Y$. Define $p_n : Y_n \to Y$ by

$$(e_1, \ldots e_n; x_1, x_2, \ldots) \to (p(e_1), \ldots p(e_n); x_1, x_2, \ldots)$$

Then each map p_n is a covering map. Next, let $\widetilde{Z} = \amalg_{n\geq 1} Y_n$ and let Z be the countably infinite sum of Y with itself. Let $q = \amalg_n p_n : \widetilde{Z} \to Z$ and let $r : Z \to Y$ be the obvious projection. Then r and q are covering maps but that the composite rq is not a covering map.

The proof is only moderately difficult, but it is also a bit lengthy and requires input involving the product topology for infinite products, and therefore the proof will be left as an exercise [*Hint:* It suffices to show that basic open sets in the product topology are not evenly covered]. Another property involving covering spaces and composites appears in the first additional exercise for this section, and in a subsequent section we shall give yet another exercise with a sufficient condition under which the composite of two covering space projections is also a covering space projection.

Munkres, Section 54

DEFAULT HYPOTHESIS. From this point on, we shall assume that *all spaces are* **Hausdorff** *and* **locally path connected** *unless explicitly stated otherwise.*

One reason for concentrating on the proof that $\pi_1(S^1, z_0) \cong \mathbf{Z}$ is that the methods have far-reaching generalizations. This might not be apparent from the treatment in Munkres, so we shall indicated how the methods yield computations for the fundamental groups of many other spaces and show that one can obtain many different groups; in fact, **every** group can be realized as the fundamental group of some arcwise connected topological space, and for certain types of groups one can say more about the types of spaces that can realize them. In this commentary we shall show that every finitely generated abelian group is the fundamental group of a compact topological manifold (Hausdorff and every point has an open neighborhood that is homeomorphic to an open subset of some $\mathbf{R}^n$), and there are compact topological manifolds whose fundamental groups are nonabelian (see Section 36 and pp. 316–318 of Munkres for the definitions and basic properties of such spaces).

The first step is an abstraction of the proof that $\pi_1(S^1, z_0)$ is infinite cyclic. This will use the notion of **covering transformations** which is defined on page 487 of Munkres; specifically, if we are given a covering space projection $p : E \to B$, then the covering transformations are all homeomorphisms $h : E \to E$ such that $p \circ h = p$, and the set $\Gamma(p)$ of all covering transformations turns out to be a group with respect to composition of functions. Note that if $b \in B$, then the elements of $\Gamma(p)$ act as a group of permutations on the set (discrete space) $p^{-1}[\{b\}]$.

The proof of Theorem 54.5 in Munkres generalizes directly to yield the following fundamental result, which is Corollary 81.4 on page 489 of Munkres.

THEOREM. *Let $p : E \to B$ be a covering space projection, and assume that E is simply connected (i.e., its fundamental group is the trivial group). Let $b \in B$, and suppose that $G \subset \Gamma(p)$ acts uniquely transitively on $p^{-1}[\{b\}]$ (if $p(e_1) = p(e_2) = b$, then there is a unique $T \in \Gamma$ such that $T(e_1) = e_2$). Then $\pi_1(B, b) \cong \Gamma(p)$.*■

We now apply this to examples.

LENS SPACES AND THE QUATERNIONIC SPACE FORM. In these cases we have a sphere S^n where $n \geq 2$, and the spaces in question are quotients of the form S^n/G for suitable finite groups G; these spaces turn out to be compact topological manifolds (see the exercises), and by the theorem we have $\pi_1(S^n/G, y_0) \cong G$. These yield the following basic facts:

(i) Every finite cyclic group is the fundamental group of a compact topological manifold.

(ii) There is at least one nonabelian finite group which can be realized as the fundamental group of a compact topological manifold.

In fact, one can realize infinitely many nonabelian finite groups as in (ii). Specifically, in side the group of unit quaternions S^3, for each positive integer $k \geq 3$ the subgroup $Q(4k)$ generated by the complex number $\exp(2\pi\mathbf{i}/q)$ and $\mathbf{j}$ turns out to be a nonabelian group of order $4k$; the previously defined quaternionic group is merely the special case where $k = 2$ and the complex number is merely $\mathbf{i}$.

This leads immediately to the first general realization statement at the beginning of the commentary for Section 54:

PROPOSITION. *Every finitely generated abelian group is the fundamental group of a compact topological manifold.*

Sketch of proof. We already know this is true if the group in question is cyclic, and we also know that every finitely generated abelian group is a product of cyclic groups. Therefore it is enough to have some means for showing that the product of two realizable groups is also realizable. One step in this process is the following fairly simple fact. If M^m and N^n are topological manifolds of dimensions m and n respectively, then their product is a topological $(m+n)$-manifold, and this product is compact if M and N are compact (why is this true?). The other step is the following basic result:

PROPOSITION. *Let (X, x) and (Y, y) be pointed topological spaces, and let $p : X \times Y \to X$ and $q : X \times Y \to Y$ be projections onto the factors. Then the associated group homomorphisms p_* and q_* define a group isomorphism from*

$$\pi_1\big(\, X \times Y,\ (x, y)\,\big)$$

to $\pi_1(X, x) \times \pi_1(Y, y)$.

Sketch of proof. We know that p_* and q_* are group homomorphisms and that they define a group homomorphism from the fundamental group of $X \times Y$ to $\pi_1(X, x) \times \pi_1(Y, y)$. To see this map is 1–1, suppose that we have closed α and β in $X \times Y$ such that $p \circ \alpha$ and $p \circ \beta$ are base point preservingly homotopic, and likewise $q \circ \alpha$ and $q \circ \beta$ are base point preservingly homotopic. If K and L are the respective homotopies, let H be the unique continuous map into $X \times Y$ such that $p \circ H = K$ and $q \circ H = L$. Then H defines a base point preserving homotopy from α to β. — To see the group homomorphism is onto, note that if γ and δ are base point preserving closed curves into X and Y respectively, then there is a unique base point preserving curve $\xi : [0, 1] \to X \times Y$ whose coordinate functions are γ and δ respectively (formally, we have $p \circ \xi = \gamma$ and $q \circ \xi = \delta$).■

The Klein bottle turns out to be an example of a space whose fundamental group is infinite and nonabelian.

COMPUTATION OF THE FUNDAMENTAL GROUP OF THE KLEIN BOTTLE. We have constructed the Klein bottle as the base space of a 2-sheeted covering space projection $T^2 \to K$. If we compose this with the standard covering space projection from $\mathbf{R}^2$ to T^2, we obtain an infinitely sheeted covering space projection φ from $\mathbf{R}^2$ to K by Exercise 4 on page 341 of Munkres. By the theorem stated above, it is only necessary to observe that there is a subgroup of covering transformations for the covering space projection $\mathbf{R}^2 \to K$ which is infinite, transitive and nonabelian.

Let $e \in K$ be the image of $(1,1) \in T^2$, and view $\mathbf{R}^2$ as the complex numbers $\mathbf{C}$. Then $\varphi^{-1}[\,\{e\}\,]$ consists of all complex numbers having the form $\frac{1}{2}m \,+\, n\mathbf{i}$ where m and n are integers. If we let χ denote complex conjugation, then covering transformations for φ are given by

$$X(z) \;=\; \chi(z) \,+\, \tfrac{1}{2}\,, \qquad Y(z) \;=\; z \,+\, n\mathbf{i}$$

and one can check directly that the subgroup generated by these transformations is transitive on $\varphi^{-1}[\,\{e\}\,]$ Note that X^2 and Y generate the group of covering transformations for the torus covering $\mathbf{R}^2 \to T^2$, and this subgroup has index 2 in the group generated by X and Y.

It follows that the group G generated by X and Y is the fundamental group of the Klein bottle. A routine computation shows that

$$Y\,X\,Y^{-1}X^{-1} \;=\; Y^2$$

and from this we can conclude that $\pi_1(K, e)$ is infinite and not abelian.■

Munkres, Section 55

The main results of this section are the no-retraction theorem (55.2 on p. 348) and the Brouwer Fixed Point Theorem for D^2 (55.6 on p. 351). In this commentary we shall describe a slightly different approach to the latter which is more standard and does not involve vector fields. The key to this is the following result.

PROPOSITION. *Suppose that there is a continuous map from D^n to itself with no fixed points for some $n > 0$. Then S^{n-1} is a retract of D^n.*

Sketch of Proof. Suppose that $f : D^n \to D^n$ is continuous but $f(\mathbf{x}) \neq \mathbf{x}$ for all $\mathbf{x}$. For each $x \in D^n$ one can define the line joining $\mathbf{x}$ to $f(\mathbf{x})$, and from this one can also define the ray $[f(\mathbf{x})\,\mathbf{x}$ — which starts at $f(\mathbf{x})$ and passes through $\mathbf{x}$. If we draw a picture, it seems clear that this ray meets the boundary sphere S^{n-1} either at $\mathbf{x}$ or at some point past $\mathbf{x}$ on the ray; the latter means that the point has the form $f(\mathbf{x}) + t\big(\mathbf{x} - f(\mathbf{x})\big)$ where $t > 1$. Furthermore, it also seems that this intersection point should vary continuously with $\mathbf{x}$. If this is the case then the map $r : D^n \to S^{n-1}$ will be a continuous mapping whose restriction to the boundary sphere is the identity and thus we shall have a retraction from the disk to the boundary sphere. If $n = 2$ we know this is impossible, so the conclusion of the Brouwer Fixed Point Theorem follows in this case.■

We should also note that a similar argument proves the fixed point theorem when $n = 1$; in this case one cannot have a retraction from D^1 onto S^0 because the former is connected but the latter is not, and the continuous image of a connected space is connected.■

In order to make the preceding argument logically rigorous, we need to justify all the assertions regarding the intersections of rays with the boundary sphere and to verify that the construction is continuous. Since this is frequently "left as an exercise to the reader" in texts (for example, on page 32 of Hatcher) and the argument is somewhat lengthy despite its elementary nature, we shall include a formal statement and proof for the sake of completeness.

LEMMA. *There is a continuous function* $\rho : D^n \times D^n - \text{Diagonal} \to S^{n-1}$ *such that* $\rho(\mathbf{x}, \mathbf{y}) = \mathbf{x}$ *if* $\mathbf{x} \in S^{n-1}$.

If we have the mapping ρ and f is a continuous map from D^n to itself without fixed points, then the retraction from D^n onto S^{n-1} is given by $\rho(\mathbf{x}, f(\mathbf{x}))$.

Proof of the Lemma. It follows immediately that the intersection points of the line joining $\mathbf{y}$ to $\mathbf{x}$ are give by the values of t which are roots of the equation

$$|\mathbf{y} \; + \; t(\mathbf{x} - \mathbf{y})|^2 \;\; = \;\; 1$$

and the desired points on the ray are given by the roots for which $t > 1$. We need to show that there is always a unique root satisfying this condition, and that this root depends continuously on $\mathbf{x}$ and $\mathbf{y}$.

We can rewrite the displayed equation as

$$|\mathbf{x} - \mathbf{y}|^2 t^2 \; + \; 2\langle \mathbf{y}, \mathbf{x} - \mathbf{y} \rangle t \; + \; (|\mathbf{y}|^2 - 1) \;\; = \;\; 0 \; .$$

If try to solve this nontrivial quadratic equation for t using the quadratic formula, then we obtain the following:

$$t \;\; = \;\; \frac{-\langle \mathbf{y}, \mathbf{x} - \mathbf{y} \rangle \; \pm \; \sqrt{\langle \mathbf{y}, \mathbf{x} - \mathbf{y} \rangle^2 \; + \; |\mathbf{x} - \mathbf{y}|^2 \cdot (1 - |\mathbf{y}|^2)}}{|\mathbf{x} - \mathbf{y}|^2}$$

One could try to analyze these roots by brute force, but it will be more pleasant to take a more qualitative viewpoint.

(a) *There are always two distinct real roots.* We need to show that the expression inside the square root sign is always a positive real number. Since $|\mathbf{y}| \leq 1$, the expression is clearly nonnegative, so we need only eliminate the possibility that it might be zero. If this happens, then each summand must be zero, and since $|\mathbf{y} - \mathbf{x}| > 0$ it follows that we must have both $\langle \mathbf{y}, \mathbf{x} - \mathbf{y} \rangle = 0$ and $1 - |\mathbf{y}|^2 = 0$. The second of these implies $|\mathbf{y}| = 1$, and the first then implies

$$\langle \mathbf{y}, \mathbf{x} \rangle \;\; = \;\; |\mathbf{y}|^2 \;\; = \;\; 1 \; .$$

If we combine this with the Cauchy-Schwarz Inequality and the basic condition $|\mathbf{x}| \leq 1$, we see that $|\mathbf{x}|$ must equal 1 and $\mathbf{x}$ must be a positive multiple of $\mathbf{y}$; these in turn imply that $\mathbf{x} = \mathbf{y}$, which contradicts our hypothesis that $\mathbf{x} \neq \mathbf{y}$. Thus the expression inside the radical sign is positive and hence there are two distinct real roots.

(b) *There are no roots t such that* $0 < t < 1$. The Triangle Inequality implies that

$$|\mathbf{y} \; + \; t(\mathbf{x} - \mathbf{y})| \;\; = \;\; |(1 - t)\mathbf{y} \; + \; t\mathbf{x}| \;\; \leq \;\; (1 - t)|\mathbf{y}| \; + \; t|\mathbf{x}| \;\; \leq 1$$

so the value of the quadratic function

$$q(t) \;\; = \;\; |\mathbf{x} - \mathbf{y}|^2 t^2 \; + \; 2\langle \mathbf{y}, \mathbf{x} - \mathbf{y} \rangle t \; + \; (|\mathbf{y}|^2 - 1)$$

lies in $[-1, 0]$ if $0 < t < 1$. Suppose that the value is zero for some t_0 of this type. Since there are two distinct roots for the associated quadratic polynomial, it follows that the latter does not take a maximum value at t_0, and hence there is some t_1 such that $0 < t_1 < 1$ and the value of the function at t_1 is positive. This contradicts our observation about the behavior of the function, and therefore our hypothesis about the existence of a root like t_0 must be false.

(d) *There is one root of $q(t)$ such that $t \leq 0$ and a second root such that $t \geq 1$.* We know that $q(0) \leq 0$ and that the limit of $q(t)$ as $t \to -\infty$ is equal to $+\infty$. By continuity there must be some $t_1 \leq 0$ such that $q(t_1) = 0$. Similarly, we know that $q(1) \leq 0$ and that the limit of $q(t)$ as $t \to +\infty$ is equal to $+\infty$, so again by continuity there must be some $t_2 \geq 1$ such that $q(t_2) = 0$.

(d) *The unique root t satisfying $t \geq 1$ is a continuous function of $\mathbf{x}$ and $\mathbf{y}$.* This is true because the desired root is given by taking the positive sign in the expression obtained from the quadratic formula, and it is a routine algebraic exercise to check that this expression is a continuous function of $(\mathbf{x}, \mathbf{y})$.

(e) *If $|\mathbf{x}| = 1$, then $t = 1$.* This just follows because $|\mathbf{y} + 1(\mathbf{x} - \mathbf{y})| = 1$ in this case.

The proposition now follows by taking

$$\rho(\mathbf{x}, \mathbf{y}) \quad = \quad \mathbf{y} + t(\mathbf{x} - \mathbf{y})$$

where t is given as above by taking the positive sign in the quadratic formula. The final property shows that $\rho(\mathbf{x}, \mathbf{y}) = \mathbf{x}$ if $|\mathbf{x}| = 1$.■

The Fixed Point Property

We shall conclude this discussion with a few general remarks.

It is not difficult to construct examples of spaces X and continuous self maps $f : X \to X$ such that f does not have a fixed point. For example, translation by a nonzero vector in $\mathbf{R}^n$ has no fixed points, and if a group G acts freely on a space X as above, then the self homeomorphisms determined by the nonzero elements of G never have fixed points (this yields compact examples). We shall say that *a space X has the* **Fixed Point Property** if every continuous mapping from X to itself has a fixed point. A fair amount of work has been done to determine which spaces have this property (related comments appear in the paragraph on p. 351 of Munkres preceding Corollary 55.7). We shall limit outselves to proving an important fact which is asserted without proof in the derivation of Corollary 55.7 on pp. 351–352 of Munkres.

PROPOSITION. *If X has the fixed point property and Y is homeomorphic to X, then Y also has the fixed point property.*

Proof. Let $f : Y \to Y$ be continuous, and let $h : X \to Y$ be a homeomorphism. Then $g = h^{-1} \circ f \circ h$ is a continuous self map of X, and as such it has a fixed point, say x. A routine computation now shows that $y = h^{-1}(x)$ is a fixed point for f.■

Munkres, Section 56

Although this section of Munkres is not listed as part of the course coverage, some comments on its content (a proof of the Fundamental Theorem of Algebra) seem worthwhile.

As noted on pp. 353–354 of Munkres, there are many proofs of the Fundamental Theorem of Algebra, and ultimately they all require some input that is intrinsically nonalgebraic and involves the geometry or topology of the complex plane. In particular, one standard approach using the theory of functions of a complex variable is mentioned at the top of page 354.

If one looks carefully at the proofs of the Fundamental Theorem of Algebra in many complex variables texts, issues about the completeness of the arguments often arise. Usually these concern path independence properties of line integrals. A logically rigorous approach to these issues normally requires some information about homotopy classes of closed curves in open subsets of the plane (the same input which appears explicitly in Munkres' proof). This commentary provides the background needed to fill in the details that are sometimes omitted in books on complex variables.

One immediate complication involves the definition of an analytic function; in some references it is defined as a complex valued function f defined on an open subset $U \subset \mathbf{C}$ such that f' exists and is continuous on U, and in other references it is taken to be a function f for which f' exists, with no *a priori* assumption of continuity. In fact, the two notions are equivalent, for the existence of f' guarantees its continuity, but this is a nontrivial fact. We shall consider both cases here, beginning with the easier one in which f' is assumed to be continuous.

Suppose we know that f' exists and is continuous. Suppose that we are given a piecewise smooth (or, more generally, a rectifiable continuous) curve γ. Write the function f in the form $f = u + v\mathbf{i}$, where u and v are functions with continuous partial derivatives satisfying the Cauchy-Riemann equations. Then the line integral $\int_\gamma f(z)\,d(z)$ is equal to

$$\int_\gamma u\,dx \;-\; v\,dy \;+\; \mathbf{i}\cdot\int_\gamma v\,dx \;+\; u\,dy \;.$$

Assume now that the region U in the complex plane is rectangular with sides parallel to the coordinate axes (all $x + y\mathbf{i}$ such that $a \leq x \leq b$ and $c \leq y \leq d$). We claim that the given line integral depends only upon the initial and final points of γ. This is shown using corresponding results from multivariable calculus about path independence. By Green's Theorem, a line integral $\int_\gamma P\,dx + Q\,dy$ over a rectangular region is path independent if we have

$$\frac{\partial Q}{\partial x} \quad = \quad \frac{\partial P}{\partial Y}$$

and using the Cauchy-Riemann equations $u_x = v_y$, $u_y = -v_x$, we see that the displayed relation holds for the integrands in the real and imaginary parts of $\int_\gamma f(z)\,dz$. This leads to the following basic result:

PROPOSITION. *Let f be an analytic function on the open set $U \subset \mathbf{C}$ in the stronger sense (f' is continuous), and let α and β be continuous rectifiable curves in U with the same endpoints such that α and β are endpoint preservingly homotopic. Then $\int_\alpha f(z)\,dz = \int_\beta f(z)\,dz$.*

This follows directly from the corresponding result for multivariable calculus which is established in the following document:

`http:math.ucr.edu/∼res/math246B/coursenotes0101.pdf`

Although these notes are for a higher level course, the content of the cited portion does not require any background beyond that developed thus far in the present course.■

Suppose we know that f' exists but we are not given any information regarding its continuity. We can use the preceding approach PROVIDED we can show that if U is a rectangular region then

$\int_\gamma f(z)\,dz$ only depends upon its endpoints. This is done in many complex variables books; for example, it appears on pp. 109–115 of the book by Ahlfors, Section 9.2 of the book by Curtiss, and Section 2.3 of the book by Fisher, all of which are listed below:

L. V. Ahlfors. *Complex Analysis* (3^{rd} Ed.), *McGraw-Hill, New York*, 1979.

J. H. Curtiss. *Introduction to Functions of a Complex Variable* (Pure and Applied Math., Vol. 44). Marcel Dekker, New York, 1978.

S. D. Fisher. *Complex Variables* (2^{nd} Ed.), *Dover, New York*, 1990.

The notion of homotopy also leads to a definitive version of the Cauchy Integral Formula for an analytic function f defined near the complex number a:

$$f(a) \quad = \quad \frac{1}{2\pi\mathbf{i}} \cdot \int_\gamma \frac{f(z)}{z-a}\,dz$$

The point is that we can give and explicit description of the type of curve γ for which the formula is valid; namely, if f is defined on the open set U and $a \in U$, then we can take γ to be an arbitrary continuous rectifiable curve in $U - \{a\}$ which is homotopic to a counterclockwise circle of sufficiently small radius centered at a.

Munkres, Section 58

The textbook describes an important class of homotopy equivalences (deformation retracts). As noted at the beginning of Chapter 0 in Hatcher, such maps arise very naturally in topology, and by Corollary 0.21 on pp. 16–17 of Hatcher, two spaces X and Y are homotopy equivalent if and only if there is a third space W containing both of them as deformation retracts (see also the final paragraph on page 365 of Munkres). This result is based upon the *mapping cylinder* construction, which we shall discuss below. However, before doing so we shall state a variant of the result from Hatcher.

PROPOSITION. *Let $f : X \to Y$ be continuous. Then there is a topological space Z and continuous maps $g : X \to Z$, $h : Z \to Y$ and $j : Y \to Z$ such that* (1) *the maps j and g are homeomorphisms onto their images,* (2) *we have $h \circ g = f$,* (3) *the subspace $j[Y]$ is a deformation retract of Z. If X and Y are Hausdorff spaces, then one can also conclude that Z is Hausdorff.*

Corollary 0.21 in Hatcher shows that if f is a homotopy equivalence then we can also conclude that $g[X]$ is a deformation retract of Z.

ANALOG FOR POINTED SPACES. *If we have a base point preserving map, then one can find a pointed space Z' and base point preserving maps g', h', j' such that the conditions in the proposition are satisfied and $j'[Y]$ is a base point preserving deformation retract of Z'.*

The mapping cylinder construction is given on page 13 of Hatcher; specifically, if $f : X \to Y$ is continuous, then the mapping cylinder $M(f)$ is a quotient space of the disjoint union $X \times [0,1] \amalg Y$, where the equivalence relation is generated by stipulating that for each $x \in X$ the points $(x,1)$ and $f(x)$ are equivalent. If we are given pointed spaces (X, x_0) and (Y, y_0), then the pointed mapping cylinder is formed from the same disjoint union, but in this case the equivalence relation is generated by the previous condition and the further stipulation that for each $t \in [0,1]$ the points (x_0, t) and $f(x_0)$ are equivalent.

Before proceeding, we shall address one point not considered in Hatcher:

LEMMA. *If in the above settings the spaces X and Y are Hausdorff, then the ordinary and pointed mapping cylinders are also Hausdorff.*

The proof of this fact is left to the exercises.■

In the course of proving the proposition we shall also need a special case of Exercise 29.11.(a) on page 186 of Munkres: *If W is a topological space, $\mathcal{R}$ is an equivalence relation on W, and $\pi : W \to W/\mathcal{R}$ is the associated quotient mapping of spaces, then $\pi \times \mathrm{id}_{[0,1]}$ is a quotient map.* (In fact, the result is true if the unit interval is replaced by an arbitrary locally compact Hausdorff space.)

Proof of Proposition. We shall only do the unpointed case; the proof of the pointed case is similar and left as an exercise. Let $q : X \times [0,1] \amalg Y \to M(f)$ be the quotient space projection. The mapping g is defined by $g(x) = q(x, 0)$, and the mapping j is defined by $j(y) = y$. To define h, start by defining $\widetilde{h}$ from $X \times [0,1] \amalg Y$ to Y by $\widetilde{h}(x,t) = f(x)$ and $\widetilde{h}(y) = y$. Since $\widetilde{h}$ is constant on the equivalence classes $q^{-1}[\{w\}]$ for all $w \in M(f)$, there is a unique continuous map h such that $h \circ q = \widetilde{h}$. It is a routine exercise to verify that the constructed maps and $Z = M(f)$ have properties (1) and (2). To prove (3), note that $h \circ j = 1_Y$, so that we need to define a homotopy from $j \circ h$ to the identity on $M(f)$.

The first step in constructing this homotopy is to define a continuous map

$$\widetilde{H} : \big(X \times [0,1] \amalg Y\big) \times [0,1] \;\cong\; X \times [0,1] \times [0,1] \amalg Y \times [0,1] \to M(f)$$

by $\widetilde{H}(x,s,t) = q(x, s+t-st)$ and $\widetilde{H}(y,t) = q(Y)$. If a and b are points of $W = X \times [0,1] \amalg Y$ such that $q(a) = q(b)$, then we have $\widetilde{H}(a,t) = \widetilde{H}(b,t)$, so $\widetilde{H}$ passes to a continuous map on the quotient space of $W \times [0,1]$ given by $(a,u) \sim (b,v)$ if and only if $a \,\mathcal{R}\, b$ and $u = v$. By the exercise in Munkres cited above, this quotient space is merely $M(f) \times [0,1]$, and therefore we see that $\widetilde{H}$ passes to a homotopy of the restrictions to $M(f) \times \{0\}$ and $M(f) \times \{1\}$. By construction, these restrictions are the identity on $M(f)$ and $j \circ f$.■

Written version of a visual proof

Example 2 on page 362 of Munkres gives a visual "proof" that the figure 8 space $S^1 \vee S^1$ is a deformation retract of the doubly punctured plane $\mathbf{R}^2 - \{\mathbf{p}, \mathbf{q}\}$. For the sake of completeness we shall give a written argument which follows the steps suggested by Figure 58.2 on the cited page.

First of all, we need to take explicit models. Let p and q be the points $(0, \pm\frac{1}{2})$. Then the first step in the figure suggests that $D^2 - \{\mathbf{p}, \mathbf{q}\}$ should be a deformation retract of $\mathbf{R}^2 - \{\mathbf{p}, \mathbf{q}\}$. This is fairly simple to check. Let

$$r : \mathbf{R}^2 - \{p, q\} \;\longrightarrow\; D^2 - \{\mathbf{p}, \mathbf{q}\}$$

be the map which sends $\mathbf{x}$ to itself if $|\mathbf{x}| \le 1$ and to $|\mathbf{x}|^{-1} \cdot \mathbf{x}$ if $|\mathbf{x}| \ge 1$. If i_0 is the inclusion map of the doubly punctured disk into the doubly punctured plane, then $r \circ i_0$ is the identity, and the map $i_0 \circ r$ is homotopic to the identity by the straight line homotopy homotopy $H_0(\mathbf{x}, t) = t\mathbf{x} + (1-t)r(\mathbf{x})$ because the image of the latter lies in the doubly punctured plane.

In this, as in many other, cases, we say that the subspace is a *strong deformation retract* of the larger space because the homotopy satisfies $H_0(y, t) = y$ for all y in the subspace.

The second step in Figure 58.2 is to show that if E is the union of two closed disks

$$\{\, \mathbf{x} \in \mathbf{R}^2 \mid |\mathbf{x} - \mathbf{p}| \le \tfrac{1}{2} \,\} \quad \text{or} \quad \{\, x \in \mathbf{R}^2 \mid |\mathbf{x} - \mathbf{q}| \le \tfrac{1}{2} \,\}$$

(note that the intersection of the disks is the origin) then $E - \{\mathbf{p}, \mathbf{q}\}$ is a strong deformation retract of $D^2 - \{\mathbf{p}, \mathbf{q}\}$. In this case the definition of the retraction is more complicated and we must divide into cases, depending upon whether the first coordinate of $\mathbf{x} = (u, v)$ is nonnegative or nonpositive. Specifically, if $0 \le u \le 1$ then let $r(u, v) = (u, v)$ if $(u - \frac{1}{2})^2 + v^2 \le \frac{1}{4}$ and if the reverse inequality holds let

$$r(u, v) \quad = \quad \left(u,\ \mathrm{sign}(v) \cdot \sqrt{\tfrac{1}{4} \ - \ \left(u - \tfrac{1}{2}\right)^2} \right) \ .$$

(Although the function sign(v) is discontinuous at 0, a direct check shows that the function defined by the displayed formula turns out to be continuous.) If i_1 denotes the associated inclusion, then $r \circ i_1$ is the identity and once again there is a straight line homotopy from the identity to $i_1 \circ r$ which is the constant homotopy on $E - \{\mathbf{p}, \mathbf{q}\}$.

Finally, in the last step we need to show that the figure 8 space given by the union of the circles with equations $|\mathbf{x} - \mathbf{p}| = \frac{1}{2}$ and $|\mathbf{x} - \mathbf{q}| = \frac{1}{2}$ is a strong deformation retract of $E - \{\mathbf{p}, \mathbf{q}\}$. Once again the definition of the retraction splits into cases depending upon the sign of the first coordinate of $\mathbf{x}$. Specifically, if $\mathbf{x} = (u, v)$ satisfies $u \ge 0$, then

$$r(\mathbf{x}) \quad = \quad \mathbf{p} \ + \ \frac{1}{2\,|\mathbf{x} - \mathbf{p}|} \cdot (\mathbf{x} - \mathbf{p})$$

while if $u \le 0$ then

$$r(\mathbf{x}) \quad = \quad \mathbf{q} \ + \ \frac{1}{2\,|\mathbf{x} - \mathbf{q}|} \cdot (\mathbf{x} - \mathbf{q}) \ .$$

One can then check that this mapping is well-defined, its restriction to the figure 8 is the identity, and there is a straight line homotopy from the composite of inclusion following retraction to the identity on $E - \{\mathbf{p}, \mathbf{q}\}$.∎

Munkres, Section 59

The triviality of $\pi_1(S^n, x_0)$ for $n \ge 2$ was previously established in the exercises, but Theorem 59.1 is fundamentally important in its own right.

One additional corollary of Theorem 59.1 is important enough to be worth mentioning at this point.

PROPOSITION. *Suppose we are given a Figure 8 space $F = C_1 \cup C_2$, where C_1 and C_2 are each homeomorphic to S^1 and $C_1 \cap C_2$ consists of one point p. Then $\pi_1(F, p)$ is generated by the images of $\pi_1(C_1, p)$ and $\pi_1(C_2, p)$.*

Proof. Let U_1 be the union of C_1 with the open semicircle V_2 in C_2 centered at p, and define U_2 similarly. Then by construction $U_1 \cup U_2 = F$ and $U_1 \cap U_2$ is homeomorphic to a union of four half open intervals which share a common endpoint and nothing else. This set is contractible (hence simply connected), and hence $\pi_1(F, p)$ is generated by the images of $\pi_1(U_1, p)$ and $\pi_1(U_2, p)$.

It will suffice to show that C_i is a strong deformation retract of U_i where $i = 1$ or 2, for this will show that the images of $\pi_1(C_i, p)$ and $\pi_1(U_i, p)$ are equal. Now $\{p\}$ is clearly a strong deformation retract of the open semicircles V_j (which are homeomorphic to open intervals), and one can define a retraction and homotopy on each $C_i \cup V_j$ (where $j \neq i$) by taking the union of the constructions on V_j with the identity on C_i.■

Munkres, Section 60

Many of the computations in this section have already been done in the commentaries or exercises. We shall give alternate approaches to the examples of nonabelian fundamental groups. As in Munkres, it will be convenient to think of the Figure 8 space as the subset of all $(z, w) \in T^2$ such that either $z = 1$ or $w = 1$; in other words, it is the union of the circles $\{1\} \times S^1$ and $S^1 \times \{1\}$, which intersect in a single point.

LEMMA. *Let F be the Figure 8 space with base point $e = (1, 1)$, let (x, x) be an arbitrary pointed space, and suppose that we are given $a, b \in \pi_1(X, x)$. Then there is a continuous base point preserving map $\varphi : (F, e) \to (X, x_0)$ such that the image of φ_* contains the subgroup generated by a and b.*

If we know this, then we can prove $\pi_1(F, e)$ is nonabelian as follows: We know there are pointed spaces (X, x) such that $\pi_1(X, x)$ is nonabelian. Take any such space, and take an arbitrary pair of elements a, b such that $ab \neq ba$. By the lemma we know that $\pi_1(F, e)$ has a homomorphic image containing the nonabelian subgroup generated by a and b. Since all homomorphic images of an abelian group are abelian, it follows that $\pi_1(F, e)$ cannot be abelian.■

Proof of Lemma. (*Sketch*) Let α and β be continuous closed curves representing a and b respectively. Note that F is the quotient space of $E = \{0\} \times [0, 1] \cup [0, 1] \times \{0\}$ by the map h which sends (s, t) to $(e^{2\pi \mathbf{i}\, s}, e^{2\pi \mathbf{i}\, t})$; by construction, this map sends E onto F, and it is closed (hence it is a quotient map). Thus if we define φ_0 on E by $\varphi_0(s, t) = \big(\alpha(s),\, \beta(t)\big)$ then φ_0 will pass to a continuous map φ on F. Furthermore, if $u \in \pi_1(F, e)$ is the class of $(e^{2\pi \mathbf{i}\, s}, 1)$ and and v is the class of $(1, e^{2\pi \mathbf{i}\, t})$, then we have $a = \varphi_*(u)$ and $b = \varphi_*(b)$. This is what we wanted to prove.■

Fundamental group of the double torus

At the end of Section 60, Munkres mentions that the fundamental group of a certain space called the *double torus* (also known as the oriented surface of genus 2) has a nonabelian fundamental group. His discussion provides motivation and some figures which indicate how one might try to prove this result. We shall use similar ideas to provide a complete proof of the given result.

Our proof requires a slight modification of the construction in Munkres. We shall remove squares rather than round disks from the two copies of T^2. Specifically, let $J \subset S^1$ be the closed arc of length $\pi/36$ $(= 5°)$ whose endpoints are $(1, 0)$ and $(\cos \pi/36, \sin \pi/36)$, and let J_0 be the open arc defined by removing the endpoints. Denote the set of endpoints by K. Then the double torus is obtained from the disjoint union of two copies of

$$T^2_0 \quad = \quad T^2 \ - \ J_0 \times J_0$$

by identifying the closed subsets $\Gamma = K \times J \cup J \times K$ of the two copies in the obvious manner (*i.e.*, $(p, 1) \sim (p, 2)$ for all $p \in \Gamma$); note that $J \times J$ is homeomorphic to the square $[0, 1] \times [0, 1]$ and Γ corresponds to the closed curve on the boundary.

A pinching map from T^2_0 to T^2, which collapses Γ to a point, is mentioned at the bottom of page 374 in Munkres. Our next step will be to define a map in the setting above which has the required property (however, our construction does not map the complement of Γ homeomorphically to the complement of $\{ (1, 1)\}$; it is not 1–1 on $J \times S^1 \cup S^1 \times J$). This is done most easily by viewing S^1 as the quotient space $[0, 1]/0 \sim 1$. Let $h : [0, 1] \to [0, 1]$ be the continuous map which is zero on $[0, 1/36]$ and maps $[1/36, 1]$ to $[0, 1]$ by a 1–1 onto increasing linear function, and let $H : [0, 1] \times [0, 1] \to T^2$ be the following map:

$$H(s, t) \quad = \quad \left(e^{2\pi \mathbf{i}\, h(s)}, \, e^{2\pi \mathbf{i}\, h(t)} \right)$$

This passes to a map of quotients from T^2 to itself (viewing T^2 as the quotient of $[0, 1] \times [0, 1]$ with $(s, 0) \sim (s, 1)$ and $(0, t) \sim (1, t)$ for all s and t). Furthermore, the disjoint union of two copies of $H|T^2_0$ passes to a map of quotient spaces from $T^2 \# T^2$ to the one point union $T^2 \vee T^2$ formed by identifying the points $(1, 1)$ in each of the two pieces of $T^2 \amalg T^2$. One can now project from $T^2 \vee T^2$ to the analogous space $S^1 \vee S^1$ formed from $S^1 \amalg S^1$ by identitying the two points 1 in the pieces. Specifically, this map is given by taking the disjoint union of two copies of the first coordinate projection map $T^2 \to S^1$ and passing to quotients. Note that the space we call $S^1 \vee S^1$ is just the space we previously called F. Consider the closed curves θ_j (where $j = 1, 2$) given by the following composites:

$$S^1 \times \{1\} \quad \subset \quad T^2_0 \quad \subset \quad T^2 \# T^2 \quad \longrightarrow T^2 \vee T^2 \quad \longrightarrow \quad S^1 \vee S^1$$

There are two possible inclusions of T^2_0 into the double torus, which is a union of two closed subspaces that are homeomorphic to T^2_0, and the index j corresponds to choosing one of the inclusions.

By construction, the images of the generators of $\pi_1(S^1)$ under θ_{j*} are represented by the closed curves $\left(e^{2\pi \mathbf{i}\, h(s)}, 1 \right)$ and $\left(1, \, e^{2\pi \mathbf{i}\, h(t)} \right)$. These also represent the generators of the images of the two basic circles in $S^1 \vee S^1$, which we know are generators for $\pi_1(S^1 \vee S^1, e)$. Therefore we see that *the associated homomorphism from $\pi_1(T^2 \# T^2, e)$ to $\pi_1(S^1 \vee S^1, e)$ is* **surjective**. — This means that the nonabelian group $\pi_1(T^2 \# T^2, e)$ is isomorphic to a quotient group of $\pi_1(S^1 \vee S^1, e)$, and therefore the fundamental group of $T^2 \# T^2$ cannot be abelian.∎

FURTHER COMMENTS ON MUNKRES, CHAPTER 9

In one of the exercises we noted that the Cantor Set does not have the homotopy type of an open subset in some Euclidean space because it has uncountably many components. Since we also know that the fundamental group of an open subset in some Euclidean space is countable, it is natural to ask if one can also construct a compact subset of, say, the plane whose fundamental group is uncountable. An example of this sort (the shrinking wedge of circles, sometimes also known as the Hawaiian earring) is described in Chapter 1 of Hatcher (see Example 1.25 on pp. 49–50).

We have seen that the fundamental group of $\mathbf{R}^2 - \{\mathbf{0}\}$ is infinite cyclic and that every finitely generated abelian group can be realized as the fundamental group of a compact topological manifold. An example of an open subset in $\mathbf{R}^2$ with an infinitely generated fundamental group is given by taking the complement U of the set of all negative integers $\{-1, -2, \cdots\}$.

Here is one way of proving that $\pi_1(U, 1)$ is not finitely generated: View U as a subset of the complex plane, and let α_k denote the closed curve in U given by the counterclockwise circle with radius $(2k+1)/4$ and center $(3-2k)/4$, so that α_k meets the real axis at the points 1 and $(1-4k)/2$. Therefore each α_k defines an element a_k of the fundamental group of $\pi_1(U, 1)$. For each positive integer j let U_j denote the complement of $\{-j\}$, and let φ_j denote the map of fundamental groups determined by the inclusion of U in U_j followed by an isomorphism from $\pi_1(U_j, 1)$ to the integers $\mathbf{Z}$. It follows that $\varphi_j(a_k)$ is a generator if $j \leq k$ and trivial if $j > k$; this is true because the point $-j$ is inside the circle α_k if $j \leq k$ and outside the circle if $j > k$ (see the additional exercise for Section 56).

We may combine the preceding homomorphisms to define a homomorphism Φ from $\pi_1(U, 1)$ to a product $\prod^\infty \mathbf{Z}$ of countably infinitely many copies of $\mathbf{Z}$ (with addition defined coordinatewise); specifically, for all j, the j^{th} coordinate of Φ is φ_j. Since the homomorphic image of a finitely generated group is finitely generated, it will suffice to show that the image of Φ is not finitely generated. This final step is purely algebraic, and it depends upon the standard structure theorems for finitely generated abelian groups.

It is a routine exercise to check that if $\{G_\alpha\}$ is a family of groups such that the only elements of finite order are the identities, then the product $\prod_\alpha G_\alpha$ also has this property. Similarly, the product is abelian if each factor is abelian. Both of these properties carry over to subgroups, and in particular they apply to the image of Φ. Therefore, if the image of Φ is finitely generated, the structure theorem for finitely generated abelian groups implies that it is a direct sum of infinite cyclic groups, and as such it has some finite rank r. For each positive integer m, let H_m denote the subgroup generated by the classes $\varphi_j(a_j)$ for $j \leq m$. Then H_m is the subgroup of $\prod^\infty \mathbf{Z}$ consisting of all elements for which the p^{th} coordinate is zero for all $p > m$. This group has rank m; it follows that for each positive integer m the image of Φ contains free abelian subgroups of rank m. Since the image of Φ has no nontrivial elements of finite order, our finite generation assumption on the image of Φ implies that the latter is free abelian and has some fixed finite rank q. General considerations involving finitely generated abelian groups imply that every subgroup of the image of Φ is also free abelian and has rank at most q. This contradicts the final sentence in the preceding paragraph. The ultimate source of this contradiction is our assumption that the image of Φ is finitely generated. As noted before, this suffices to show that the fundamental group of U is not finitely generated.■

In fact, it is possible to show that every countably generated group can be realized as the fundamental group of an open subset in $\mathbf{R}^n$ if $n \geq 4$, but proving this would require methods and results which are outside the scope of this course.

LAST BUT NOT LEAST. Here are a couple of results that are important to know but have not yet been mentioned in the text or commentaries. First of all, a covering space projection is an open mapping; a proof of this fact appears near the end of `math205Bhints1.pdf`. Also, the following result could/should have been included earlier:

PROPOSITION. *Let $p : E \to B$ be a covering space projection, and assume that B is (nonempty and)connected. Then the cardinality of $p^{-1}[\{b\}]$ is the same for all $b \in B$.*

Proof. Define an equivalence relation on B by $x \sim y$ if and only if $p^{-1}[\{x\}]$ and $p^{-1}[\{y\}]$ have the same cardinality. By the definition of covering space, the equivalence classes of this relation are all open, and since they are pairwise disjoint it follows that they are also closed. Since B is connected, there can only be a single equivalence class.∎

Munkres, Section 61

The central objective of Chapter 10 in Munkres is to prove the Jordan Curve Theorem and some of its consequences. This result, which states that a simple closed curve in the plane separates the latter into two connected components, has been known empirically since at least the Late Stone Age (to quote the prominent topologist L. Siebenmann, it is "a fact that shepherds have relied on since time immemorial"), but known efforts to prove the general result mathematically only date back to the 19$^{\text{th}}$ century, and as noted in Munkres the first complete proof was published in 1905. Munkres also notes that the standard approach to proving this result is to view it a special case of a more general result (the *Jordan-Brouwer Separation Theorem*, which states that if $A \subset \mathbf{R}^n$ is homeomorphic to S^{n-1}, then A separates $\mathbf{R}^n$ into two components), and to prove the latter using another algebraic construction on spaces called **homology theory** (which is covered in Mathematics 246A and 246B).

One obvious motivation for Munkres' proof of the Jordan Curve Theorem is to illustrate how algebraic constructions like the fundamental group can be effectively used to answer "difficult questions concerning the topology of the plane that arise quite naturally in the study of analysis ... [for which the answers] seem geometrically quite obvious but turn out to be surprisingly hard to prove [Munkres, p. 376]." In particular, many of the arguments are extremely delicate. In these commentaries we shall describe simpler arguments which prove a conclusion that is weaker but yields the Jordan Curve Theorem for curves that are relatively well-behaved; in particular, it applies to broken line curves and more generally to curves that are piecewise smooth, but we cannot say *a priori* that it applies to objects such as *fractal curves* (an interesting topic which is outside the scope of this course sequence).

Locally flat curves

We shall begin by defining the restricted class of curves for our setting and indicating why this class contains all the "standard" examples.

Definition. Let X be a topological space, and let $\gamma : [0,1] \to X$ be a (continuous) parametrized curve. We shall say that γ is a *simple curve* if either γ is 1–1 or if $\gamma(t_1) = \gamma(t_2)$ and $t_1 \neq t_2$, then $\{t_1, t_2\} = \{0, 1\}$. If X is Hausdorff, this means that either γ maps $[0,1]$ homeomorphically onto its image or else γ passes to a continuous closed curve $\widetilde{\gamma}$ from $S^1 \cong [0,1]/0 \sim 1$ which maps S^1 homeomorphically onto its image.

In the first case, the curve is often called a *simple compact arc*, and in the second case the curve is often called a *simple closed curve.*

Definition. A parametrized curve $\gamma : [a,b] \to \mathbf{R}^n$ is said to be *locally flat* if for each $t \in [a,b]$ there is an open neighborhood U of $\gamma(t)$, an open interval J containing t, an open neighborhood V of the origin in $\mathbf{R}^{n-1}$, and a homeomorphism $h : U \to J \times V$ such that for all $s \in [a,b] \cap J$ we have $h \circ \gamma(s) = (s, 0)$.

One can use the Inverse Function Theorem to prove the following:

PROPOSITION. *If $\varepsilon > 0$ and $\gamma : (a-\varepsilon, b+\varepsilon) \to \mathbf{R}^n$ is a curve such that γ' always exists and a* **nonzero** *continuous function on the given (open) interval, then the restriction of γ to $[a,b]$ is locally flat.*

Sketch of proof. Let t be given, choose vectors $\mathbf{v}_2, \cdots, \mathbf{v}_n$ such that $\{\gamma'(t), \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ is a basis for $\mathbf{R}^n$, and define $g : (a-\varepsilon, b+\varepsilon) \times \mathbf{R}^{n-1} \to \mathbf{R}^n$ as follows:

$$g(x_1, \cdots, x_n) \quad = \quad \gamma(x_1) \; + \; \sum_{i=2}^{n} x_i \, \mathbf{v}_i$$

The g has continuous partial derivatives wherever it is defined, and the linear transformation $Dg(t, \mathbf{0})$ is an isomorphism because it takes the standard basis of $\mathbf{R}^n$ to $\{\gamma'(t), \mathbf{v}_2, \cdots, \mathbf{v}_n\}$. By the Inverse Function Theorem and the properties of the product topology there are open sets $J \subset (a-\varepsilon, b+\varepsilon)$ containing t and $V \subset \mathbf{R}^{n-1}$ containing $\mathbf{0}$ such that g maps $J \times V$ homeomorphically onto an open subset $U \subset \mathbf{R}^n$ (in fact, the inverse homeomorphism's coordinate functions have continuous partial derivatives, but we shall not need this). If h is the inverse homeomorphism, then for all $s \in J$ we have $h \circ \gamma(s) = (s, 0)$ as required.■

REMARKS. There is no assumption in the proposition that the curve γ is 1–1, but the conclusion implies that a curve satisfying the given conditions will be *locally* 1–1. The "lazy figure 8" curve $\gamma(t) = (\cos t, \sin 2t)$ is a curve which satisfies the conditions of the proposition but is not globally 1–1, even if one restricts to the open interval $(0, 2\pi)$; its values at $t = \frac{\pi}{2}$ and $t = \frac{3\pi}{2}$ are the same, but the tangent vectors for these parameter values are different. Many further examples can be constructed using other Lissajous curves of the form $(\cos p\,t, \sin q\,t)$, where p and q are positive integers.

Definitions. A *locally flat simple compact arc* in $\mathbf{R}^n$ is a simple compact arc which has a locally flat parametrization. A *locally flat simple closed curve* in $\mathbf{R}^n$ is a simple closed curve which has a *periodic* locally flat parametrization $\alpha : \mathbf{R} \to \mathbf{R}^n$ such that (i) the restriction to $[0,1]$ parametrizes the simple closed curve, (ii) for all integers n we have $\alpha(t+n) = \alpha(n)$. — Of course, given any parametrization $\gamma : [0,1] \to X$ of a simple closed curve there is a unique associated periodic parametrization $\gamma^* : \mathbf{R} \to X$.

If we define a *piecewise smooth* parametrization of a curve to be a function $\gamma : [0,1] \to \mathbf{R}^n$ for which there is a partition of $[0,1]$ into subintervals with endpoints $0 = t_0 < \cdots < t + m = 1$ such that each restriction $\gamma|[t_{i-1}, t_i]$ has continuous nonzero derivatives at all points, then one can in fact prove that *every curve with a piecewise smooth parametrization is locally flat.* For the sake of simplicity, we shall only state and prove this result for broken line curves (as defined in the exercises).

PROPOSITION. *If γ is a broken line curve in $\mathbf{R}^n$ and γ is either a simple compact arc or a simple closed curve, then γ is locally flat.*

Proof. By the definition of broken line curves, there is a parametrization γ for which there is a partition of $[0,1]$ into subintervals with endpoints $0 = t_0 < \cdots < t+m = 1$ such that each restriction $\gamma|[t_{i-1}, t_i]$ is a straight line segment. Local flatness at all interior points of the subintervals follows immediately from the construction (linear parametrizations are globally flat!), so it is only necessary to check what happens at the points t_i. In both cases (compact arcs and closed curves) it is necessary to consider the points t_i for which $i \neq 0, m$, and in the closed case it is also necessary to consider the cases where $i = 0$ or $i = m$.

CLAIM: If $i \neq 0, m$, then $\gamma(t_i) - \gamma(t_{i-1})$ and $\gamma(t_{i+1}) - \gamma(t_i)$ are **not** positive multiples of each other, and if γ is closed then $\gamma(t_m) - \gamma(t_{m-1})$ and $\gamma(t_1) - \gamma(t_0)$ are also not positive multiples of each other. — If the first of these happens, then there is a point $\gamma(u)$ such that $0 < u < 1$ such that some open neighborhood of $\gamma(u)$ in $\Gamma = \text{Image}\,(\gamma)$ is homeomorphic to a half-open interval, and if the second happens then there is a similar neighborhood of $\gamma(0) = \gamma(1)$. Each of these contradicts the assumption that γ maps $[0,1]$ or S^1 (in the separate cases) homeomorphically onto Γ.

We shall first prove that in both cases γ is locally flat at all points t_i such that $0 < i < m$; this will prove the result in the non-closed case, and in the closed case we shall need to modify the argument to show the periodic extension of γ is locally flat at 0 (equivalently, at 1).

There are two subcases, depending upon whether or not $\gamma(t_i) - \gamma(t_{i-1})$ and $\gamma(t_{i+1}) - \gamma(t_i)$ are **negative** multiples of each other. If the latter is true, then one can construct the homeomorphism h as follows: Construct an orthonormal basis for $\mathbf{R}^n$ by taking $\mathbf{u}_1$ to be a positive multiple of $\gamma(t_{i+1}) - \gamma(t_i)$, and extending this to an orthonormal basis for $\mathbf{R}^n$. Let $\mathbf{a} = \gamma(t_i)$, and let h_0 be the isometry which sends $\mathbf{a}, \mathbf{a}+\mathbf{u}_1, \cdots, \mathbf{a}+\mathbf{u}_n$ to the standard configuration $\mathbf{0}, \mathbf{e}_1, \cdots, \mathbf{e}_n$, where the $\mathbf{e}_j$ are the standard unit vectors. Define positive real numbers

$$c \quad = \quad \frac{t_i - t_{i-1}}{|\gamma(t_i) - \gamma(t_{i-1})|} \qquad d \quad = \quad \frac{t_{i+1} - t_i}{|\gamma(t_{i+1}) - \gamma(t_i)|}$$

respectively, and define a homeomorphism h_1 from $\mathbf{R}^n$ to itself by setting $h_1(x_1, \cdots, x_n) = (t_i + cx_1, x_2, \cdots, x_n)$ if $x_1 \geq 0$ and $h_1(x_1, \cdots, x_n) = (t_i + cd_1, x_2, \cdots, x_n)$ if $x_1 \leq 0$; note that the two definitions agree on the overlapping set where $x_1 = 0$. $t_1 > 0$. Then a restriction of $h_1 \circ h_0$ to a suitably small neighborhood of $\gamma(t_i)$ will have all the required properties.

In the remaining cases, the vectors $\gamma(t_i) - \gamma(t_{i-1})$ and $\gamma(t_{i+1}) - \gamma(t_i)$ are *linearly independent*. Let $\mathbf{v}$ and $\mathbf{w}$ be unit vectors which are positive multiples of $\gamma(t_i) - \gamma(t_{i-1})$ and $\gamma(t_{i+1}) - \gamma(t_i)$ respectively, and (if $n \geq 3$) choose orthonormal vectors $\mathbf{u}_3, \cdots, \mathbf{u}_n$ such that these vectors together with $\mathbf{v}$ and $\mathbf{w}$ form a basis for $\mathbf{R}^n$. Let $\mathbf{u}_2$ be a unit vector which is a positive multiple of $\mathbf{v}+\mathbf{w}$, and let $\mathbf{u}_1$ be a unit vector which is perpendicular to $\mathbf{u}_2$ and satisfies $\langle \mathbf{w}, \mathbf{u}_1 \rangle > 0$. — In order to justify the final statement, we need to check that if $\mathbf{y}$ is nonzero and perpendicular to $\mathbf{u}_2$, then the dot product of $\mathbf{y}$ and $\mathbf{w}$ is nonzero, but this can be checked as follows: By construction, no two of the vectors $\mathbf{v}$, $\mathbf{w}$ and $\mathbf{u}_2$ are scalar multiples of each other, for the the first two vectors are linearly independent third vector is a linear combination of $\mathbf{v}$ and $\mathbf{w}$ such that the coefficients of both are nonzero. By the Schwarz Inequality we then know that $-1 < \langle \mathbf{u}_2, \mathbf{w} \rangle < +1$, and therefore it follows that the absolute value of $\langle \mathbf{y}, \mathbf{w} \rangle$ is equal to

$$|\mathbf{y}| \cdot \sqrt{1 \; - \; \langle \mathbf{u}_2, \mathbf{w} \rangle^2} \quad > \quad 0 \; .$$

The figure in the document `flattening.pdf` illustrates the statements made in this paragraph and the next one.

We shall construct a piecewise linear homeomorphism from $\mathbf{R}^n$ to itself which is the identity on the subspace spanned by the vectors $\mathbf{u}_i$ for $i \geq 2$ and sends $\mathbf{w}$ and $\mathbf{v}$ to $\mathbf{u}_1$ and $-\mathbf{u}_1$ respectively.

As indicated in `flattening.pdf`, this map flattens the angle $\angle \mathbf{v0w}$ into the 1-dimensional vector subspace spanned by $\mathbf{u}_1$, and (as also indicated in the cited document) it is easier to describe the inverse, which is given as follows: If $\mathbf{x} = x_1\mathbf{u}_1 + x_2\mathbf{u}_2$, then $k(\mathbf{x}) = x_1\mathbf{w} + x_2\mathbf{u}_2$ if $x_1 \geq 0$ and $k(\mathbf{x}) = |x_1|\mathbf{v} + x_2\mathbf{u}_2$ if $x_1 \leq 0$; this map sends the x-axis to $\angle \mathbf{v0w}$, it maps the upper half plane to the interior of this angle, and it maps the lower half plane to the exterior of this angle. The two definitions reduce to the identity on the span of $\mathbf{u}_2$, each vertical line of the form $c\mathbf{u}_1 + t\mathbf{u}_2$ (where c is constant and t runs through all real numbers) is sent to itself, and the function is a translation on each vertical line. If we now take h_0 and h_1 as in the first case, then the homeomorphism $h = h_1 \circ k^{-1} \circ h_0$ will have the required properties. This completes the proof for simple arcs.

It remains to complete the proof in the case of simple closed curves at the initial and terminal point $\gamma(0) = \gamma(1)$. Given such a broken line curve γ, let γ^* be its periodic extension to a curve defined over the entire real line; next, let $\gamma_1 : [0,1] \to U$ be given by $\gamma_1(s) = \gamma^*(t_1 + s)$, where t_1 is the first partition point in the description of γ as a broken line curve. Then γ_1 is also a closed simple broken line curve, but we now have $\gamma_1(0) = \gamma_1(1) = \gamma(t_1)$, so by the preceding discussion we know that γ_1 is locally flat at $\gamma_1(1 - t_1) = \gamma(1) = \gamma(0)$. Since the parametrizations of γ and γ_1 are related by the elementary change of variables $s = t - t_1$, it follows that γ is locally flat at $\gamma(0) = \gamma(1)$.■

If we have a locally flat closed curve defined on $[0,1]$, then we know that it can be extended to the real line. In some situations it is helpful to have the following analogous fact for curves whose endpoints are not the same.

PROPOSITION. *Let W be an open subset of some Euclidean space $\mathbf{R}^n$. If $\gamma : [0,1] \to W$ is a locally flat curve, then there is some $\varepsilon > 0$ such that γ extends to a locally flat curve on $(-\varepsilon, 1+\varepsilon)$. Furthermore, if γ is $1-1$, it is possible to find a $1-1$ extension for some (potentially smaller) ε.*

Proof. Since γ is locally flat at 0, there is an open neighborhood U of $\gamma(0)$, an open interval J containing t, an open neighborhood V of the origin in $\mathbf{R}^{n-1}$, and a homeomorphism $h : U \to J \times V$ such that for all $s \in [0,1] \cap J$ we have $h \circ \gamma(s) = (s, 0)$. We might as well assume that the interval J has the form $(-\delta, \delta)$ for some $\delta > 0$. We may now extend γ to $(-\delta, 1]$ by setting $\gamma(s) = h^{-1}(s, 0)$ for $s \leq 0$; this and the original definition agree when $s = 0$, and therefore we have a locally flat extension of γ to some interval of the form $(-\delta, 1]$. A similar argument shows that we can extend γ further to a locally flate curve on some interval of the form $(-\delta, 1 + \delta')$ for a suitable choice of $\delta' > 0$. Finally, if we take ε to be the smaller of δ and δ', we obtain an extension of the original curve γ to $(\varepsilon, 1 + \varepsilon)$.

It remains to check that we can choose $\varepsilon' > 0$ so that the restriction to $(\varepsilon', 1 + \varepsilon')$ is 1–1. Since a locally flat curve is locally 1–1, it follows that there is some $\eta \in (0, \frac{1}{2}$ such that the restrictions of γ to $(-\eta, \eta)$ and $(1 - \eta, 1 + \eta)$ are both 1–1. Furthermore, since $\gamma(0) \neq \gamma(1)$ one can also choose η so that the images of these restrictions are disjoint. Let A be the image of γ restricted to $[\eta, 1 - \eta]$. Since the restriction of γ to the original interval $[0,1]$ is 1–1 and A is compact, there will be some $M > 0$ such that the images of γ restricted to $\left[-\frac{1}{M}\eta, \frac{1}{2}\eta\right]$ and $\left[1 - \frac{1}{2}\eta, 1 + \frac{1}{M}\eta\right]$ will be disjoint from A. If we now take $\varepsilon' = \frac{1}{M}\eta$, then the restriction of γ to $(\varepsilon', 1 + \varepsilon')$ will be 1–1.■

Complements of locally flat simple arcs

Our objective is to prove the Jordan Curve Theorem for locally flat closed curves. This will first require some information about the complements of locally flat simple arcs. In fact, we have the following strong conclusion.

COMPLEMENT THEOREM. *If $\Gamma \subset \mathbf{R}^2$ is a locally flat simple arc and we view $\mathbf{R}^2$ as a subset of S^2 in the usual fashion, then $S^2 - \Gamma$ is homeomorphic to $\mathbf{R}^2$.*

The proof of this result will require some preliminary machinery. We begin by using the preceding result to prove a strengthening of local flatness.

BOX LEMMA. *Let $\gamma : (a, b) \to \mathbf{R}^2$ be 1–1 and locally flat, and let $x_0 \in (a, b)$. Then there is a closed interval $(x_0 - \delta, x_0 + \delta) \subset (a, b)$ and a homeomorphism H from $[-2, 2] \times [-1, 1]$ onto a closed subset of $\mathbf{R}^2$ such that the image of $(-2, 2) \times (-1, 1)$ is open and we have the following:*

(i) The restriction of γ to $\left[x_0 - \frac{2}{3}\delta, x_0 + \frac{2}{3}\delta\right]$ is given by

$$H\left(\frac{3(t - x_0)}{2\delta} + x_0,\ 0\right) .$$

(ii) If $\eta \in \left(0, \frac{1}{2}(b - a)\right)$ and $x_0 \in [a + \eta, b - \eta]$ and K_η is the image of $[a + \eta, b - \eta]$ under γ, then the intersection of the image of H with K_η is equal to the image of $\left[x_0 - \frac{2}{3}\delta, x_0 + \frac{2}{3}\delta\right]$ under γ.

NOTE. The argument below also shows that the conclusions of (i) and (ii) remain true if we replace $\left[x_0 - \frac{2}{3}\delta, x_0 + \frac{2}{3}\delta\right]$ by an arbitrary closed subinterval $[y, z]$.

Proof of the Box Lemma. Since γ is locally flat and $x_0 \in (a, b)$, there is an open neighborhood U of $\gamma(x_0)$, an open interval J containing x_0, an open neighborhood V of the origin in $\mathbf{R}$, and a homeomorphism $h : U \to J \times V$ such that for all $s \in J$ we have $h \circ \gamma(s) = (s, 0)$. We might as well assume that $J = (x_0 - \delta, x_0 + \delta)$ for some $\delta > 0$ and that $V = (-\theta, \theta)$ for some $\theta > 0$. Let J_0 be the closed interval of length $\frac{4}{3}\delta$ centered at x_0.

Let B denote the compact set $K_\eta - J$, and let B' denote its image under γ. Since γ is 1–1, we can find an open neighborhood C of $\gamma[J_0]$ that is disjoint from B'. The image of this open neighborhood under h contains a set of the form $J_0 \times [-\theta', \theta']$ for some $\theta' \in (0, \theta)$. The desired homeomorphism H is constructed by first taking the homeomorphism from $\left[x_0 - \frac{2}{3}\delta, x_0 + \frac{2}{3}\delta\right] \times [-\theta', \theta']$ to $[-2, 2] \times [-1, 1]$ given by a product of linear maps on the two factors, and then composing with the inverse to the homeomorphism h mentioned in the first paragraph of this argument.■

The next result will yield the Complement Theorem for flat curves.

LOCAL SHRINKING LEMMA. *There is a continuous mapping φ from $\mathbf{R}^2$ to itself with the following properties:*

(i) The map φ onto, and it is the identity on the complement of the open set $(0, 2) \times (0, 2)$.

(ii) If π_2 denotes projection onto the second coordinate, then for all (x, y) we have $\pi_2 \circ \varphi(x, y) = y$.

(iii) The map φ is $1 - 1$ on the complement of the closed segment $C = [0, 1] \times \{1\}$, and it maps this compact set to $(0, 1)$.

(iv) The restriction of φ to $\mathbf{R}^2 - C$ maps the latter homeomorphically to $\mathbf{R}^2 - \{\mathbf{0}\}$.

A proof of this result is described in the file `shrinkmap.pdf`. In fact, the restriction in (iv) is the identity off a bounded set and hence if we extend the map to S^2 by sending the point at infinity to itself, we obtain a homeomorphism from $S^2 - C$ to $S^2 - \{pt\} \cong \mathbf{R}^2$.■

Proof of the Complement Theorem. Intuitively, the idea is simple. Given a simple locally flat arc Γ, we partition the arc into finitely many small pieces Γ_i and use the preceding results to show that the complement of $\cup_{i \leq k} \Gamma_k$ is homeomorphic to the shorter arc $\cup_{i<k} \Gamma_k$ for each $k > 1$, and at the final step when $k = 1$ we show that the complement of Γ_1 is homeomorphic to the complement of its left hand endpoint.

Here is a more formal version of the approach; at a few points some basically routine but tedious details are omitted and left to the reader. Extend γ to $(-\varepsilon, 1+\varepsilon)$ as before, and for each $x \in [0,1]$ let δ_x be as in the proposition above. Then there is a Lebesgue number ξ for the set of all intervals $\left(x - \frac{1}{3}\delta_x, x + \frac{1}{3}\delta_x\right)$ where $x \in [0,1]$, and choose a positive integer n such that $\frac{1}{n} < \frac{1}{2}\xi$. Then we can apply the Box Lemma (and the accompanying note) to find well-behaved compact subsets Q_k which are homeomorphic to products of two closed intervals and contain the intervals

$$\left[\frac{k}{n}, \frac{k+2}{n}\right] .$$

Let H_k denbote the homeomorphism from $[-2,2] \times [-1,1]$ to Q_k given by the Box Lemma, and let ψ be the homeomorphism from $[0,2] \times [0,2]$ to $[-2,2] \times [-1,1]$ sending (u,v) to $(2u-2, v-1)$ so that each map $H_k \circ \psi$ defines a homeomorphism ψ_k onto its image E_k and ψ_k maps the frontier of the square to the frontier of E_k. For each k such that $1 \leq k \leq n$ we may now define a continuous map from $\mathbf{R}^2$ to itself such that λ_k is given by $\psi_k \circ \varphi \circ \psi_k^{-1}$ on E_k and the identity otherwise. Less formally, λ_k corresponds to φ under the homeomorphism between $[0,2] \times [0,2]$ and E_k, and it is the identity at all other points. Once again, since λ_k is the identity off a bounded set, there is a unique continuous extension to S^2 given by sending the point at infinity to itself.

For all $k \geq 0$, let Γ_k denote the image of the restriction of γ to the interval

$$\left[\frac{k}{n}, \frac{k+1}{n}\right]$$

and it $k > 0$ let Σ_k denote the union $\Gamma_1 \cup \ \cdots \ \cup \Gamma_{k+1}$ for $k > 0$. By construction, if $k > 0$ then the map λ_k defines a homeomorphism λ'_k from $\mathbf{R}^2 - \Sigma_k$ to $\mathbf{R}^2 - \Sigma_{k-1}$, and if $k = 0$ then the map λ_0 defines a homeomorphism λ'_0 from $\mathbf{R}^2 - \Sigma_0$ to $\mathbf{R}^2 - \{\gamma(0)\}$; as before, one also obtains similar homeomorphisms if $\mathbf{R}^2$ is replaced by S^2. Since $\Gamma = \Sigma_{n-1}$ we may combine the preceding statements to conclude that $\mathbf{R}^2 - \Gamma$ is homeomorphic to $\mathbf{R}^2 - \{\gamma(0)\}$, and likewise that $S^2 - \Gamma$ is homeomorphic to $S^2 - \{\gamma(0)\}$. Since the latter is homeomorphic to $\mathbf{R}^2$, the conclusion of the Complement Theorem follows.∎

The separation theorem

The main result in Section 61 of Munkres is the following theorem.

THEOREM. *Let C be a locally flat simple closed curve in $\mathbf{R}^2 \subset S^2$. Then the complement of C in S^2 is* not *connected.*

Proof. View C as the image of a continuous map from S^1 to $\mathbf{R}^2$, and let C_+ and C_- be the images of the upper and lower semicircles. Then C_+ and C_- are locally flat simple arcs, and hence $S^2 - C_\pm$ is homeomorphic to $\mathbf{R}^2$.

We know that S^1 is not homeomorphic to S^2 because the former becomes disconnected if two points are removed and the latter does not. Thus C is a proper subset of S^2 and $S^2 - C$ has at least one component. Suppose that it has only one component. Since the sets $S^2 - C_\pm$ are arcwise connected by the Complement Theorem and

$$S^2 - C \quad = \quad S^2 - C_+ \quad \cap \quad S^2 - C_-$$

it would follow that

$$S^2 - (C_+ \cap C_) \quad = \quad S^2 - C_+ \ \cup \ S^2 - C_-$$

would be simply connected. However, $C_+ \cap C_-$ consists of two points, and $S^2 - \{x, y\} \cong \mathbf{R}^2 - \{v\} \cong S^1 \times \mathbf{R}$ is not simply connected, so we have a countradiciton. The source is our assumption that the complement of C is connected, and hence it must have at least two components.∎

At this point we only know that $S^2 - \Gamma$ has at least two components if Γ is a locally flat simple closed curve. In the commentary to Section 63 we shall prove that there are exactly two components, and that Γ is the set of frontier points for each of these components.

Munkres, Section 63

The following result is needed in our proof of the Jordan Curve Theorem.

LIMIT POINT LEMMA. *Let $n > 0$, let A be a nonempty compact subset of $\mathbf{R}^n$, and let C be a connected component of $\mathbf{R}^n - A$. If $\mathbf{Lim}(C)$ denotes the set of limit points for C in $\mathbf{R}^n$, then $\mathbf{Lim}(C) \cap A \neq \emptyset$.*

Proof of the Limit Point Lemma. Suppose to the contrary that $\mathbf{Lim}(C) \cap A = \emptyset$. Then we must have $\mathbf{Lim}(C) \cap A \subset \mathbf{R}^n - A$. Now the latter is open in $\mathbf{R}^n$ and hence is locally connected, so that each component C is both open and closed in $\mathbf{R}^n - A$ and also open in $\mathbf{R}^n$. Combining this with the previous conclusions, we see that $\mathbf{Lim}(C) \cap A$ is contained in C, which implies that C must be closed in $\mathbf{R}^n$. Since A is a nonempty proper subset of $\mathbf{R}^n$, this is a contradiction. The source of this contradiction is the hypothesis that $\mathbf{Lim}(C) \cap A = \emptyset$, and therefore the intersection must be nonempty as claimed.∎

For the sake of completeness, here is a statement of the main result.

JORDAN CURVE THEOREM. *Let $\Gamma \subset \mathbf{R}^2$ be homeomorphic to S^2. Then $\mathbf{R}^2 - \Gamma$ and $S^2 - \Gamma$ have exactly two components, and Γ is the frontier of each component.*

The component of $S^2 - \Gamma$ containing the point at infinity is called the *exterior region* determined by the curve, and the other component is called the *interior region* determined by the curve. There is an expected relation between the components of $\mathbf{R}^2 - \Gamma$ and $S^2 - \Gamma$. Namely, if U_1 and U_2 are the components of $S^2 - \Gamma$ and U_2 contains the point at infinity, then the components of $\mathbf{R}^2 - \Gamma$ are given by U_1 and $U_2 - \{\infty\}$ (see the first additional exercise for Section 61).

Proof. Since Γ is locally flat, for each $x \in \Gamma$ there is an open neighborhood U of x in $\mathbf{R}^2$ such that $U - \Gamma$ has two components, and in fact such neighborhoods form a neighborhood base at x. Furthermore, one can find a neighborhood base of this type such that if $V \subset U$ then the inclusion map determines a 1–1 correspondence from components of $V - \Gamma$ to components of $U - \Gamma$. It follows that for each $x \in \Gamma$ there are components $C_1(x)$ and $C_2(x)$ of $S^2 - \Gamma$ (which may be the same) such that for all sufficiently small neighborhoods W of x the set $W - \Gamma$ is contained in $C_1(x) \cup C_2(x)$. Thus if **SubCC** is the set of connected components of $S^2 - \Gamma$, we obtain a map of sets C from Γ to **SubCC**. The local flatness condition implies that this mapping is locally constant, and therefore it is continuous if Γ has the subspace topology and **SubCC** is given the discrete topology. Since Γ is connected, it follows that C is a constant map.

The reasoning of the preceding paragraph shows that there are connected components C_1 and C_2, which for all we know at this point may be the same, such that for some open neighborhood V of Γ we have $V - \Gamma \subset C_1 \cup C_2$. In order to complete the proof we must show that (i) $S^2 - \Gamma$ has no components other than C_1 and C_2, (ii) we cannot have $C_1 = C_2$.

If (i) is true, then (ii) follows because otherwise $S^2 - \Gamma$ would be connected and we know this is not the case, so everything reduces to proving (i). But suppose that C_3 is a component of $S^2 - \Gamma$. By the Limit Point Lemma there is a limit point y of C_3 which lies in Γ. By the definition of limit point, there is some basic open neighborhood W_0 of y such that W_0 contains a point z of C_3; since $C_3 \cap \Gamma = \emptyset$, this point must lie in $W_0 - \Gamma$. However, we know that the latter set is contained in $C_1 \cup C_2$ and since components are pairwise disjoint this means that C_3 must be equal to C_1 or C_2. This completes the proof that the complement has two components. The statement about frontier points is an immediate consequence of this fact and local flatness (observe that $t \in (a, b)$ implies that t is a limit point of both $(a, b) \times (0, h)$ and $(a, b) \times (-h, 0)$, and combine this with local flatness to see that every point of Γ is a limit point of C_1 and C_2).■

Concluding remarks

For most of the standard simple closed curves $\Gamma \subset S^2$, it is apparent (and sometimes very easy to check) that the interior regions in $S^2 - \Gamma$ are homeomorphic to open disks, and in fact the closures of these regions are homeomorphic to closed disks. A fundamental theorem of plane topology known as the **Schoenflies Theorem** proves this is true for arbitrary Γ (without a local flatness assumption). One proof of this result appears in Section IV.20 of the following book:

> **G. E. Bredon.** *Topology and Geometry*. Graduate Texts in Mathematics Vol. 139. Springer-Verlag, New York–*etc.*, 1993. ISBN: 0–387–97926–3

In particular, the Schoenflies Theorem implies that a simple closed curve in the plane is locally flat, and using this result one can also show that a simple compact arc in the plane is locally flat at all interior points of the interval. Similarly, one can extend the Complement Theorem to simple arcs which are not assumed to be locally flat; a discussion of the latter appears at the end of the file `complements.pdf`.

On the other hand, it is possible to construct arcs in $\mathbf{R}^n$ which are not locally flat for all $n \geq 3$. A 3-dimensional example is discussed on page 231 of the book by Bredon. Many other examples (including curves in $\mathbf{R}^n$ for $n \geq 4$) are contained in the following book:

> **T. B. Rushing.** *Topological Embeddings.* Pure and Applied Mathematics, Vol. 52. Academic Press, New York and London, 1973.

Munkres, Section 64

This section discusses some results on classes of topological spaces that are called *finite linear graphs* in Munkres and *finite edge-vertex graphs* on pp. 3–4 of `math205Bhints2.pdf`. There is a slight difference in these definitions; in Munkres it is assumed that two edges meet in just one endpoint, but in the other document the intersection is also allowed to be both vertices. In the first paragraph of page 395 in Munkres there is a passing comment that every object of the first type can be expressed as an object of the first type. We shall begin by justifying this assertion.

LEMMA. *Let Γ be a finite edge-vertex graph, and let $\mathcal{E}$ be the collection of edges determining the graph structure of Γ. Then there is another family of closed subsets $\mathcal{E}'$ such that the following hold:*

(i) *The family $\mathcal{E}'$ is a collection of edges for a possibly different graph structure on Γ.*

(ii) *Each element of $\mathcal{E}'$ is contained in a unique element of $\mathcal{E}$ such that one endpoint of $\mathcal{E}'$ is also an endpoint for $\mathcal{E}$ but another is not, and each edge in $\mathcal{E}$ is a union of two edges in $\mathcal{E}'$.*

(iii) *The intersection of two distinct edges in $\mathcal{E}'$ is a single point which is a common vertex.*

Proof. For each edge $E \in \mathcal{E}$, pick a point $b_E \in E$ that is not an endpoint. It follows that $E - \{b_E\}$ has two connected components, each of which contains exactly one endpoint of E. If x is an endpoint of E define the set $[x, E]$ to be the closure of the component of $E - \{b_E\}$ which contains x. If $\mathcal{E}'$ denotes the set of all such subsets $[x, E]$, then it follows immediately that $\mathcal{E}'$ has the properties stated in the lemma. Note that by construction the endpoints of a given edge $[x, E]$ are x and b_E.■

The family $\mathcal{E}'$ is frequently called the *derived* graph structure associated to $\mathcal{E}$.

As noted in Exercise 55.A4, many examples of edge-vertex graphs are suggested by ordinary letters and numerals. The main results in Section 64 of Munkres give examples of graphs which cannot be realized as subsets of $\mathbf{R}^2$.

Strictly speaking, the results in this commentary yield Theorems 64.2 and 64.4 in Munkres only for a restricted class of graphs; specifically, we need an analog of the local flatness conditions in Sections 61 and 63.

Definition. Let $\Gamma \subset \mathbf{R}^2$ be an edge-vertex graph with edge structure $\mathcal{E}$. We shall say that Γ is a *locally tame subset* or that *the embedding of Γ is locally tame* (with respect to $\mathcal{E}$) if the following hold:

(1) The restriction of Γ to an open edge (an edge with the endpoints removed) is locally flat.

(2) If x is a vertex of Γ and $\{ E_\alpha \}$ is the set of edges which contain x as an endpoint, then there is a homeomorphism h from an open neighborhood U of x to an open neighborhood V of $\mathbf{0}$ such that $h(x) = \mathbf{0}$ and there is a corresponding (finite) set of distinct rays S_α beginning at the origin such that for each α the image $h[E_\alpha \cap U]$ lies on S_α.

The results of Section 61 show that Γ is locally tame with respect to $\mathcal{E}$ if and only if it is locally tame with respect to $\mathcal{E}'$. Furthermore, if we construct a simple closed curve using edges from $\mathcal{E}$ (or the derived structure), then this simple closed curve will also be locally flat by the results of Section 61.

Alternate proof of Lemma 64.1

The proofs of Lemma 64.1, Theorems 64.2 and 64.4 in Munkres depend upon a result from Section 63 (Theorem 63.5) that was not covered in this course. We shall modify the approach in

Munkres to prove the first two of these results for locally tame edge-vertex graphs in the plane. In fact, the proof of Theorem 64.2 in Munkres will go through for locally tame graphs homeomorphic to the utilities graph if we can prove Lemma 64.1 for locally tame graphs in the plane which are homeomorphic to the Figure Theta graph, so this will be our key objective. Similar methods will yield alternate proofs of Lemma 64.3 (which analyzes spaces homeomorphic to the complete graph on four vertices) and Theorem 64.4 (which involves spaces homeomorphic to the complete graph on five vertices) for locally tame graphs in the plane, but we shall not do so in order to limit the length and complexity of the discussion.

Proof of Munkres, Lemma 64.1, for locally tame graphs. We adopt the same notation as on page 395 of Munkres: The space X is the union of three simple locally flat compact arcs A, B, C such that the intersection of any pair is the two point set $\{a, b\}$. Then the locally flat simple closed curve $A \cup C$ separates S^2 into two components U and V, and we shall assume that V contains the point at infinity (viewing S^2 as the one point compactification of $\mathbf{R}^2$). Let $A_0 = A - \{a, b\}$, and define B_0 and C_0 similarly.

STEP 1. This is a preliminary reduction of the proof to a special case. By definition, the set B_0 is a connected subset of $U \cup V$. We claim it suffices to consider the case where $B_0 \subset U$. If $B_0 \subset V$ instead, then let $c \in U$ be arbitrary, and consider the homeomorphism h from $S^2 = \mathbf{R}^2 \cup \{\infty\}$ to itself which sends z to $1/(z - c)$ if $z \neq c, \infty$ and switches these two exceptional points (one needs to verify this map is continuous, but it is elementary to do so). Then the complement of $h[A \cup C]$ has two components, the component $h[U]$ is the unbounded component, and $h[B_0]$ lies in the other component $h[V]$, which must be the bounded component. If we can prove the result for $h[X]$, then the result will also be true for X, and therefore there it is enough to prove the result when B_0 lies in the bounded component U of $S^2 - (A \cup C)$.

STEP 2. We shall analyze the relationship between the unbounded components of the complements of the closed curves $A \cup C$, $A \cup B$ and $B \cup C$. The condition $B_0 \subset U$ in the preceding paragraph implies that

$$S^2 - X = (U \cup V) - B = (U - B) \cup V$$

and therefore V is contained in the unbounded component of $S^2 - X$. This open connected set is contained in the unbounded components of $S^2 - (A \cap B)$ and $S^2 - (B \cap C)$, and hence V is contained in the unbounded components V_1 and V_2 of both $S^2 - (A \cup B)$ and $S^2 - (B \cup C)$.

Let U_1 and U_2 be the bounded components of the complements of $A \cup B$ and $B \cup C$ respectively. We want to prove that the sets U_1, U_2 and V are pairwise disjoint connected open sets whose union is $S^2 - X$.

STEP 3. We shall prove that small neighborhoods of points in $A_0 \cup C_0$ are contained in the union of X, U_1, U_2 and V. Suppose that $x \in A_0$. Then by the argument proving the Jordan Curve Theorem we know that x has a small open neighborhood W such that (i) $W - X$ has two components, one of which is contained in U and the other of which is contained in V, (ii) $W - X$ has two components, one of which is contained in U_2 and the other of which is contained in V_2. Suppose that P and Q are the two components of $W - X$, and let P be the component that is contained in V. Then $V \subset V_2$, so P is also the component of $W - X$ which is contained in V_2. It follows that the other component must be contained in U_2. Combining these observations, we obtain (iii) $W - X$ has two components, one of which is contained in U_2 and the other of which is contained in V.

Similar reasoning yields the following analog for C_0: If $x \in C_0$, then there is a small open neighborhood W of x such that $W - X$ has two components, one of which is contained in U_1 and the other of which is contained in V.

STEP 4. We shall need a similar result for B_0, but this will require some preparation. First, we claim that $U_2 \cup C_0 \subset V_1$ and $U_1 \cup A_0 \subset V_2$. To prove the first of these, use the previous step to show that every point of C_0 is a limit point of V, and therefore $V \cup C_0$ is a connected subset of $S^2 - (A \cup B)$; since V_2 is the maximal connected set containing V, it follows that $C_0 \subset V_1$. To prove that U_2 is contained in V_1, it will suffice to show that some point in U_2 can be connected to a point in V by a continuous curve whose image is disjoint from $A \cup B$. Let N be an open arcwise connected neighborhood of some point $z \in C_0$ such that N is disjoint from $A \cup B$ and $N - X$ has two components, one of which is contained in U_2 and one of which is contained in V. Clearly it is possible to connect some point in U_2 with some point in V by a continuous curve whose image lies entirely inside N, which is disjoint from $A \cup B$, and therefore it follows that U_2 and V must lie in the same component of $S^2 - (A \cup B)$. Similar considerations prove the second assertion at the beginning of this paragraph.

One important consequence of the preceding discussion is that U_1 and U_2 must be disjoint (since $U_2 \subset V_1$). Suppose now that $y \in B_0$. Then as in the previous step we know that there is an open neighborhood W of y such that $W - Y$ has two components and (i) one component is contained in U_1, (ii) one component is contained in U_2. Since U_1 and U_2 are disjoint, it follows that each of these sets contains one component of $W - X$.

STEP 5. We shall now use local tameness to give a similar analysis of sufficiently small neighborhoods of the points a and b in $A \cap B \cap C$. By this hypothesis, locally there are continuous changes of coordinates such that near a and b the graph looks like three rays emanating from the origin. Let N be such a neighborhood for either a or b. It follows immediately that $N - X$ is a union of three pairwise disjoint open connected subsets $Y_1 \cup Y_2 \cup Y_3$. More precisely, there is a homeomorphism taking a neighborhood of either a or b to a small disk centered at the origin so that the vertex in question is mapped to the origin, the set $N \cap A$ corresponds to the ray $\theta = 0$ (and $r \geq 0$), the set $N \cap B$ corresponds to the ray $\theta = \beta$ (and $r \geq 0$), and the set $N \cap C$ corresponds to the ray $\theta = \gamma$ (and $r \geq 0$), where $0 < \beta < \gamma < 2\pi$; the open sets Y_i then correspond to the connected open subsets defined by the inequalities $0 < \theta < \beta$ (for Y_1), $\beta < \theta < \gamma$ (for Y_2), and $\gamma < \theta < 2\pi$ (for Y_3). For $i = 1, 2, 3$ let $Y_i' \subset N$ correspond to Y_i under the continuous change of coordinates.

Passing to sufficiently small neighborhoods of a or b and the origin if necessary, by the preceding steps and the proof of the Jordan Curve Theorem we may assume that $N - (A \cup B)$ is contained in $U_1 \cup V_1$, $N - (B \cup C)$ is contained in $U_2 \cup V_2$, and $N - (A \cup C)$ is contained in $U \cup V$.

On the other hand, we know that

$$N \; - \; (A \cup C) \;\; = \;\; Y_1' \; \cup \; Y_2' \; \cup \; Y_3' \; \cup \; B_0$$

and its connected components are Y_3' and $Y_1' \cup Y_2' \cup B_0$. By the discussion above, this means that one of the latter sets is contained in U and the other in V. Since we know that $B_0 \subset U$, it follows that the first component is contained in V and the second in U; in other words, we have $Y_3' \subset V$. Similarly, we know that the connected components of $N - (A \cup B)$ are given by Y_1' and $Y_2' \cup C_0 \cup Y_3'$, with one of these sets contained in U_1 and the other in V_1. Since we already know that $Y_3 \subset V$ and $V \subset V_1$, it follows that Y_1' must be contained in U_1. Similar reasoning shows that Y_2' must be contained in U_2.

We may restate the conclusions of the preceding paragraph as follows: *The points a and b have open neighborhoods L and M such that $L - X$ and $M - X$ are contained in $U_1 \cup U_2 \cup V$.*

STEP 6. We now have enough machinery to prove the theorem. The preceding steps show that every point of X has an open neighborhood which is contained in

$$\Omega \;=\; X \,\cup\, U_1 \,\cup\, U_2 \,\cup\, V$$

and since the last three summands on the right hand side are open it follows that Ω is an open subset of S^2. Furthermore, since the closures of the sets U_i and V are contained in their unions with the compact subset X, it also follows that the closure of this union

$$\overline{\Omega} \;=\; X \,\cup\, \overline{U_1} \,\cup\, \overline{U_2} \,\cup\, \overline{V}$$

is contained in Ω and hence Ω is also closed in S^2. By connectedness we must have $\Omega = S^2$, and since the summands in the union expression are pairwise disjoint it follows that $S^2 - X$ is the union of the connected, pairwise disjoint open subsets U_1, U_2 and V. This proves that $S^2 - X$ has three components, and they are given by the three open subsets in the preceding sentence. The assertion about the frontiers of these open subsets follows immediately from their descriptions in terms of the locally flat closed curves $A \cup B$, $B \cup C$ and $A \cup C$.■

Embedding graphs in $\mathbf{R}^3$

In contrast to the results in Munkres, which show that some graphs cannot be realized as subsets of $\mathbf{R}^2$, we have the following general result in higher dimensions:

THEOREM. *If Γ is a finite linear graph in the sense of Munkres, then Γ is homeomorphic to a subset of $\mathbf{R}^3$. In fact, there is a homeomorphism such that each edge of Γ corresponds to a closed line segment in $\mathbf{R}^3$.*

Proof. The first step is to prove the following general fact: *There is a countably infinite set of isolated points in $\mathbf{R}^3$ such that no four are coplanar.* — The idea is simple; we can find a set of four points by simply taking a basis together with $\mathbf{0}$, and if we have a set S of n points with the given property, then we can obtain a set S' of the same type with one more point by picking some point which is not in the (finite) union of the planes determined by triples of vectors in S (this uses the fact that a finite union of planes in $\mathbf{R}^3$ cannot be all of $\mathbf{R}^3$, which follows because planes are nowhere dense in $\mathbf{R}^3$).

Label the points in the given set as $\mathbf{s}_1, \mathbf{s}_2, \cdots$ and label the vertices of Γ as $v_1, v_2, \cdots v_M$. By hypothesis, for each edge E there is a unique pair of distinct integers i and j such that the endpoints of E are v_i and v_j. For each $i < j$, let $E(i,j)$ denote the edge in $\mathcal{E}$ with vertices v_i and v_j, provided that there is an edge in $\mathcal{E}$ with the given vertices, and let $h_{i,j}$ denote a homeomorphism from $E(i,j)$ to $[0,1]$ such that v_i is mapped to 0 and v_j to 1. Define $f_{i,j}$ on $E(i,j)$ by

$$f_{i,j}(x) \;=\; h_{i,j}(x)\cdot\mathbf{s}_j \;+\; \big(1 - h_{i,j}(x)\big)\cdot\mathbf{s}_i \;.$$

In words, this map sends $E(i,j)$ to the segment joining $\mathbf{s}_i$ to $\mathbf{s}_j$ such that v_i and v_j are sent to $\mathbf{s}_i$ and $\mathbf{s}_j$ respectively. These mappings fit together to define a continuous map f from Γ to $\mathbf{R}^3$. We claim this mapping is 1–1; by construction it is 1–1 on each edge, so suppose that we have points y and z which go to the same point in $\mathbf{R}^3$. Suppose that y is on the edge whose vertices map to $\mathbf{a}$ and

b, and suppose that z is on the edge whose vertices map to **c** and **d**. It follows that the lines **ab** and **cd** meet at the point $f(y) = f(z)$. Since two intersecting lines are always contained in a plane (why?), it follows that **a**, **b**, **c** and **d** cannot be distinct (otherwise they would be noncoplanar). Interchanging the roles of **a** and **b** and of **c** and **d** if necessary, we might as well assume that **a** = **c**. If we also have **b** = **d**, then the two edges will be identical, and since f is 1–1 on an edge it will follow that $y = z$. On the other hand, if **b** and **d** are distinct, then the three points **a**, **b** and **d** must be noncollinear (if we throw in another point **e** from our infinite set we shall obtain a coplanar subset of four points). But this means that the two lines **ab** and **ad** only meet at a single point and this point must be **a**. In other words, in this case the relation $f(y) = f(z)$ implies that y and z are both endpoints of the distinct edges E and E'; since two distinct edges have at most one point in common, this means that $y = z$ as required.∎

Kuratowski's Theorem

At the end of Section 64, Munkres mentions a celebrated result of C. Kuratowski, which states that every graph which cannot be realized as a subset of $\mathbf{R}^2$ must contain a copy of either the utilities network or the complete graph on five vertices. Here is an online reference for the proof:

`http://cs.princeton.edu/~ymakaryc/papers/kuratowski.pdf`

FURTHER COMMENTS ON MUNKRES, CHAPTER 10

We shall discuss some additional results on the space which is studied in Munkres, Section 61 (the *closed topologist's sine curve*, also known as the *Polish circle*), and described more explicitly in `polishcircle.pdf`.

None of this material will be used subsequently in topics to be covered on examinations, so it can be skipped without loss of continuity. However, it does illustrate some approaches and methods that appear frequently in more advanced topology courses, using only material within the setting of this course and its prerequisite. At one point in the discussion we shall need a result that might not received much attention in 205A; namely, the *Tietze Extension Theorem*, which states that if A is a closed subset of a metric space X and $f : A \to \mathbf{R}^n$ is continuous, then f extends to a continuous function on X (see Theorem 35.1 on pp. 219–222 of Munkres).

We begin with a couple of basic observations.

PROPOSITION. *Let P be the Polish circle as described as in the references cited above, and let $V_n \subset \mathbf{R}^2$ be the open rectangular region*

$$\left(0,\ \frac{2}{(4n+3)\pi}\right) \ \times\ \left(-\frac{3}{2},\ \frac{3}{2}\right) \ .$$

Then $P - V$ is homeomorphic to a closed interval and hence is contractible.∎

This follows immediately from the construction, and the formal proof is left to the reader as an exercise.∎

PROPOSITION. *In the setting of the previous result, if K is a compact, connected and locally connected space, and $f : K \to P$ is continuous, then there is some $n > 0$ such that the image of f is contained in $P - V_n$.*

Sketch of proof. The main step involves the following standard observation: If $(x, 0) \in P$ such that $x \geq -1$, then for every sufficiently small open neighborhood W of $(x, 0)$ in P, the connected component of $(x, 0)$ in $P \cap W$ is contained in the y-axis. — This is the basic reason why the Polish circle is not locally connected.

Combining this with the local connectedness of K, we see that for every $y \in K$ there is an open neighborhood W_y and a positive integer $n(y)$ such that f maps W_y into $P - V_{n(y)}$. By compactness there is some $m > 0$ such that f maps K into $P - V_m$.■

COROLLARY. *If $x_0 \in P$, then $\pi_1(P, x_0)$ is trivial.*

Similar considerations show that if X is an arbitrary arcwise connected and locally arcwise connected space, the every continuous map from X to P is homotopic to a constant map.

Proof. If γ is a closed curve in P, then by the previous proposition we know that the image of γ is contained in a set of the form $P - V_m$ for some m. However, these sets are contractible by the first proposition above, and therefore the class of γ in the fundamental group of P must be trivial.■

In contrast, it turns out that P is not a contractible space. This will be an immediate consequence of the following result, which reflects the similarities between P and the standard circle S^1:

THEOREM. *If P is the Polish Circle, then there is a continuous map from P to S^1 which is not homotopic to a constant.*

COROLLARY. *The space P is not contractible.*

Prof. If P were contractible, then for every space Y, all continuous maps from P to Y would be homotopic to constant mappings.■

Proof of Theorem. We shall use the setting and terminology of `polishcircle.pdf` freely in the discussion below. Define a mapping r_1 from B_1 to the boundary G of the square with vertices $(1, -1)$, $(0, -1)$, $(0, -2)$, and $(1, -2)$ such that G sends (x, y) to $\big(x, m(y)\big)$, where $m(y)$ is the lesser of y and -1. By construction, for every positive integer n the restriction of r_1 to the simple closed curve C_n is onto, it is 1–1 off the set $\{1\} \times [-1, \sin 1)$, and it is constant on that exceptional interval. If we compose G with standard homeomorphisms $S^1 \cong C_n$ and $G \cong S^1$, we obtain a mapping g_n from S^1 to itself. Furthermore, if $\varphi : [0, 1] \to S^1$ is the usual map $\varphi(t) = \exp(2\pi \mathbf{i} t)$, then $g_n \circ \varphi$ is a map such that g_n is onto and there are points $a_n < b_n$ in the open interval $(0, 1)$ such that $g_n \circ \varphi$ is 1–1 on both $[0, a_n]$ and $[b_n, 1]$, while it is constant on $[a_n, b_n]$. Furthermore, this function is 1–1 on the complement of $[a_n, b_n)$.

CLAIM. *The mapping g_n is homotopic to a homotopy equivalence.* — Let $t_n \in \mathbf{R}$ be such that $g_n(1) = p(t_n)$, where p denotes the standard covering map from $\mathbf{R}$ to S^1, and let γ_n denote the unique lifting of $g_n \circ \varphi$ such that $\gamma_n(0) = t_n$. Since γ_n is 1–1 on $[0, a_n]$ it follows that it is either strictly increasing or decreasing there. We shall only consider the case where γ_n is increasing. In the other case, the curve $-\gamma_n$ is an increasing lifting of the complex conjugate curve $\overline{g_n}$, and by the increasing case we know that this conjugate curve is homotopic to a homeomorphism; taking conjugates, we see that g_n will also be homotopic to a homeomorphism.

So we assume that γ_n is strictly increasing on $[0, a_n]$. Since $g_n \circ \varphi$ is constant on $[a_n, b_n]$, it follows that the same is true for γ_n. Next, we claim that γ_n must be strictly increasing on $[b_n, 1]$. Since $g_n \circ \varphi$ is 1–1 on this interval, the same must be true for γ_n, which means that the latter is either strictly increasing or decreasing on the interval. If it were decreasing, this would contradict the previously stated injectivity properties of g_n. Therefore γ_n is nondecreasing and nonconstant, so there is a positive integer d such that $\gamma_n(1) = t_n + d$. If $d = 1$ then it will follow that g_n is homotopic to the identity, so it is only necessary to show that d cannot be greater than 1. But if this were the case, then there would be some $s \in (0, 1)$ such that $\gamma_n(s) = 1$ and hence if $z_s = p(s)$, then $z_s \neq 1$ and $g_n(z_s) = g_n(1)$. However, by construction the function g_n is 1–1 off the image of the subinterval $[a_n, b_n]$ under φ, and this image does not contain $g_n(1)$. Hence we see that d must be equal to 1, and as noted before this proves the claim.

Returning to the proof of the theorem, let r_n denote the restriction of r_1 to $B_n \subset B_1$, so that the previous discussion implies that $r_n|C_n$ is a homotopy equivalence. It follows immediately that for each n the map r_n cannot be homotopic to a constant mapping; if this were so then $r|C_n$ would be homotopic to a constant and the same would be true of the associated homotopy equivalence g_n from S^1 to itself. Since no such map is homotopic to a constant, the assertion regarding r_n follows.

We shall now prove that $r|P$ also is not homotopic to a constant mapping. Assume the contrary, and let H be a homotopy from $r|P$ to a constant map. Extend H to a continuous map H' on $P \times B_1 \times [0, 1]$ by letting H' be given by r on $B_1 \times \{0\}$ and by the appropriate constant map on $B_1 \times \{1\}$. Now apply the Tietze Extension Theorem to construct a continuous extension of H' to a continuous map K_0 from $\mathbf{R}^2 \times [0, 1]$ to $\mathbf{R}^2$. Let W_0 denote the inverse image of $\mathbf{R}^2 - \{\mathbf{0}\}$ with respect to K_0. Then W_0 is an open neighborhood of $\{0\} \times [-1, 1] \times [0, 1]$, and by the Tube Lemma it contains a subset of the form

$$\left[0, \frac{1}{(2k+1)\pi}\right] \times [-1, 1] \times [0, 1]$$

for some positive integer k. If $\mathbf{U}$ is the usual retraction from $\mathbf{R}^2 - \{\mathbf{0}\}$ to S^1 which sends $\mathbf{v}$ to $|\mathbf{v}|^{-1}\,\mathbf{v}$, then on the set $B_k \times [0, 1]$ the map $K(x, t) = \mathbf{U}\big(K_0(x, t)\big)$ defines a homotopy from r_k to a constant map from B_k to S^1.

The preceding sentence contradicts our earlier conclusion that r_k is not homotopic to a constant; the source of the contradiction is our assumption that $r|P$ is homotopic to a constant, and therefore this must be false and the assertion in the theorem must be true.■

Further results

As before, this material may be skipped without loss of continuity.

There is a construction which is dual to the fundamental group called the *Bruschlinsky group* that we shall now discuss.

Given a nonempty topological space X, define $\pi^1(X)$ to be the set of all homotopy classes $[X, S^1]$. Previous results and exercises show that the canonical map from $\pi_1(s^1, 1)$ to $\pi^1(X)$ is an isomorphism. By one of the exercises, the group structure on $\pi_1(S^1, 1)$ can be defined by taking the pointwise product of two closed curves in S^1, and more generally pointwise multiplication defines an abelian group structure on $\pi^1(X)$; specifically, if u and v are represented by $f : X \to S^1$ and $g : X \to S^1$, then $u \cdot v$ is represented by their pointwise product $f \cdot g$ (a little work is needed to show that the product is continuous and its homotopy class depends only on the homotopy classes of f and g). This construction has several properties that are analogous to those of the fundamental group.

(1) *If $h : Y \to X$ is a continuous map, then there is a homomorphism $h^* : \pi^1(X) \to \pi^1(Y)$ such that h^* takes the homotopy class of $f : X \to S^1$ to the homotopy class of $f \circ h$.*

(2) *In the preceding construction, if h_0 and h_1 are homotopic maps, then $h_0^* = h_1^*$.*

(3) *If h is the identity map on X, then h^* is the identity homomorphism. If $k : W \to Y$ is another continuous map, then $(h \circ k)^* = k^* \circ h^*$.*

(4) *If h is a constant map, then h^* is the trivial homomorphism.*

(5) *If "$\amalg$" denotes the disjoint union, then there is a canonical isomorphism from the group $\pi^1\big(X \amalg Y\big)$ to $\pi^1(X) \times \pi_1(Y)$ such that the algebraic coordinate projections correspond to the homomorphisms induced by the standard inclusions of X and Y in $X \amalg Y$.*

If $\alpha \in \pi_1(X)$ and $u \in \pi_1(X, x_0)$ with representatives f and γ, then the composite $f \circ \gamma$ defines an element of $[S^1, S^1] \cong \mathbf{Z}$, and this yields a canonical map assigning to α a homomorphism $K(\alpha) : \pi_1(X, x_0) \to \mathbf{Z}$. The set of such homomorphisms is an abelian group with respect to pointwise multiplication, and the mapping $\alpha \to K(\alpha)$ is a homomorphism with respect to this group structure. Basic results of algebraic topology state that the map K from $\pi^1(X)$ to the homomorphism group $\mathcal{H}(X, x_0)$ is an isomorphism if X is a sufficiently "nice" arcwise connected space (for example, a connected *cell complex* in the sense of Hatcher). On the other hand, the discussion above shows that K is not an isomorphism if X is the Polish circle (since $\pi_1(P, x_0)$ is trivial but $\pi^1(P)$ is not). In fact, if $f : P \to S^1$ is the map described in the theorem, then the argument proving the latter can be extended to show that the map

$$f^* : \mathbf{Z} \;\cong\; \pi^1(S^1) \;\longrightarrow\; \pi^1(P)$$

is an isomorphism.

Roughly speaking, the condition for X to be "sufficiently nice" is that each point should have a neighborhood base of open, contractible sets. Note that this condition holds for topological manifolds, edge-vertex graphs, and many of the other examples that have been considered in the course. isomorphism is that X should be arcwise connected.

Munkres, Section 67

Much if not all of the material in this section is covered in the Mathematics 201 graduate algebra sequence. A few comments on certain aspects of the material will appear in the commentary on the next section.

Munkres, Section 68

Given two groups H and K, the goal of this section is to construct the most general group which has subgroups isomorphic to H and K; in other words, we want a group G such that if L is any group which is generated by subgroups isomorphic to H and K, then L is a homomorphic image of G, and the homomorphism maps the subgroups generated by H and K in G to their counterparts in L.

One objective of the preceding section was to illustrate the basic idea in the case of **abelian** groups. Specifically, given two abelian groups H and K, one wants the most general abelian group A such that A contains isomorphic copies of H and K. This is fairly easy to do.

PROPOSITION. *Let G be an abelian group, let $i : H \to G$ and $j : K \to G$ be inclusion homomorphisms of abelian groups, and suppose that G is generated by the images $i[H]$ and $j[K]$. Then there is a unique surjective homomorphism $\varphi : H \times K \to G$ such that for all (h,k) we have $\varphi(h,k) = i(h) \cdot j(k)$.*

Verification of this result is a straightforward elementary exercise.∎

One can formulate a corresponding question for arbitrary indexed families of groups $\{ H_\alpha \}$. In this case the appropriate "universal example" is the **direct sum**

$$\bigoplus_{\alpha \in A} H_\alpha$$

which consists of all elements $\mathbf{x} \in \prod_\alpha H_\alpha$ such that all but finitely many of the coordinates x_α are the trivial elements of the respective groups. Once again, the formulation and the proof of an analog to the previous proposition are straightforward.

The **free product** is the corresponding construction for groups that are not necessarily abelian; in particular, even if the subgroups H_α are abelian this free product is supposed to be the most general group for **all** groups generated by isomorphic copies of the subgroups H_α, not just for all abelian groups.

A striking and fundamental result of A. G. Kurosh gives an elegant description of the subgroups of a free product. One proof of this result (using topological constructions as in this course) is given on pages 392–393 of J. Rotman, *An Introduction to the Theory of Groups* (Fourth Ed., Springer-Verlag, 1995).

Munkres, Section 69

The *free groups* in this section may be viewed as special cases of the examples in the preceding section in which the groups H_α are all infinite cyclic groups. The **universal mapping property** in Lemma 69.1 gives an important characterization of such groups in terms of homomorphisms.

Here is an alternate approach to a basic result on free groups.

THEOREM. *Let F be a group, and let X and Y be subsets of F such that X and Y are free generators for F. Then X and Y have the same cardinality.*

Proof. There are two cases, depending on whether or not one of the subsets is finite. Let $|X|$ and $|Y|$ be the respective cardinalities. If F is freely generated by X, then by Lemma 69.1 it follows that F admits exactly $2^{|X|}$ homomorphisms into the cyclic group $\mathbf{Z}_2$. Therefore we must have $2^{|X|} = 2^{|Y|}$. If either $|X|$ or $|Y|$ is finite, this means that the other is also finite and these cardinal numbers are equal. On the other hand, if the cardinalities are both infinite, then it is straightforward to check that the direct construction of a free product of $|X|$ cyclic groups has cardinality $|X|$. Therefore if both X and Y are infinite it follows that $|X| = |F| = |Y|$. — Note that the argument in the first case breaks down if X and Y are infinite because one can construct models for set theory in which two unequal cardinal numbers α and β satisfy $2^\alpha = 2^\beta$ (see the first unit of the 205A notes).■

As noted in Munkres, a subgroup H of a free group G is free, and by the preceding discussion we know that if $|G| \leq \alpha$ then every set B of free generators for H also satisfies this inequality (note that G must be infinite if it is a free group on one or more generators). In particular, if A is an infinite set of free generators for G, then we must have $|B| \leq |A|$. However, if G is a free group on n generators for some integer $n \geq 2$, then one cannot conclude that H has a set of generators B such that $|B| \leq n$. For example, a free group on two generators contains a subgroup which is a free group on an infinite set of generators.

At the end of this section, Munkres introduces the notion of a *finitely presented group.* The reason for interest in these groups is that they are precisely the groups that can be realized as fundamental groups for certain "nice" classes of spaces, including (i) compact topological manifolds of dimension ≥ 4, (ii) finite simplicial complexes of dimension ≥ 2 (see Hatcher, p. 107). We shall prove a closely related result in the commentary for the next section.

A curious example

Page 22 of Hatcher describes a property of the free product $\mathbf{Z}_2 * \mathbf{Z}_2$ that is not immediately obvious but turns out to be quite important in certain contexts. Namely, this group is isomorphic to the *infinite dihedral group* D_∞ which has two generators x and y such that $x^2 = 1$ and $xyx^{-1} = y^{-1}$. The element y generates an infinite cyclic subgroup which has index 2 and (hence) is normal. Every element of this group has a unique description as a product $x^\varepsilon y^n$, where n is an integer and $\varepsilon = 0$ or ± 1. The reason for the name involves the ordinary dihedral groups of order $2n$, where $n \geq 3$ is an integer. These groups are the subgroups of the group $\mathbf{O}(2)$ of 2×2 orthogonal matrices which send the standard regular n-gon — which has vertices $\exp(2\pi \mathbf{i} k/n)$, where $1 \leq k \leq n$ — to itself. Generators for this group are given by the matrix A which acts by counterclockwise rotation through an angle of $360/n$ degrees, and the matrix B which acts by reflection with respect to the x-axis. These two matrices satisfy the relation $BAB^{-1} = A^{-1}$, and in fact the dihedral group of order $2n$ is isomorphic to a quotient of D_∞ via the map sending x to B and y to A (its kernel is the subgroup generated by y^n). Note that if $n = 2$ the analogous group is just the Klein Four Group $\mathbf{Z}_2 \times \mathbf{Z}_2$.

Munkres, Section 70

The central objective of this section is to prove the Seifert-van Kampen Theorem, which describes how one can construct the fundamental group of a space X out of the fundamental groups of two arcwise connected open subsets U and V such that $U \cap V$ is also arcwise connected. If we let $i : U \to X$ and $j : V \to X$ denote the inclusion mappings and similarly define $i_0 : U \cap V \to V$ and $j_0 : U \cap V \to U$, then on fundamental groups we clearly have $j_* \circ i_{0*} = i_* \circ j_{0*}$, and this is a nontrivial constraint on the relation between the images of i_* and j_*. The Seifert-van Kampen Theorem states that the fundamental group of X is the most general group G for which we have homomorphisms $\alpha : \pi_1(U, x_0) \to \pi_1(X.x_0.$ and $\beta : \pi_1(V, x_0) \to \pi_1(X.x_0$ such that $\beta \circ i_{0*} = \alpha \circ j_{0*}$. More precisely, given any $(G, \alpha\beta)$ as above, *there is a unique homomorphism*

$$\Phi : \pi_1(X, x_0) \longrightarrow G$$

such that $\alpha = \Phi \circ i_*$ *and* $\beta = \Phi \circ j_*$.

In the language of category theory, one says that the triple $(\pi_1(X, x_0), i_*, j_*)$ is the **pushout** of the diagram associated to $(\pi_1(U \cap V, x_0), i_{0*}, j_{0*})$.

Constructing the abstract pushout is straightforward. We simply take the quotient of the free product $\pi_1(U, x_0) * \pi_1(V, x_0)$ by the normal subgroup N generated by all elements of the form

$$i_{0*}(y) \cdot \big(i_{0*}(y) \big)^{-1}$$

where $y \in \pi_1(U \cap V, x_0)$. It follows immediately that there is a unique homomorphism Φ from this pushout Γ to $\pi_1(X, x_0)$ such that Φ maps the images of $\pi_1(U, x_0)$ and $\pi_1(V, x_0)$ in Γ to their images in $\pi_1(X, x_0)$ via i_* and j_*. The difficult part is to show that Φ is an isomorphism. The proof takes nearly five pages in Munkres (pp. 426–430). At this point we shall merely assume this result is true for spaces that satisfy the conditions in the definition below, and we shall give an alternate proof for such spaces in the commentaries for Chapter 13.

Definition. An arcwise connected, locally arcwise connected topological space X is said to be *semilocally simply connected* if every point has at least one simply connected neighborhood.

This definition is slightly stronger that that of Munkres; both arise often in mathematical writings. Either holds for arcwise connected spaces that are *locally simply connected* in the sense that every point has an open neighborhood base of simply connected sets. Note that open sets in Euclidean spaces and edge-vertex graphs have this property (as do topological manifolds).

Generalizations of the Seifert-van Kampen Theorem

It is natural to ask if the Seifert-van Kampen Theorem can be extended to situations with weaker hypotheses. We shall discuss two possibilities:

(i) An arcwise connected compact Hausdorff topological space X with is a union of two closed arcwise connected subspaces $A \cup B$ where $A \cap B$ is also arcwise connected.

(ii) A semilocally simply connected Hausdorff topological space X which is a union of two semilocally simply connected open subsets U and V where $U \cap V$ is not necessarily (arcwise) connected.

In the first case, it is possible to construct examples such that conclusion of the Seifert-van Kampen Theorem does not hold. These are related to a subset $C \subset \mathbf{R}^3$ called the *Alexander Horned Sphere*. This set is homeomorphic to S^2, and $S^3 - C$ has two components U and V whose frontiers are equal to C; for the sake of definiteness, suppose that $\infty \in U$. Let A and B be the closed sets $U \cup C$ and $V \cup C$. Then $A \cap B = C$, and if the Seifert-van Kampen Theorem held for this example it would follow that the fundamental groups of A and B would both be trivial. However, the fundamental group of B is nontrivial. Further information about this example appears on pages 169–170 of Hatcher and also in the following online sites:

`http://mathworld.wolfram.com/AlexandersHornedSphere.html`

`http://en.wikipedia.org/wiki/Alexander_Horned_Sphere`

On the other hand, there are many situations in which a space X is a union of two arcwise connected closed subspaces A and B with arcwise connected intersection such that (i) the subsets A and B are deformation retracts of the open subsets U and V, (ii) the intersection $A \cap B$ is a deformation of the intersection $U \cap V$. Under these conditions the analog of the Seifert-van Kampen Theorem holds for the fundamental groups of X, A, B and $A \cap B$. This follows from the validity of the result for the fundamental groups of groups of X, U, V and $U \cap V$ and the fact that the inclusion maps $A \subset U$, $B \subset V$ and $A \cap B \subset U \cap V$ are all homotopy equivalences.

In the second case, there is a generalization which states that the *fundamental groupoid* of X (see Additional Exercise 52.9) is a suitably defined pushout of the diagram of fundamental groupoids

$$\Pi(U) \longleftarrow \Pi(U \cap V) \longrightarrow \Pi(V)$$

where the arrows are induced by inclusions of subspaces. One reference for this result is the following book:

R. (= Ronald) **Brown.** *Elements of Modern Topology.* McGraw-Hill, New York, 1968.

Extensively revised versions of this book also exist (one published by Ellis Horwood in 1988, and another by BookSurge in 2006). Since there is more than one R. Brown who has worked in algebraic topology during the past few decades, we note that the home page for the book's author is `http://www.bangor.ac.uk/~mas010/welcome.html` and the home page of the other topologist (Robert F. Brown) is `http://www.math.ucla.edu/~rfb/`. Both have written topology books of potential interest to graduate students.

Munkres, Section 71

One can use the results of this section to give another proof that the fundamental group of the genus two surface in Section 60 is not abelian. This follows because we can write down the fundamental group of the surface explicitly. We may view the torus as the quotient of $[0,1] \times [0,1]$ by the equivalence relation which identifies $(x,0)$ to $(x,1)$ and $(0,y)$ to $(1,y)$ for all x and y, and we may view the genus two surface as given by the union of two pieces U_1 and U_2 such that U_1 and U_2 are homeomorphic to the images of $[0,1] \times [0,1] - \{(\frac{1}{2}, \frac{1}{2})\}$ and their intersections are the images of the sets

$$\left\{(u,v) \in [0,1] \times [0,1] \mid (u - \tfrac{1}{2})^2 \ + \ (v - \tfrac{1}{2})^2 \ < \ \tfrac{1}{16} \right\}$$

where $V \subset U_1$ is identified with $V \subset U_2$ by the identity map. Now $S^1 \vee S_1$ is a deformation retract of U_i, and the generator of $\pi_1(V) \cong \mathbf{Z}$ maps to $xyx^{-1}y^{-1}$, where x and y are given by the two circles whose union is $S^1 \vee S^1$. By the Seifert-van Kampen Theorem we now have the following result.

THEOREM. *The fundamental group of the genus two surface X is isomorphic to the quotient of the free group on generators x_1, y_1, x_2, y_2 by the normal subgroup (normally) generated by the element $x_1 y_1 x_1^{-1} y_1^{-1} x_2 y_2 x_2^{-1} y_2^{-1}$, and it is nonabelian.*

The phrases *normal subgroup generated* and *subgroup normally generated* by a set X mean the subgroup generated by the union of the sets $\chi_g[X]$, where χ_g is the inner automorphism sending x to $g\,x\,g^{-1}$ and g runs through all the elements of G. By construction, this group is automatically a normal subgroup of G.

Proof. The fundamental groups of U_1 and U_2 are freely generated with free generators x, y and z, w respectively. By the comments preceding the statement of the theorem and the Seifert-van Kampen Theorem, it follows that the fundamental group of the surface is isomorphic to the quotient of the free product of these groups — which is just a free group on the four elements — by the normal subgroup normally generated by $xyx^{-1}y^{-1}(zwz^{-1}w^{-1})_{-1}$. The latter is equal to

$$xyx^{-1}y^{-1}wzw^{-1}z^{-1}$$

and the first conclusion of the theorem follows by labeling these free generators appropriately.

To prove the fundamental group is nonabelian, it suffices to show that it admits a homomorphism onto a free group on two generators. There is a standard surjective homomorphism from the free group on the x_i and y_i to the free group on x_1 and x_2 which sends the x_i to the obvious elements and sends the y_i to the identity. This will pass to a (surjective) homomorphism on the quotient group $\pi_1(X, u_0)$ if and only if it sends the element $x_1y_1x_1^{-1}y_1^{-1}x_2y_2x_2^{-1}y_2^{-1}$ to the identity. However, it is easy to check this is the case, and consequently it follows that the homomorphism from the free group on four generators to the free group on two generators does factor through the fundamental group of X.■

Munkres, Section 72

One important consequence of the main result of this section is the following result which was mentioned earlier.

THEOREM. *If G is a finitely presented group, then there is a space X such that X is obtained from a finite wedge of spheres by adjoining finitely many 2-cells (as in Munkres) and the fundamental group of X is isomorphic to G.*

Proof. Suppose that G is given by generators $g_1, \cdots, g_n$ and relations $r_1, \cdots, r_m$. Let X_1 be a wedge of n circles C_i, and choose closed curves γ_j representing the words r_j. Define a map $\mathbf{r}$ from $S^1 \times \{1, \cdots, m\}$ to X_1 whose restriction to $S^1 \times \{j\}$ is equal to γ_j for each j, and let Y_1 be the mapping cylinder of $\mathbf{r}$. Let $K_j \subset Y_1$ correspond to the circle $S^1 \times \{j\}$, take the disjoint union of Y_1 with $D^2 \times \{1, \cdots, m\}$, and identify the circle K_j in the first space with the circle $S^1 \times \{j\}$ in the second. Let B_j be the union of Y_1 with the first j disks.

If we now combine Theorem 72.1 in Munkres with an inductive argument to show that for each $k \leq m$, we see that the fundamental group of B_k is given by generators g_i with relations $r_1, \cdots, r_k$. If $k = m$ then $B_m = X$ and thus we have shown that the fundamental group of X is isomorphic to G.■

Since we have said we are particularly interested in spaces which are semilocally simply connected, we should note that the objects constructed above do have this property. Most of the work is contained in the following basic step.

LEMMA. *Suppose that X and Y are spaces that are Hausdorff, locally arcwise connected, and semilocally simply connected, and let $f; X \to Y$ be continuous. Then the (unpointed) mapping cylinder of f is also Hausdorff, locally arcwise connected, and semilocally simply connected,*

Knowing this, we can prove the spaces in the theorem are semilocally simply connected as follows: If we remove the center points $\{\mathbf{0}\} \times \{j\}$ from the disks, we obtain a set which is homeomorphic to an open subset in Y_1 and hence we have semilocal simple connectivity at all points except the center points of the disks. However, each center point has an open neighborhood homeomorphic to $\mathbf{R}^2$, and therefore the semilocal simple connectivity condition also holds at these points. For the sake of completeness, we should note that this set is compact by construction, and it is Hausdorff because it is a finite union of closed subspaces which are Hausdorff; namely, the mapping cylinder Y_1 (previously shown) and the disk subspaces homeomorphic to $D^2 \times \{j\}$.■

Proof of the Lemma. We have already shown that the mapping cylinder is Hausdorff if X and Y are. Suppose now that X and Y are locally arcwise connected, let $M(f)$ denote the mapping cylinder of f, and let $q : X \times [0,1] \amalg Y \to M(f)$ be the defining quotient map for $M(f)$. We know that q maps $X \times [0,1)$ homeomorphically onto an open subset of $M(f)$; this follows because $X \times [0,1)$ is open in $X \times [0,1] \amalg Y$ and each one point set in $X \times [0,1)$ is an equivalence class for the equivalence relation defining $M(f)$. Therefore we know that $M(f)$ is locally arcwise connected at all points coming from $X \times [0,1)$, and we only need to check that the same is true for points which come from Y.

The statement in the preceding sentence will be derived from the following more general result: *Suppose we are given a continuous mapping $f : X \to Y$ with mapping cylinder $M(f)$. Let q be the quotient projection from $X \times [0,1] \amalg Y$ to $M(f)$, let $y \in Y$, and let $\mathcal{V}$ be a neighborhood base for y in Y. Then a neighborhood base fo $q(y)$ is given by open sets of the form $q[V] \cup q[W]$, where $V \in \mathcal{V}$ and $W \subset f^{-1}[V] \times [0,1]$ is a saturated open subset (with respect to the equivalence relation associated to q) which contains $f^{-1}[\{y\}] \times \{1\}$ and is* upwardly vertically convex *— in other words, if $(x,t) \in W$, then $\{x\} \times [t,1] \subset W$.*

To prove the assertion in the previous paragraph, note that a typical open neighborhood of $q(y)$ is the image under q of a set having the form $V_0 \amalg W_0$, where V_0 is an open neighborhood of y in Y and W_0 is a saturated open subset of $X \times [0,1]$ containing $f^{-1}[\{y\}] \times \{1\}$. If $V \in \mathcal{V}$ is such that $y \in V \subset V_0$, then by intersecting the given open neighborhood with $f^{-1}[V] \times [0,1] \amalg V$ we obtain a new open neighborhood of the form $V \amalg W_1$, where W_1 is a saturated open subset of $X \times [0,1]$ containing $f^{-1}[\{y\}] \times \{1\}$ and contained in $f^{-1}[V] \times [0,1]$. Let $U_1 \subset X$ be chosen such that $U_1 \times \{1\}$ is the saturated set $W_1 \cap \times \{1\}$ (hence U_1 is open in X). For each $x \in U_1$ take an open neighborhood of $(x,1)$ having the form $N_x \times (\varepsilon_x, 1]$, where N_x is an open subset of U_1 containing x and $N_x \times (\varepsilon_x, 1] \subset W_1$, Let W be the union of all these open sets; observe that by construction we have $W \cap X \times \{1\} = U_1 \times [0,1]$. This open set fulfills the conditions in the preceding paragraph and it is contained in the image of $V \cup W_1$, and thus we have shown that $q(y)$ has a neighborhood base of the type described above.

Completion of the proof of the lemma. Suppose now that X and Y are locally path connected and $y \in Y$. Then we know that y has a neighborhood base $\mathcal{V}$ consisting of arcwise connected open subsets. Consider a basic neighborhood of y as in the previous paragraph for this choice of neighborhood base $\mathcal{V}$. By upward vertical convexity it follows that every point in such a neighborhood lies in the same arc component as a point in $q[V]$, where $V \in \mathcal{V}$. On the other hand, we know that V is arcwise connected, so there is only one arc component in the basic open neighborhood of $q(y)$. Therefore $M(f)$ is locally arcwise connected.

Finally, we need to show that $M(f)$ is semilocally simply connected if X and Y are. Once again, this follows immediately for points in the image of $X \times [0,1)$ under Q, for if (x,t) is such a point and U is a simply connected neighborhood of x, then $U \times [0,1)$ is a simply connected neighborhood of (x,t). Thus it only remains to show that every point of the form $q(y)$ has a simply connected neighborhood. Let V be a simply connected neighborhood of y in Y, and consider the open neighborhood V^* of $q(y)$ whose inverse image under q is the saturated open set $f^{-1}[V] \times [0,1] \amalg V$. We know that V is homeomorphic to $q[V]$, so it is only necessary to check that $q[V]$ is a strong deformation retract of V^*. One quick way of seeing this is to check directly that V^* is merely the mapping cylinder of the map F_V from $f^{-1}[V]$ to V defined by f (so that $f_V(x) = f(x)$ for all $x \in f^{-1}[V]$); this is immediate on the set-theoretic level, and on the topological level it follows because of the following basic observation: If $q : A \to B$ is a quotient map and U is open in B, then the continuous map from $q^{-1}[U]$ to U defined by q is also a quotient map.■

FURTHER COMMENTS ON MUNKRES, CHAPTER 11

In the final two paragraphs of Section 69 in Munkres (see p. 425), the *isomorphism decision problem* for finitely presented groups is mentioned. This problem asks whether there is some uniform, totally systematic procedure for determining whether two finitely presented groups are isomorphic. One criterion for such a procedure is that it should lead to a computer program which could, after a finite amount of time, determine whether or not two finite presentations (of generators and relations) define isomorphic groups. As noted in Munkres, one can prove that no such procedure exists. This is one of several decision problems about groups that were shown to be unsolvable during the nineteen fifties. Further information on such questions appears in Chapter 12 of the previously cited book by Rotman, and the unsolvability of the isomorphism question appears as Corollary 12.34 on page 469 of that reference. The following Wikipedia reference discusses the central question (the *Word Problem*) starting from first principles:

`http://en.wikipedia.org/wiki/Word_problem_for_groups`

Munkres, Section 79

Several points in the introduction to Chapter 13 are important enough to note. First of all, there is a default hypothesis that the domains and codomains of covering spaces are Hausdorff and (arcwise) connected unless explicitly stated otherwise; one reason for this is that arbitrary covering space projections (on locally arcwise connected spaces) split into disjoint unions of covering space projections with arcwise connected domains and codomains. Also, we shall need the full force of Theorem 54.6. This result states that if $p : (E, e_0) \to (B, b_0)$ is a base point preserving covering space projection, then the associated map of fundamental groups is injective, and if H denotes its image then path lifting defines a 1–1 correspondence between $p^{-1}[\{b_0\}]$ and the quotient of $\pi_1(B, b_0)$ by the equivalence relation generated by $g \sim h\,g$ for all $h \in H$.

Change of base point lemma

Let p be as above, suppose that $e_1 \in \pi_1(E, e_1)$ and let H_1 denote the image of $\pi_1(E, e_1)$ in $\pi_1(B, b_0)$. As noted in Munkres, one basic issue is to understand the relationship between the isomorphic subgroups H and H_1. The ultimate result is that $H_1 = g\,H\,g^{-1}$ for some g in the fundamental group of B. One key step in this argument can be restated in the following general form:

PROPOSITION. *Let (X, x_0) and (Y, y_0) be pointed spaces, let $f : X \to Y$ be base point preserving, and suppose that we are given $x_1 \in X$ and $y_1 \in Y$ such that $f(x_1) = y_1$. For $i = 0, 1$ let $f_*^{(i)}$ denote the associated map from $\pi_1(X, x_i)$ to $\pi_1(Y, y_i)$. If γ is a continuous curve in X joining x_0 to x_1, then the following diagram is commutative; in other words, we have $f_*^{(1)} \circ \gamma^* = (f \circ \gamma)^* \circ f_*^{(0)}$.*

$$\begin{array}{ccc} \pi_1(X, x_0) & \xrightarrow{\gamma^*} & \pi_1(X, x_1) \\ \big\downarrow f_*^{(0)} & & \big\downarrow f_*^{(0)} \\ \pi_1(Y, y_0) & \xrightarrow{[f \circ \gamma]^*} & \pi_1(Y, y_1) \end{array}$$

In the example involving covering spaces, the map f is the covering space projection $p : E \to B$, the points x_i are points e_i such that $p(e_0) = p(e_1) = b_0$, and the map $[f \circ \gamma]^*$ is an inner automorphism of $\pi_1(B, b_0)$. Combining this with the earlier discussions, we see that the subgroups H and H_1 are conjugates of each other.■

The proof of the proposition is a straightforward exercise.■

Munkres, Section 80

It is useful to consider some examples of universal, simply connected coverings spaces. If X is the n-torus T^n, then the universal covering is given by the map $\prod^n(p)$ from $\mathbf{R}^n$ to T^n, where $p : \mathbf{R} \to S^1$ is the usual covering space projection. If X is the projective space $\mathbf{RP}^n$ for $n \geq 2$, then the universal covering space is homeomorphic to S^n.

If M^n is a topological n-manifold and $p : E \to M$ is a covering space projection, then it follows immediately that E is also a topological n-manifold, and the results of Section 82 will imply that every topological n-manifold has a simply connected universal covering space. Furthermore, a major result of 2-dimensional topology states that every simply connected topological 2-manifold is homeomorphic to either S^2 or $\mathbf{R}^2$, and the universal covering of a 2-manifold is homeomorphic to S^2 if and only if the manifold is homeomorphic to S^2 or $\mathbf{RP}^2$. In particular, it follows that the universal covering of every connected open subset of $\mathbf{R}^2$ is homeomorphic to $\mathbf{R}^2$, and $\mathbf{R}^2$ is also (homeomorphic to) the universal covering space of the genus two surface that was previously considered.

Munkres, Section 81

Since we shall only need a weak version of the results in this section and the proof in this case simplifies, we shall state and prove the main result that we shall need. This is a converse to the earlier construction of spaces S^n/G whose fundamental groups were the groups G for various choices of n and G.

THEOREM. *Let $p : E \to B$ be a covering space projection such that E is simply connected, and let $\mathbf{Aut}(p)$ denote the set of covering transformations of p (homeomorphisms $T : E \to E$ such that $p = p \circ T$). Then $\mathbf{Aut}(p)$ is isomorphic to $\pi_1(B, b_0)$.*

It is elementary to verify that the set of all covering transformations of p is a group with respect to composition of functions (the identity qualifies, and the set is closed under taking composites and inverses).

Proof. Define a map φ from $\mathbf{Aut}(p)$ to $\pi_1(B, b_0)$ as follows: If T is a covering transformation, let γ be a continuous curve from e_0 to $T(e_0)$ and take $\varphi(T)$ to be the class of $[p \circ \gamma]$ in $\pi_1(B, b_0)$. The simple connectivity of E implies that $\varphi(T)$ does not depend upon the choice of γ.

Given an arbitrary element g of the fundamental group, we know it has the form $[p \circ \gamma]$ for some curve γ. Let $e \in E$ be the point over b_0 such that $\gamma(1) = e$. Then by the general lifting criterion we know that there is some continuous lifting $T : E \to E$ of the map p such that $T(e_0) = e$, and similarly there is some continuous lifting S of the map p such that $S(e) = e_0$. It follows that $S \circ T$ and $T \circ S$ are liftings of p such that $S \circ T(e_0) = e_0$ and $T \circ S(e) = e$. By the uniqueness property of liftings it follows that $S \circ T$ and $T \circ S$ are identity mappings, and hence T is a homeomorphism; it follows that T lies in $\mathbf{Aut}(p)$ and by construction we have $\varphi(T) = g$. Therefore φ is surjective.

Suppose now that $\varphi(T) = \varphi(S)$, and let α and β be the curves in the construction of φ. It then follows that $p \circ \alpha$ and $p \circ \beta$ represent the same element of the fundamental group, and thus we must have $\alpha(1) = \beta(1)$. Since the left hand side is equal to $T(e_0)$ and the right hand side is equal to $S(e_0)$, it follows that these two elements of E are the same. We can now apply the uniqueness property of liftings to show that $S = T$.

Finally, we need to show that the two groups are isomorphic. Let S and T be arbitrary covering transformations, and as in the preceding paragraph let α and β be the curves in the construction of $\varphi(S)$ and $\varphi(T)$ respectively. It then follows that $\alpha(1) = S(e_0)$ and $\beta(1) = T(e_0)$, and consequently the curve $\alpha + S \circ \beta$ joins e_0 to $S \circ T(e_0)$. Thus we have

$$\varphi(S \circ T) \quad = \quad [p \circ (\alpha + S \circ \beta)] \quad = \quad [p \circ \alpha + p \circ \beta] \quad = \quad [p \circ \alpha]\,[p \circ \beta] \quad = \quad \varphi(S) \cdot \varphi(T)$$

so that the mapping φ is a group isomorphism.■

In the next section we shall also need the following generalization of the preceding result:

THEOREM. *Let $p : E \to B$ be a covering space projection such that the image H of the map $p_* : \pi_1(E, e_0) \to \pi_1(B, b_0)$ is a normal subgroup, and let $\mathbf{Aut}(p)$ denote the set of covering transformations of p. Then $\mathbf{Aut}(p)$ acts transitively on $F = p^{-1}[\{b_0\}]$, and $\mathbf{Aut}(p)$ is isomorphic to $\pi_1(B, b_0)/H$.*■

Munkres, Section 82

The construction of a universal simply connected covering space is somewhat complicated, but the existence of such objects for locally well-behaved spaces itself is fundamentally important. One advantage of the Munkres definition of semilocally simply connected spaces is apparent in the corollary near the end of the section, which states that this is the most general situation in which one can construct a simply connected covering space.

In practice, the spaces for which we want simply connected covering spaces usually satisfy a local contractibility condition which states that each point has a neighborhood base of contractible open subsets. This applies in particular to topological n-manifolds (hence to open subsets of $\mathbf{R}^n$) and to the edge-vertex graphs that we have considered.

Realizing subgroups by covering spaces

If we say that two covering space projections $p : E \to B$ and $p' : E' \to B$ are equivalent if and only if there is a homeomorphism $h : E' \to E$ such that $p' = p \circ h$, then the results of Munkres give a complete proof that the equivalence classes of covering spaces are in 1–1 correspondence with the conjugacy classes of subgroups of the fundamental group of B provided B has a simply connected covering space $\widetilde{B}$. Given a subgroup H of the fundamental group, one method for constructing this covering space is to take the quotient E_H of $\widetilde{B}$ by the equivalence relation generated by the condition $x \sim h \cdot x$ for all $x \in \widetilde{B}$ and $h \in H$. It follows immediately that image of the fundamental group $\pi_1(E_H, e)$ in $\pi_1(B, b_0)$ is equal or at least conjugate to H.

Application to the Seifert-van Kampen Theorem

Using universal covering spaces and the preceding remarks, we may give an alternate proof of the Seifert-van Kampen Theorem for arcwise connected spaces X which are expressible as unions of arcwise connected open subspaces U and V such that $U \cap V$ is arcwise connected and each of U and V has a simply connected covering space.

Let G and H denote the fundamental groups of U and V respectively, and let $\widetilde{U}$ and $\widetilde{V}$ denote their universal coverings. As before, let N be the normal subgroup of $G * H$ which is normally generated by elements of the form

$$i_{0*}(y) \cdot \left(i_{0*}(y) \right)^{-1}$$

where $y \in \pi_1(U \cap V, x_0)$ and $i_0 : U \cap V \to V$, $j_0 : U \cap V \to U$ are the inclusion mappings. We are then interested in the group

$$\Gamma \quad = \quad (G * H)/N$$

and we know that there is a canonical homomorphism from Γ to $\pi_1(X, x_0)$. Since the images of G and H generate the fundamental group, we know that this canonical homomorphism is onto. We shall prove that the map is 1–1 by constructing a covering space E of X whose fundamental group is isomorphic touch that the image of $\pi_1(E, e)$ is equal to the image of Γ.

Digression. The construction of the desired covering space involves some general concepts that are also important in other contexts.

NOTATIONAL CONVENTION. If we are given a group L which acts on a space X and a homomorphism $j : L \to M$, then we define $M \times_L X$ to be the quotient of $M \times X$ modulo the equivalence relation defined by $(g, x) \sim (g \cdot j(h), h^{-1} \cdot x)$ for all (g, x) in $M \times X$ and $h \in L$. It follows immediately that if $p : E \to B$ is a covering space projection (where E is *not necessarily connected*) and L acts as a group of covering space transformations on E, then $M \times_L E$ is a **not necessarily connected** covering space over B with projection p_M sending $[g, x]$ to $p(x)$ for all (g, x). This is often called a *balanced product construction.*

This construction has two important properties:

TRANSITIVITY PROPERY. *In the setting above, if we are also given a homomorphism $k : M \to N$, then there is a canonical homeomorphism from $N \times_M (M \times_L X)$ to $N \times_L X$. If $E \to B$ is a covering space and L acts as a group of covering space transformations on E, then the canonical homeomorphism is in fact an equivalence of covering spaces.*■

RESTRICTION PROPERY. *Suppose that $E \to B$ is a covering space projection such that E is simply connected and L acts on E by covering space transformations. Suppose that A is an arcwise connected, locally arcwise connected subspace of B which has a simply connected covering space, and let $j_* : \pi_1(A.b_0) \to \pi_1(B, b_0)$ be induced by the inclusion of A in B. If E_A is the inverse image of A in E and $p_A : E_A \to A$ is the restricted covering space (which need not be connected), then there is an equivalence of covering spaces from E_A to $\pi_1(B, b_0) \times_{\pi_1(A,b_0)} \widetilde{A}$, where as usual $\widetilde{A} \to A$ denotes the universal covering space projection.*■

Both proofs are straightforward and left as exercises.

Completion of the proof of the Seifert-van Kampen Theorem. Consider the spaces

$$U_\Gamma \;=\; \Gamma \times_G \widetilde{U} \;, \qquad V_\Gamma \;=\; \Gamma \times_H \widetilde{V} \;.$$

By the Transitivity and Restriction Properties, the restrictions of these covering spaces to $U \cup V$ are canonically equivalent to

$$\Gamma \times_{\pi_1(U \cap V)} (U \cap V)^\sim$$

(where $(\ )^\sim$ denotes the universal covering space), and if we take the quotient of $U_\Gamma \amalg V_\Gamma$ formed by identifying points in these two open subsets via the equivalence of covering spaces, we obtain a space E, a covering space projection $E \to X$, and an action of Γ on E by covering space transformations. This action is transitive on the inverse image F of the base point x_0; in other words, if $e_0 \in F$ is the base point of E and $e_1 \in E$, then there is a (necessarily unique) covering transformation T such that $T(e_0) = e_1$.

We claim that E is arcwise connected. In fact, it suffices to show that the inverse image of F lies in a single arc component of E, for if $y \in E$ then one has a continuous curve γ joining $p(y) \in X$ to the base point of X, and if we take the unique lifting of γ which starts at y we obtain a curve joining y to a point in F; if all of F lies in one arc component of E, it then follows that every point of E lies in this arc component.

By construction, if e_1 and e_2 are two points in F such that $g \cdot e_1 = e_2$ for some g in the image of the fundamental group of U, it follows that e_1 can be joined to e_2 by a continuous curve whose image lies in the inverse image of U in E; a similar conclusion holds if we replace U be V in the preceding statement. By the construction of Γ and the transitivity of Γ on F, we know that if e

and e_0 are in Γ then there is some g such that $g \cdot e_0 = e$, and we also know that g can be written as a product $g_1 \ \cdots \ g_k$ where each g_i comes from the fundamental group of U or V. It follows by induction that for all j the point $g_j \ \cdots \ g_k \cdot e_0$ lies in the same arc component as e_0. If we take $j = 1$ we see that e and e_0 must lie in the same arc component of E. This completes the proof that E is arcwise connected.

Since Γ acts as a group of covering transformations on E and it is transitive on F, the results of Section 81 imply that the image J of $\pi_1(E, e_0)$ in $\pi_1(B, b_0)$ is a normal subgroup and the quotient group is isomorphic to Γ. In fact, the projection map $\partial : \pi_1(B, b_0) \to \Gamma$ is given by taking a closed curve γ representing an element g of the fundamental group of B, forming the unique lifting $\widetilde{\gamma}$ starting at e_0, and defining $\partial(g)$ so that $\widetilde{\gamma}(1) = g \cdot e_0$. One must use the fact that J is normal in the fundamental group to prove that ∂ is a homomorphism.

Combining this with previous observations, we obtain the diagram of morphisms displayed below, in which the square is commutative (all compositions of morphisms between two objects in this part of the diagram are equal).

$$\begin{array}{ccccccc}
\pi_1(U \cap V) & \longrightarrow & \pi_1(U \cap V) & & & & \\
\downarrow & & \downarrow J(U) & & & & \\
\pi_1(U \cap V) & \xrightarrow{J(V)} & \Gamma & \xrightarrow{\Phi} & \pi_x(X) & \xrightarrow{\partial} & \Gamma
\end{array}$$

The map Φ is the homomorphism given by the universal mapping property of the pushout group Γ (see the commentary to Section 70). If we can show that $\partial \circ \Phi$ is the identity, then it will follow that Φ is injective. Since we already know that Φ is surjective (see Section 70), it will follow that Φ is an isomorphism, and the proof will be complete.

By the construction of the covering space E and the map ∂, it follows immediately that $\partial \circ i_{U*} = \partial \circ \Phi \circ J(U) = J(U)$ and $\partial \circ i_{V*} = \partial \circ \Phi \circ J(V) = \partial \circ J(V)$. Since the identity 1_Γ on Γ satisfies $1_\gamma \circ J(U) = J(U)$ and $1_\gamma \circ J(V) = J(V)$, It follows that the identity and $\partial \circ \Phi$ agree on the images of $J(U)$ and $J(V)$. Since these sets generate Γ, it follows that $\partial \circ \Phi = 1_\Gamma$, and as noted before this suffices to complete the proof of the Seifert-van Kampen Theorem.■

In fact, the covering space E constructed in the proof is simply connected. This will follow if we can show that ∂ is an isomorphism, for the latter will imply that the kernel of ∂ — which is isomorphic to the fundamental group of E — must be trivial; to see the assertion regarding ∂, note that the proof implies that Φ is an isomorphism, and since $\partial \circ \Phi$ is the identity it follows that $\partial = \Phi^{-1}$.■

FURTHER COMMENTS ON MUNKRES, CHAPTER 13

There are some noteworthy similarities between the classification of based covering spaces over a topological space and the classification of subfields in a Galois extension of a field. In both cases, the objects of interest are classified by subgroups of a naturally associated group — the fundamental group $\pi_1(B, b_0)$ in the case of covering spaces and the Galois group G in the case of a Galois extension K/F. In the case of topological spaces, if $\pi_1(E_1, e_1)$ is contained in $\pi_1(E_2, e_2)$, then there is a covering space projection $q : E_2 \to E_1$ such that the respective covering space projections $p_i : E_i \to B$ satisfy $p_2 = p_1 \circ q$, and in the case of Galois extensions if H_1 is a subgroup of H_2 then the corresponding subfields E_i satisfy $E_2 \subset E_1$. In both cases the objects which correspond to normal subgroups are characterized by special properties. For Galois extensions, H is normal in the Galois group G if and only if the corresponding subfield E is a Galois extension of F, and in this case the Galois group of E/F is isomorphic to G/H. For covering spaces, H is normal in the fundamental group if and only if the corresponding covering space is *regular* or *normal*, which means that the covering space has a group of covering space transformations which are transitive on the inverse image of the base point, and this group of covering transformations is isomorphic to the quotient $\pi_1(B, b_0)/H$. Because of this similarity, some writers say that the group of covering space transformations of a regular covering space projection is the *Galois group* of the covering space data. Further information and reference are given at the following online site:

`http://planetmath.org/encyclopedia/ClassificationOfCoveringSpaces.html`

Munkres, Section 83

Unless specifically stated otherwise, all graphs will be finite, and we shall also assume that we are working with finite linear graphs in the sense of Munkres except in some exercises.

Definition. Let $(X, \mathcal{E})$ be a finite linear graph. A *subgraph* $(X_0, \mathcal{E}_0)$ is given by a subfamily $\mathcal{E}_0 \subset \mathcal{E}$ such that X_0 is the union of all the edges in $\mathcal{E}_0$. It is said to be a *full subgraph* if two vertices $\mathbf{v}$ and $\mathbf{w}$ lie in X_0 and there is an edge $E \in \mathcal{E}$ joining them, then $E \in \mathcal{E}_0$.

The following basic result is an elaboration of the first lemma in Section 84:

PROPOSITION. *If $(X, \mathcal{E})$ is a finite linear graph, then X is connected if and only if for each pair of distinct vertices $\mathbf{v}$ and $\mathbf{w}$ there is an* edge-path sequence *$E_1, \cdots, E_n$ such that $\mathbf{v}$ is one vertex of E_1, $\mathbf{w}$ is one vertex of E_n, for each k satisfying $1 < k \leq n$ the edges E_k and E_{k-1} have one vertex in common, and $\mathbf{v}$ and $\mathbf{w}$ are the "other" vertices of E_1 and E_n. Furthermore, X is a union of finitely many components, each of which is a full subgraph.*

Proof. First of all, since every point lies on an edge, it follows that every point lie in the connected component of some vertex. In particular, there are only finitely many connected components. Define a binary relation on the set of vertices such that $\mathbf{v} \sim \mathbf{w}$ if and only if the two vertices are equal or there is an edge-path sequence as in the statement of the proposition. It is elementary to check that this is an equivalence relation, and that vertices in the same equivalence class determine the same connected component in X.

Given a vertex $\mathbf{v}$, let $Y_\mathbf{v}$ denote the union of all edges containing vertices which are equivalent to $\mathbf{v}$ in the sense of the preceding paragraph. If we choose one vertex $\mathbf{v}$ from each equivalence class, then we obtain a finite, pairwise disjoint family of closed connected subsets whose union is X, and it follows that these sets are must be the connected components of X. In fact, by construction each of these connected component is a full subgraph of $(X, \mathcal{E})$.■

Frequently it is convenient to look at edge-path sequences that are *minimal* or *simple* in the sense that one cannot easily extract shorter edge-path sequences from them. Specifically, if we have an edge-path sequence $E_1, \cdots, E_n$ such that for some $i < j$ the common vertices $\mathbf{v}_i \in E_i \cap E_{i-1}$ and $\mathbf{v}_j \in E_j \cap E_{j-1}$ are equal, then clearly we can obtain a shorter path by eliminating $E_i, \cdots E_{j-1}$. Therefore we shall say that an edge-path sequence is *reduced* if the vertices $\mathbf{v}_i$ as above are distinct.

IMPORTANT: If we let $\mathbf{v}_0$ and $\mathbf{v}_n$ denote the vertices of E_1 and E_n that are not given by intersections of adjacent edges as in the preceding paragraph, then we are **NOT** making any assumptions about whether or not these two vertices are equal. If they are, then we shall say that the edge-path sequence is *closed* or that it is a *simple circuit* or *simple cycle*.

Munkres, Section 84

If we restrict attention to finite graphs, then the existence of a maximal tree can be established without using Zorn's Lemma; it suffices to take a tree with the largest number of edges (this is possible since the whole graph only has finitely many edges).

We shall need the following strengthening of Theorem 84.3:

THEOREM. *If $(T, \mathcal{E})$ is a tree and $\mathbf{v}$ is a vertex of this graph, then $\{\mathbf{v}\}$ is a strong deformation retract of X.*

Proof. This is trivial for graphs with one edge because they are homeomorphic to the unit interval. Suppose now that we know the result for trees with at most n edges, and suppose that $(T, \mathcal{E})$ has $n+1$ edges.

By Lemma 84.2 we may write $T = T_0 \cup A$ where A is an edge and T_0 is a tree with n edges such that $A \cap T_0$ is a single vertex $\mathbf{w}$. Let $\mathbf{y}$ be the other vertex of A. The proof splits into cases depending upon whether or not the vertex $\mathbf{v}$ of T is equal to $\mathbf{y}$, $\mathbf{w}$ or some other vertex in T_0.

We shall need the following two results on strong deformation retracts; in both cases the proofs are elementary:

(1) *Suppose X is a union of two closed subspaces $A \cup B$, and let $A \cap B = C$. If C is a strong deformation retract of both A and B, then C is a strong deformation retract of X.*

(2) *Suppose X is a union of two closed subspaces $A \cup B$, and let $A \cap B = C$. If C is a strong deformation retract of B, then A is a strong deformation retract of X.*

Suppose first that the vertex is $\mathbf{w}$. Then $\{\mathbf{w}\}$ is a strong deformation retract of both A and T_0, so by the first statement above it is a strong deformation retract of their union, which is T.

Now suppose that the vertex is $\mathbf{y}$. Then the second statement above implies that A is a strong deformation retract of T. Since $\{\mathbf{y}\}$ is a strong deformation retract of A, it follows that $\{\mathbf{y}\}$ is also a strong deformation retract of T.

Finally, suppose that the vertex $\mathbf{v}$ lies in T_0 but is not $\mathbf{w}$. Another application of the second statement implies that T_0 is a strong deformation retract of T, and since $\{\mathbf{v}\}$ is a strong deformation retract of T_0, it follows that $\{\mathbf{v}\}$ is also a strong deformation retract of T.∎

We shall also give a slightly different proof of Theorem 84.7 in which the role of Lemma 84.6 is replaced by the following:

PROPOSITION. *Suppose that the connected graph $(X, \mathcal{E})$ contains a maximal tree T such that X is the union of T with a single edge E^*. Then X is homotopy equivalent to S^1.*

Proof. Since T is a maximal tree, the vertices of E^* lie in T. If $\mathbf{a}$ and $\mathbf{b}$ are these vertices, then there is a reduced edge-path sequence $E_1, \cdots, E_n$ joining $\mathbf{a}$ to $\mathbf{b}$, and if we let Γ be the union of the $E - i$'s and E^*, it follows that Γ must be homeomorphic to S^1. By construction Γ determines a subgraph of X. For the sake of uniformity, set $\mathbf{v}_0 = \mathbf{a}$ and $\mathbf{v}_n = \mathbf{b}$.

We claim that Γ is a strong deformation retract of X. Let Y be the subgraph obtained by removing the edges E^* and E_i from $\mathcal{E}$, and for each i let Y_i be the component of $\mathbf{v}_i$. By our assumptions it follows that Y and the subgraphs Y_i are trees. It will suffice to prove that if $i \neq j$ then $\mathbf{v}_j \notin Y_i$, for then we have $Y_i \cap \Gamma = \{\mathbf{v}_i\}$ and we can repeatedly apply the criteria in the previous argument to show that Γ is a strong deformation retract of X.

Suppose now that $\mathbf{v}_j \notin Y_i$ for some $j \neq i$. Then there is some reduced edge-path sequence $F_1, \cdots, F_m$ joining $\mathbf{v}_i$ to $\mathbf{v}_j$ in Y_i. Since the vertices of the edges F_r contain at least one $\mathbf{v}_j$ other than $\mathbf{v}_i$, there is a first edge in the sequence F_s which contains such an edge, say $\mathbf{v}_k$. Of course, none of the edges F_r lies in Γ. However, we also know that there is a reduced edge path sequence in $\Gamma \cap T$ which joins $\mathbf{v}_j$ to $\mathbf{v}_k$, and we can merge this with the edge-path sequence $F_1, \cdots, F_s$ (whose edges lie in T but not Γ) to obtain a reduced cycle in T. Since T is a tree, this is a contradiction, and therefore we must have $Y_i \cap \Gamma = \{\mathbf{v}_i\}$. As noted before, this suffices to complete the proof.■

Alternate proof of Theorem 84.7. Let T be a maximal tree in the connected graph X, and let $F_1, \cdots, F_b$ denote the edges of X which do not lie in T. Let $W \subset X$ be the open set obtained by deleting exactly one non-vertex point from each of the edges F_i, and let $U_j = W \cup F_j$. It then follows that each U_j is an open subset of X and if $i \neq j$ then $U_i \cap U_j = W$. Furthermore T is a strong deformation retract of W and for each subset of indices $i_1, \cdots, i_k$ the set $F_{i_1} \cup F_{i_k}$ is a strong deformation retract of $U_{i_1} \cup U_{i_k}$. In particular, we know that the sets W and U_i are all arcwise connected. By the preceding result we know that F_1 and U_1 are homotopy equivalent to S^1, and we claim by induction that the fundamental groups of $F_1 \cup \cdots \cup F_t$ and $U_1 \cup \cdots \cup U_t$ are free on t generators. For if the result is true for $t \geq 1$, then we have

$$\bigcup_{i \leq t+1} U_i \;=\; \left(\bigcup_{i \leq t} U_i\right) \cup U_{t+1}\,, \qquad W \;=\; \left(\bigcup_{i \leq t} U_i\right) \cap U_{t+1}$$

so that the Seifert-van Kampen Theorem implies that the fundamental group of $U_1 \cup \cdots \cup U_{t+1}$ is the free product of the fundamental groups of $U_1 \cup \cdots \cup U_t$ and U_{t+1}. By induction the group for the first space is free on t generators while the group for the second is infinite cyclic, and this completes the proof of the inductive step.■

The Euler characteristic of a graph

If $(X, \mathcal{E})$ is a connected graph, then the preceding discussion shows that the fundamental group of X is a free group on a finite set of free generators. We would like to have a formula for the number of generators which can be read off immediately from the graph structure and does not require us to find an explicit maximal tree inside the graph.

Definition. The *Euler characteristic* of $(X, \mathcal{E})$ is the integer $\chi(X, \mathcal{E}) = v - e$, where e is the number of edges in the graph and v is the number of vertices.

If there is exactly one edge, then clearly $v = 2$, $e = 1$, and the Euler characteristic is equal to $1 = 2 - 1$. The first indication of the Euler characteristic's potential usefulness is an extension of this to arbitrary trees.

PROPOSITION. *If $(T, \mathcal{E})$ is a tree, then $\chi(T, \mathcal{E}) = 1$.*

Proof. Not surprisingly, this goes by induction on the number of edges. We already know the formula if there is one edge. As before, if we know the result for trees with n edges and T has

$n+1$ edges we may write $T = T_0 \cup A$, where T_0 is a tree, A is a vertex, and their intersection is a single point. For each subgraph Y let $e(Y)$ and $v(Y)$ denote the numbers of edges and vertices in Y. Then we have $e(T) = e(T_0) + 1$, $v(T) = v(T_0) + 1$, and hence we also have

$$\chi(T) \quad = \quad v(T) - e(T) \quad = \quad [v(T_0) + 1] - [e(T_0) + 1] \quad = \quad v(T_0) - e(T_0) \quad = \quad 1$$

which is the formula we wanted to verify.■

THEOREM. *If $(X, \mathcal{E})$ is a connected graph, then the fundamental group of X is a free group on $1 - \chi(X, \mathcal{E})$ generators.*

Proof. We adopt the notational conventions in the preceding argument. Let T be a maximal tree in X, and suppose that there are k edges in X which are not in T, so that the fundamental group is free on k generators. By construction we know that $v(T) = v(X)$ and $e(X) = e(T) + k$, and by the preceding result we know that the Euler characteristic of T is 1. Therefore we have

$$\chi(X, \mathcal{E}) \quad = \quad v(X) - e(X) \quad = \quad v(T) - e(T) - k \quad = \quad 1 - k$$

so that $k = 1 - \chi(X, \mathcal{E})$ as required.■

In the exercises we note that the theorem is also valid for the edge-path graphs as defined in the files for this course.

COROLLARY. *If two connected graphs X and X' are base point preservingly homotopy equivalent as topological spaces, then they have the same Euler characteristics.*

In particular, the corollary applies if X and X' are homeomorphic. For this reason we often suppress the edge decomposition and simply use $\chi(X)$ when writing the Euler characteristic.

Proof. If X and X' are homotopy equivalent, then their fundamental groups are isomorphic, and hence they are both free groups with the same numbers of generators. Since the Euler characteristics can be expressed as functions of these numbers of generators, it follows that the Euler characteristics of X and X' must be equal.■

The following special case will be used in the commentary for Section 85:

EXAMPLE. Let $\mathbf{M}$ denote the grid subset of $\mathbf{R}^2$ defined by all (x, y) such that at least one coordinate is an integer, and for integers $a < b$ and $c < d$ let $\mathbf{M}(a, b; c, d)$ denote the intersection of $\mathbf{M}$ with the rectangular region $[a, b] \times [c, d]$. It follows that this set is a finite graph whose edges are given by the edges of squares having the form $[p, p+1] \times [q, q+1]$, where $a \leq p < b$ and $c \leq q < d$. If one counts edges and vertices, then an application of the preceding theorem will show that the fundamental group of $\mathbf{M}(a, b; c, d)$ is a free group on $(b-a) \cdot (d-c)$ generators. Checking the details will be left as an exercise.

Munkres, Section 85

The goal of the commentary to this section is to prove some results on subgroups of free groups using the fundamental groups of graphs. We begin with the following result on subgroups of finite index:

PROPOSITION. *Let F be a free group on k generators, and let H be a subgroup of index n. Then H is free on $nk - n + 1$ generators.*

Proof. Let $(X, \mathcal{E})$ be a connected graph whose fundamental group is free on k generators; one method of constructing such a graph is to take edges A_i, B_i and C_i for $1 \leq i \leq k$, where the edges of A_i are x, p_i, and q_i, the edges of B_i are x, r_i, and s_i, and the edges of B_i are x, u_i, and v_i (topologically, X is a union of k circles such that each pair intersect at x and nowhere else). By the formula relating the number of generators for F and the Euler characteristic, we know that $k = 1 - \chi(X)$, or equivalently $\chi(X) = 1 - k$. Let Y be the connected covering space of X corresponding to the subgroup H. Then Y is a graph, and the fundamental group of Y is H, so that H is a free group.

We know that the number of free generators for H is given by $1 - \chi(Y)$, so it is only necessary to compute this Euler characteristic. Let e and v be the number of edges and vertices for $(X, \mathcal{E})$, so that $n = 1 - \chi(X)$, where $\chi(X) = v - e$. Since Y is an n-sheeted covering of X, if we take the associated edge decomposition of Y (such that each edge of Y maps homeomorphically to an edge of X) we see that the numbers of vertices and edges for Y are nv and ne respectively, so that

$$\chi(Y) \quad = \quad n \cdot \chi(X) \ .$$

Therefore the number of generators for the fundamental group of Y is given by

$$1 - \chi(Y) \quad = \quad 1 - n \cdot \chi(X) \quad = \quad 1 - n(1 - k) \quad = \quad nk - n + 1$$

which is what we wanted to prove.■

We shall conclude this commentary by proving the following basic result:

THEOREM. *Let F be a free group on two generators, and let $[F, F]$ be its commutator subgroup. Then $[F, F]$ is a free group on a countably infinite set of generators.*

A somewhat different subgroup $H \subset F$ which is free on infinitely many generators is described in one of the exercises.

The proof of the theorem uses a good model for the covering space of $S^1 \vee S^1$ associated to the commutator subgroup $[F, F]$, and this is based on the following observation:

PROPOSITION. *Given a pointed space (X, x_0), let $A \subset X$ be a subset such that $x_0 \subset A$ and the map induced by inclusion i_* from $\pi_1(A, x_0)$ to $\pi_1(X, x_0)$ is onto. Assume that both X and A are arcwise connected and semilocally simply connected. If $p : \widetilde{X} \to X$ denotes the universal covering space data, then the restricted projection $p^{-1}[A] \to A$ defines an arcwise connected covering space which corresponds to the subgroup* Kernel (i_*).

Proof. Let F denote the inverse image of $\{x_0\}$. Since the map of fundamental groups is onto, it follows that each point in F can be connected to the base point of $\widetilde{X}$ and $p^{-1}[A]$ by a continuous curve in the latter space. Since we are assuming that A is arcwise connected and locally arcwise connected (in fact, A is semilocally simply connected), it follows that $p^{-1}[A]$ is arcwise connected. The action of $\pi_1(X)$ on $\widetilde{X}$ by covering transformations sends $p^{-1}[A]$ to itself, and by construction we know that the map $\partial : \pi_1(A, x_0) \to \pi_1(X, x_0)$ is merely the mapping i_*. Therefore by the results of Section 81 we know that the fundamental group of $p^{-1}[A]$ is equal to the kernel if i_*.■

COROLLARY. *If D denotes the commutator subgroup of $\pi_1(S^1 \vee S^1, e)$, then the covering space of $S^1 \vee S^1$ corresponding to D is given by the set $\mathbf{M}$ of all points $(x, y) \in \mathbf{R}^2$ such that at least one coordinate is an integer.*

This follows because D is the kernel of the standard map from $pi_1(S^1 \vee S^1, e)$ to $\pi_1(T^2, e)$ and the universal covering space of T^2 is given by $p \times p : \mathbf{R} \times \mathbf{R} \to S^1 \times S^1$.■

By the preceding discussion, the proof of the theorem amounts to showing that the fundamental group of $\mathbf{M}$ is a free group on a countably infinite set of free generators. We shall do this by showing that the fundamental groups of the graphs $\mathbf{M}(a, b; c, d)$ from the preceding section yield good approximations to the fundamental group of $\mathbf{M}$.

The first step in this process is to see what happens to $\pi_1\big(\mathbf{M}(a, b; c, d)\big)$ if we increase the length or width of the rectangle:

(1) *Suppose that p is an integer such that $a < p < b$. Then the fundamental group of $\mathbf{M}(a, b; c, d)$ is isomorphic to the free product of the fundamental groups $\mathbf{M}(a, p; c, d)$ and $\mathbf{M}(p, b; c, d)$ such that the subgroup inclusions correspond to the maps of fundamental groups induced by $\mathbf{M}(a, p; c, d) \subset \mathbf{M}(a, b; c, d)$ and $\mathbf{M}(p, b; c, d) \subset \mathbf{M}(a, b; c, d)$.*

(2) *Suppose that q is an integer such that $c < q < d$. Then the fundamental group of $\mathbf{M}(a, b; c, d)$ is isomorphic to the free product of the fundamental groups $\mathbf{M}(a, b; c, q)$ and $\mathbf{M}(a, b; q, d)$ such that the subgroup inclusions correspond to the maps of fundamental groups induced by $\mathbf{M}(a, b; c, q) \subset \mathbf{M}(a, b; c, d)$ and $\mathbf{M}(a, b; q, d) \subset \mathbf{M}(a, b; c, d)$.*

Since the second statement can be derived from the first by switching the roles of the first and second coordinates, it is only necessary to prove the first statement. The first step is to thicken the sets $\mathbf{M}(a, p; c, d)$ and $\mathbf{M}(p, b; c, d)$ to open subsets U and V such that the inclusions are strong deformation retracts. Specifically, take U to be the set of all points in $\mathbf{M}(a, b; c, d)$ whose first coordinate is less than $p + \frac{1}{2}$ and take V to be the set of all points in $\mathbf{M}(a, b; c, d)$ whose first coordinate is greater than $p - \frac{1}{2}$. It then follows that the simply connected set $\{p\} \times [c, d]$ is a strong deformation retract of $U \cap V$, and therefore the fundamental group of $\mathbf{M}(a, b; c, d)$ is isomorphic to the free product of the fundamental groups of U and V such that the subgroup inclusions correspond to the maps of fundamental groups induced by the inclusion mappings. Since the latter are isomorphic to the fundamental groups of $\mathbf{M}(a, p; c, d)$ and $\mathbf{M}(p, b; c, d)$, the conclusion in (1) follows immediately. As noted above, similar reasoning establishes (2).

Using (1) and (2), we shall establish the next step:

(3) *For all positive integers n, the inclusion of $\mathbf{M}(-n, n; -n, n)$ in $\mathbf{M}(-n-1, n+1; -n-1, n+1)$ is a monomorphism of free groups such that the image of a set of free generators for the fundamental group of $\mathbf{M}(-n, n; -n, n)$ extends to a set of free generators for the fundamental groupo of $\mathbf{M}(-n-1, n+1; -n-1, n+1)$.*

This can be shown by applying the previous observations to the following sequence of inclusions:

$$\mathbf{M}(-n, n; -n, n) \quad \subset \quad \mathbf{M}(-n, n+1; -n, n) \quad \subset \quad \mathbf{M}(-n-1, n+1; -n, n) \quad \subset$$

$$\mathbf{M}(-n-1, n+1; -n, n+1) \quad \subset \quad \mathbf{M}(-n-1, n+1; -n-1, n+1)$$

By the results of the preceding section the fundamental group of $\mathbf{M}(-n, n; -n, n)$ is free on $4n^2$ generators. Let F_∞ be the free group on a countably infinite set of generators x_i, and identify the fundamental group of $\mathbf{M}(-n, n; -n, n)$ with the first $4n^2$ generators in the sequence; by (3), this can be done coherently for all n.

Using (3) and the preceding paragraph, we may construct a group homomorphism from F_∞ to $\pi_1(\mathbf{M})$ such that for each n the restriction to the subgroup generated by the first $4n^2$ elements of the sequence $\{x_i\}$ corresponds to the map of fundamental groups associated to the inclusion of $\mathbf{M}(-n, n; -n, n)$ in $\mathbf{M}$. We claim this map is an isomorphism.

The key fact in proving the claim is that every compact subset of $\mathbf{M}$ is contained in some $\mathbf{M}(-n, n; -n, n)$. This immediately implies that every element of $\pi_1(\mathbf{M})$ comes from the fundamental group of some subspace $\mathbf{M}(-n, n; -n, n)$, and the latter in turn implies that the map from F_∞ to $\pi_1(\mathbf{M})$ is onto. Similarly, if we have a closed curve in some $\mathbf{M}(-n, n; -n, n)$ which is nullhomotopic in $\mathbf{M}$, then there is some $m \geq n$ such that the closed curve is nullhomotopic in $\mathbf{M}(-m, m; -m, m)$. Since the inclusion mappings of fundamental groups are monomorphisms, it follows that the closed curve must already by nullhomotopic in $\mathbf{M}(-n, n; -n, n)$. This implies that the map from F_∞ to $\pi_1(\mathbf{M})$ is 1–1, and if we combine this with the previous observations we see that this homomorphism must be an isomorphism.■

Comparison with Munkres

As indicated by the title of Section 85, the main result in this section of the text (Theorem 85.1) describes the isomorphism types of all subgroups of a free group (namely, they are free). Its proof requires the use of *infinite graphs*, which are defined and studied in the Munkres but not in these commentaries. The two main steps in the argument are similar in nature to results discussed here:

(1) Proof that the fundamental group of an infinite graph is a free group (see Theorem 84.7).

(2) Construction of a graph structure on a covering space of an arbitrary graph (see Theorem 83.4).

One can view Theorem 85.1 as a special case of the previously mentioned Kurosh Subgroup Theorem.

A fundamentally important construction relating group theory and algebraic topology is described in Section 1.B of Hatcher. The Wikipedia articles

`http://en.wikipedia.prg/wiki/Geometric_group_theory`

`http://en.wikipedia.prg/wiki/Group_cohomology`

contain more detailed additional information (and further links) concerning the ways in which topology and group theory — particularly infinite group theory — have interacted with each other in mathematical research during the past century.