Chapter 12

The Brouwer Fixed Point Theorem

In this chapter we will prove the Brouwer Fixed Point Theorem in the plane. We will use three lemmas, which are proved in the first three sections. The first two lemmas are standard ingredients in many proofs of the Brouwer theorem, while the third is a little nonstandard.

12.1 Retractions

Definition 12.1 *Let* $(X, \mathbf{K})$ *be a topological space, and suppose* $A \subset X$. *A* **retraction** *from* X *to* A *is a continuous function* $r : X \to A$ *such that* $r(a) = a$ *for all* $a \in A$.

Example 12.2 If $(X, \mathbf{K})$ is any topological space, and if $x \in X$, then the constant function which takes every point of X to x is a retraction from X to $\{x\}$.

Example 12.3 The function $r : \mathbb{R}^2 \to \{x \in \mathbb{R} \mid x_2 = 0\}$ defined by $r(x_1, x_2) = (x_1, 0)$ is a retraction. (This function projects the plane onto one of its "coordinate axes.")

The boundary of the unit disk is the unit circle

$$S^1 \stackrel{\text{def}}{=} \{x \in \mathbb{R}^2 \mid d(x, 0) = 1\}.$$

Example 12.4 The function $r : B^2 - \{0\} \to S^1$ defined by $r(x) = x/d(x, 0)$ is a retraction.

Retractions are important for the Brouwer theorem because of the following counterfactual:

Figure 12.1: Proof of the Retraction Lemma.

Lemma 12.5 (Retraction Lemma) *If there exists a continuous function* $f : B^2 \to B^2$ *with no fixed points, then there exists a retraction from* B^2 *to* S^1.

Proof. Let f be a continuous function from B^2 to B^2 with no fixed points. For each $x \in B^2$, draw a ray from $f(x)$ through x (there is a unique such ray because $f(x) \neq x$), and let $r(x)$ be the intersection of this ray with S^1. (See Fig. 12.1.)

I claim that r is a retraction. By definition, $r(x) = x$ for $x \in S^1$. Also, it is very plausible that r is continuous; if x moves a little bit, then $f(x)$ moves a little bit, so $r(x)$ moves a little bit. We will omit the formal proof of this fact. (One way to prove it is to find an explicit formula for $r(x)$ in vector notation and apply the Continuous Operations Theorem and similar facts.) □

The Retraction Lemma tells us that to prove the Brouwer theorem in the plane, it is enough to show that there is no retraction from B^2 to its boundary. But it is not hard to see intuitively that it is impossible to retract a disk to its boundary without "poking a hole" in the disk. The rest of this chapter will be devoted to a rigorous proof of this fact.

Incidentally, the Retraction Lemma generalizes easily to show that if there is a continuous function from B^n to itself, then there is a retraction from B^n to

$$\partial B^n = \{x \in \mathbb{R}^n \mid d(x, 0) = 1\},$$

an $(n-1)$-dimensional sphere. In fact, we did the case $n = 1$ at the end of Chapter 4.

It turns out that the unit disk is homeomorphic to a filled-in square, and in some situations the square is easier to work with. To this end, we will now prove a more useful version of the Retraction Lemma.

Define

$$\begin{aligned} I^2 &\stackrel{\text{def}}{=} \{x \in \mathbb{R}^2 \mid |x_1| \leq 1 \text{ and } |x_2| \leq 1\} \\ &= [-1, 1] \times [-1, 1]. \end{aligned}$$

This is a filled-in square centered at the origin. Also,

$$\begin{aligned}\partial I^2 &= \{x \in I^2 \mid |x_1| = 1 \text{ or } |x_2| = 1\} \\ &= \Big([-1,1] \times \{-1,1\}\Big) \cup \Big(\{-1,1\} \times [-1,1]\Big).\end{aligned}$$

It doesn't matter that this is the actual boundary; we just want to use the notation ∂I^2 to refer to this set.

Lemma 12.6 *There is a homeomorphism $h : B^2 \to I^2$ such that*

$$x \in S^1 \iff h(x) \in \partial I^2.$$

Proof. Define $h(0) = 0$. If $x \in B^2$ and $x \neq 0$, draw a ray from 0 through x, and let y be the intersection of this ray with ∂I^2; define $h(x) = d(y, 0)x$. We leave it to the reader to check that h is a bijection and that

$$x \in S^1 \iff h(x) \in \partial I^2.$$

We will omit the tedious proof that h and h^{-1} are continuous. □

Lemma 12.7 (Retraction Lemma 2) *To prove the Brouwer theorem in the plane, it is enough to show that there is no retraction from I^2 to ∂I^2.*

Proof. Suppose there exists no retraction from I^2 to ∂I^2. Suppose also that there exists a map $f : B^2 \to B^2$ with no fixed point. By the Retraction Lemma, there exists a retraction $r : B^2 \to S^1$. Let h be the homeomorphism of Lemma 12.6. Then hrh^{-1} is a retraction from I^2 to ∂I^2 (why?), which is a contradiction. □

Exercises

1. Check that the examples given of retractions are in fact retractions.

2. Fill in the details in the proofs of Lemma 12.6 and the Retraction Lemma 2.

3. Let $(X, \mathbf{K})$ be a connected topological space, and suppose there is a retraction from X to A. Show that A is connected. (In a certain sense that we will not define here, the unit circle is "less connected" than the unit disk, and this is a fundamental reason why there is no retraction from B^2 to S^1.)

12.2 Uniform continuity

The next ingredient we need is a special form of continuity which only makes sense in metric spaces.

Suppose X and Y are subsets of Euclidean spaces. Recall that a function $f : X \to Y$ is continuous if

$$(\forall x \in X)\ (\forall \varepsilon > 0)\ (\exists \delta > 0)\ (\forall y \in X)\ \ d(x,y) < \delta \Longrightarrow d(f(x), f(y)) < \varepsilon.$$

The definition of uniform continuity looks almost the same; only the order of the quantifiers is changed.

Definition 12.8 *Let X and Y be subsets of Euclidean spaces. A function $f : X \to Y$ is* **uniformly continuous** *if*

$$(\forall \varepsilon > 0)\ (\exists \delta > 0)\ (\forall x \in X)\ (\forall y \in X)\ \ d(x,y) < \delta \Longrightarrow d(f(x), f(y)) < \varepsilon.$$

To prove continuity, you are given an $x \in X$ and $\varepsilon > 0$, and you need to find a $\delta > 0$. To prove uniform continuity, you are given an $\varepsilon > 0$, and you need to find a *single* $\delta > 0$ that works for *every* $x \in X$.

The proofs of the next two examples are left as exercises.

Example 12.9 Any uniformly continuous function is continuous.

Example 12.10 Any isometry is uniformly continuous. (Recall that an isometry is a function $f : \mathbb{R}^n \to \mathbb{R}^n$ such that $d(f(x), f(y)) = d(x,y)$ for all $x, y \in \mathbb{R}^n$.)

Example 12.11 The function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$ is continuous but not uniformly continuous.

Proof. By the Continuous Operations Theorem, f is continuous. To show that f is not uniformly continuous, we must check that

$$(\exists \varepsilon > 0)\ (\forall \delta > 0)\ (\exists x \in X)\ (\exists y \in X)\ \ d(x,y) < \delta \ \text{ and }\ d(f(x), f(y)) \geq \varepsilon.$$

Set $\varepsilon = 1$. (Actually any $\varepsilon > 0$ will work.) To prove that $\varepsilon = 1$ works, let $\delta > 0$ be given. We must find $x, y \in \mathbb{R}$ such that $|x - y| < \delta$ and $|x^2 - y^2| \geq 1$. The former condition will be satisified if we require that $y = x + \delta/2$. To satisfy the latter condition, we need only choose x sufficiently large that $|x + y| \geq 2/\delta$ (e.g. by setting $x = 1/\delta$); then

$$|x^2 - y^2| = |x + y||x - y| \geq (2/\delta)(\delta/2) = 1. \qquad \square$$

For some subsets $A \subset \mathbb{R}^n$, it turns out that every continuous function from A to a metric space is uniformly continuous as well. These turn out to be the sets that are closed and "bounded" (see Exercise 1). These are examples of what are called **compact sets**. For the Brouwer theorem we just need the following.

Theorem 12.12 (Uniform Continuity Lemma) *If $f : I^2 \to \mathbb{R}^n$ is continuous, then f is uniformly continuous..*

Figure 12.2: An infinite sequence of nested squares.

Proof. Suppose f is continuous, and let $\varepsilon > 0$ be given. To simplify the writing, if A is any subset of I^2 and $\delta > 0$, we will use the phrase "δ works for A" to mean that

$$(\forall x \in A)\ (\forall y \in I^2)\ d(x, y) < \delta \Longrightarrow d(f(x), f(y)) < \varepsilon.$$

The proof will be complete if we can find $\delta > 0$ that works for I^2.

Suppose to the contrary that no $\delta > 0$ works for I^2. We can divide I^2 into four quadrants:

$$I^2 = \Big([-1,0] \times [-1,0]\Big) \cup \Big([-1,0] \times [0,1]\Big) \cup \Big([0,1] \times [-1,0]\Big) \cup \Big([0,1] \times [0,1]\Big).$$

I claim that there is at least one of these quadrants for which no $\delta > 0$ works. Otherwise, if δ_1 works for the first quadrant, and δ_2 works for the second quadrant, and so on, then $\min\{\delta_1, \ldots, \delta_4\}$ will work for all of I^2. So let $[a_1, b_1] \times [c_1, d_1]$ be a quadrant for which no $\delta > 0$ works. We can divide $[a_1, b_1] \times [c_1, d_1]$ into four quadrants, and by the above argument, we can choose one of these quadrants, which we'll call $[a_2, b_2] \times [c_2, d_2]$, for which no $\delta > 0$ works. Repeat this process to get an infinite sequence of nested squares $[a_1, b_1] \times [c_1, d_1], [a_2, b_2] \times [c_2, d_2], [a_3, b_3] \times [c_3, d_3], \ldots$, such that for each of these squares, no $\delta > 0$ works.

By the Principle of Nested Closed Intervals, we can choose $p_1 \in \bigcap_{n \in \mathbb{Z}^+} [a_n, b_n]$ and $p_2 \in \bigcap_{n \in \mathbb{Z}^+} [c_n, d_n]$. Then the point $p = (p_1, p_2)$ is in each of the infinitely many squares. Clearly $p \in I^2$. Since f is continuous, there exists $\delta > 0$ such that

$$(\forall q \in I^2)\ d(p, q) < \delta \Longrightarrow d(f(p), f(q)) > \varepsilon/2.$$

Observe also that we can choose n sufficiently large that

$$[a_n, b_n] \times [c_n, d_n] \subset B(p, \delta/2).$$

(Proof: by induction, we see that the maximum distance between two points in $[a_n, b_n] \times [c_n, d_n]$ is $(2\sqrt{2})2^{-n}$. For n sufficiently large, we will have $(2\sqrt{2})2^{-n} < \delta/2$, and since $p \in [a_n, b_n] \times [c_n, d_n]$, every other point in $[a_n, b_n] \times [c_n, d_n]$ must be within distance $\delta/2$ of p.)

I claim that $\delta/2$ works for $[a_n, b_n] \times [c_n, d_n]$, which will be a contradiction. To prove this, suppose $x \in [a_n, b_n] \times [c_n, d_n]$, $y \in I^2$, and $d(x, y) < \delta/2$. Since $x \in [a_n, b_n] \times [c_n, d_n]$,

$$d(p, x) < \delta/2.$$

By the triangle inequality,

$$d(p, y) \leq d(p, x) + d(x, y) < \delta/2 + \delta/2 = \delta.$$

By our choice of δ, $d(f(p), f(x)) < \varepsilon/2$ and $d(f(p), f(y)) < \varepsilon/2$. So by the triangle inequality,

$$d(f(x), f(y)) \leq d(f(x), f(p)) + d(f(p), f(y)) < \varepsilon/2 + \varepsilon/2 = \varepsilon. \qquad \square$$

Exercises

1. A subset $A \subset \mathbb{R}^2$ is **bounded** if there exists a real number R such that $d(x, 0) \leq R$ for all $x \in A$. Equivalently, every bounded set is contained in some square. Modify the proof of Theorem 12.12 to show that the Uniform Continuity Lemma holds not just for I^2, but for any subset of the plane which is closed and bounded. Where do you need to use the fact that your set is closed?

12.3 The game of Hex

The final ingredient in our proof of the Brouwer theorem is the game of Hex. Figure 12.3 shows a typical game of Hex, in progress. The players, Black and White, take turns filling in the hexagons with their respective colors. White wins if there is a chain of white hexagons connecting the top and bottom edges of the board. To be precise, such a chain is a sequence of white hexagons $h_1, h_2, \ldots, h_n$ such that h_1 is on the top of the board, h_n is on the bottom of the board, and h_i shares an edge with h_{i+1} for $1 \leq i < n$. Black wins if there is a chain of black hexagons connecting the left and right sides of the board. In the position shown, it is Black's turn, but if White is clever, he or she can win, no matter what Black does.

A game of Hex can never end in a draw; no matter how stupidly the players play, someone must eventually win. This is a consequence of the following fact, which we will prove: if all the hexagons are colored black or white, then either black or white has a winning chain. If you try playing the game, you will see empirically that this is true, although it is not obvious why. (At least it was

Figure 12.3: A game of Hex.

not immediately obvious to me; see Exercise 3, for instance.) The proof below is adapted from "The Game of Hex and the Brouwer Fixed Point Theorem" by David Gale, *American Mathematical Monthly* #86, 1979.

Theorem 12.13 (Hex Lemma) *If every hexagon of a Hex board is colored black or white, then one of the two players has a winning chain.*

Proof. To start, draw four additional regions on the outside of the Hex board, colored as in Figure 12.4. We will call the boundaries between pairs of regions **edges**. We will call the points at which edges meet **vertices**. An edge is **incident** to a vertex if the vertex is an endpoint of the edge. Notice that every vertex has exactly three edges incident to it.

Let us mark all the edges that separate oppositely colored regions. (In the figures to follow, marked edges are shown in bold.) Observe that every vertex has either 0 or 2 marked edges incident to it. This is because each vertex is a corner of three regions, and either all three regions are the same color, or one region is colored differently from the other two. (These cases are shown in Figure 12.5.)

We will find the chain we are looking for by traversing some of the marked edges. Let e_0 represent the marked edge which separates the outer left and outer bottom regions; let v_1 be the vertex where this edge meets the lower left corner of the Hex board. We now define edges $e_1, e_2, \ldots$ and vertices $v_2, v_3, \ldots$ inductively as follows. Given a marked edge e_{n-1} and a vertex v_n which is an endpoint of this edge, define e_n to be the other marked edge incident to v_n, and let v_{n+1} be the endpoint of e_n opposite v_n. This is just a formal way of saying

Figure 12.4: To prove the Hex Lemma, we add four outer regions to the Hex board.

Figure 12.5: Each vertex has either 0 or 2 marked edges incident to it.

Figure 12.6: Proof that W_n is adjacent to W_{n+1} and B_n is adjacent to B_{n+1}.

that we start at the lower left edge and follow the marked edges around the board. Where do we end up?

Observe that we can never visit the same vertex twice (Exercise 4). Hence we must eventually reach one of the other three edges on the outside of the board; call this edge e_N. For $0 \leq n \leq N$, let W_n and B_n be the white and black regions that edge e_n separates. Note that B_n is adjacent to (i.e. shares an edge with) B_{n+1}, and W_n is adjacent to W_{n+1}. To see this, one need only consider the two cases shown in Figure 12.6, and their mirror images, which behave the same way. (The mirror images do not actually occur, although we will not prove this.) Thus $B_0, B_1, \ldots, B_N$ and $W_0, W_1, \ldots, W_N$ are chains of adjacent hexagons.

Note that B_0 is the outer left region and W_0 is the outer bottom region. If e_N is the lower right edge then B_N is the outer right region and $B_1, \ldots, B_{N-1}$ is a winning chain of black hexagons. If e_N is the upper left edge then $W_1, \ldots, W_{N-1}$ is a winning chain for White. If e_N is the upper right edge then both players have winning chains. (Actually, this can't happen, although we won't prove that here.) $\square$

Exercises

1. Play Hex with your friends.

2. Why is White guaranteed to win in the position shown in Figure 12.3?

3. The game of Squ is just like the game of Hex, except that it is played on a square grid instead. Is there a Squ Lemma like the Hex Lemma, or not?

4. Prove that the path of marked edges constructed in the proof of the Hex Lemma can never visit the same vertex twice. *Hint:* Suppose that some vertex is visited twice. Prove that the *first* vertex to be visited twice has at least three marked edges incident to it, which is a contradiction.

5. Suppose a game has the following properties:

   (a) There are two players who alternate moves.

   (b) Nothing is hidden from either player.

   (c) There exists a positive integer N such that every game ends in N or fewer turns.

   (d) At the end of the game, one player or the other wins, and which player wins depends only on the moves made (no cards or dice).

   Prove that either the player who goes first or the player who goes second has a winning strategy. *Hint:* use induction on N.

   Does your proof give a good indication of what the winning strategy actually is?

6. Show that in a game of Hex, the player who goes first has a winning strategy. *Hint:* Use a proof by contradiction. Assume that the first player does not have a winning strategy, so that by the previous exercise, the second player has a winning strategy instead. Then show how the first player can "plagiarize" the second player's strategy to win, contradicting our assumption. It is essential that the board has the same horizontal and vertical dimensions.

7. (challenge problem) Prove that it is impossible for a Hex board to have winning chains in both colors simultaneously.

12.4 Proof of the Brouwer theorem

By the Retraction Lemma 2, it is enough to show that there is no retraction from I^2 to ∂I^2. Suppose that r is such a retraction; we will deduce a contradiction.

By the Uniform Continuity Lemma, we can choose $\delta > 0$ such that

$$|x - y| < \delta \Longrightarrow |r(x) - r(y)| < 2.$$

Hence no two points whose distance is less than δ are mapped to opposite sides of ∂I^2. Choose a positive integer n such that $2\sqrt{2}/n < \delta$, and divide I^2 into squares of side length $2/n$; then no two points in the same square are mapped by r to opposite sides of ∂I^2.

After dividing I^2 into squares of side length $2/n$ and adding some diagonal lines, as shown in Figure 12.7, I^2 is equivalent to a Hex board. The corners of the squares correspond to hexagons, and two hexagons are adjacent if and only if there is a single line segment connecting the corresponding points.

Number each hexagon 1, 2, 3, 4, according to which side of ∂I^2 the corresponding point is mapped to by r. (See Figure 12.7 for the numberings of the sides of ∂I^2.) If the point is mapped to two adjacent sides of ∂I^2 (i.e. a corner),

Figure 12.7: I^2 can be thought of as a Hex board.

number it either way. Color a hexagon white if it is numbered 1 or 3, and black if it is numbered 2 or 4. By the Hex Lemma, some player has a winning chain.

Without loss of generality, there is a chain of white hexagons connecting the top and bottom. Since r is a retraction, any hexagon in the chain at the top is numbered 1, and any hexagon in the chain at the bottom is numbered 3. This means that somewhere in the chain, there must be a hexagon numbered 1 adjacent to a hexagon numbered 3. But then these hexagons correspond to two points in the same square that are mapped to opposite sides of ∂I^2, and this is a contradiction. □

Exercises

1. Prove that if $f : B^2 \to B^2$ is continuous, and if $f(x) = x$ for all $x \in S^1$, then f is surjective. *Hint:* If f is not surjective then I can find a retraction from B^2 to S^1.

   This is sort of like a two-dimensional version of the Intermediate Value theorem, and it can be generalized somewhat. Suppose $f : B^2 \to B^2$ is continuous, and suppose f maps points in S^1 to points in S^1. Roughly speaking, let the variable x revolve counterclockwise around S^1 once, and count the number of times that $f(x)$ goes around S^1, where counterclockwise revolutions count as +1 and clockwise revolutions count as -1. If this number, called the "degree" of f, is nonzero, then f is surjective.

2. It is intuitively obvious that if one path goes from the left edge of a square to the right edge, and if another path goes from the bottom edge to the top edge, then the two paths must cross. Using the Brouwer fixed point theorem, one can give a rigorous proof of this fact. (Compare to Exercise 12.3.7.)

Let φ and ψ be continuous functions from $[0,1]$ to $[0,1]$. These give rise to two paths in $[0,1]\times[0,1]$; the first path takes $s\in[0,1]$ to $(s,\varphi(s))$, and the second path takes $t\in[0,1]$ to $(\psi(t),t)$. Show that these two paths cross; i.e., there exist $s,t\in[0,1]$ such that $(s,\varphi(s))=(\psi(t),t)$. *Hint:* first show that $[0,1]\times[0,1]$ has the fixed point property.

This also works for more general paths of the form $(\varphi_1(s),\varphi_2(s))$ and $(\psi_1(t),\psi_2(t))$, although this is not as easy to prove.

12.5 Another proof of the Brouwer theorem*

The proof we just gave is pleasing because of the surprising use of the game of Hex. But at the same time, one can criticize the proof for being *ad hoc*. The proof appears to be correct, but how did we ever think of it, and what does it mean? What is the main idea of the proof? What are the underlying mechanisms? If we want to try to prove other statements along the same lines, how should we proceed?

These questions are hard to answer, so we will now suggest a more conceptual proof of the Brouwer theorem. The interested reader can work out the details as a (long) exercise.

The main ingredient is the notion of the **degree** of a continuous function $f:S^1\to S^1$. Roughly speaking, let the variable x revolve counterclockwise around S^1 once, and count the number of times that $f(x)$ goes around S^1, where counterclockwise revolutions count as +1 and clockwise revolutions count as -1. This number is the degree of f, denoted by $\deg(f)$. For the example, the degree of the identity is 1; the degree of a constant function is 0; and the degree of the function $f(x)=-x$ is 1.

We just want to give an intuitive sketch here. But for reference, a more precise definition of $\deg(f)$ can be given as follows. Divide S^1 into three intervals X, Y, and Z, going around the circle counterclockwise. Also choose a large integer N and divide S^1 into N equal intervals $I_1,\ldots I_N$, in counterclockwise order. We can choose N sufficiently large that for a given k with $1\le k\le n$, the image $f(I_k)$ will intersect no more than two of the intervals X,Y,Z. This basically follows from Exercise 12.2.1. Now let $x_1,\ldots,x_N$ be the N endpoints of $I_1,\ldots,I_N$. Without loss of generality, $f(x_1)\in X$. We then define $\deg(f)$ to be the number of k such that $f(x_k)\in Y$ and $f(x_{k+1})\in Z$, minus the number of k such that $f(x_k)\in Y$ and $f(x_{k+1})\in X$. Intuitively, we are counting the number of times we cross Y going counterclockwise, minus the number of times we cross Y going clockwise. With some tedious work, one can prove that this number is well defined, independent of the choices we made to define it.

Now suppose there exists a retraction $f:B^2\to S^1$. We can think of $B^2-\{0\}$ as a union of concentric circles C_r of radius r, with $0<r\le 1$. For each r, there is a function $g_r:S^1\to C_r$ which simply "projects" $S^1=C_1$ to C_r,

by moving along rays emanating from the origin. Now define one more function, $d : (0, 1] \to \mathbb{Z}$, by $d(r) = \deg(f \circ g_r)$. Roughly speaking, to find $d(r)$, we let x rotate around a circle of radius r and count how many times $f(r)$ goes around S^1.

Since f is a retraction, $d(1) = 1$. On the other hand, since f is continuous around the origin, one can show that $d(r) = 0$ if r is sufficiently small. This is plausible because if x goes around a very small circle, then $f(x)$ does not have time to go around S^1 at all. Finally, one can use the uniform continuity of f to show that d is continuous. But this gives a contradiction. For example, by the Intermediate Value Theorem, there exists r such that $d(r) = 1/2$, but this is impossible since $d(r)$ is defined to be an integer.

Right now, the degree proof might not appear any more intuitive than the Hex proof. However, the concept of degree is useful for proving many other theorems, as you can learn in a more advanced topology course.