The Banach-Tarski Paradox and the Sierpinski-Mazurkiewicz Paradox: Use Set-Theoretic Geometry to Duplicate Sets!

Sophia Tatar

Ross Mathematics Program Lecture 3

The Purpose of this Series

As the title suggests, the purpose of this lecture is, of course, to teach you about duplication, but there is also a greater goal for this lecture and the series as a whole .The purpose of this series is to show you something surprising and beautiful so that you ***stretch the limits of your mathematical intuition and imagination***. In those moments where one senses their mathematical intuition expanding, it is common for them to think to themselves the following phrase: "wow." I hope this series makes you say "wow."

Recall what we started with way back in our first lecture. We began with just two fundamental objects: points and lines, and focused on their properties within the scope of Incidence Geometry. By starting from something as integral to the human experience as points and lines, and then tickling our brains with something that seems to break rules we normally take for granted (that parallel lines don't intersect), you may have been pleasantly surprised to see how beautifully calculations with these rules work out. Moving to even more surprising results, like duality in the real projective plane, we can enjoy the dramatic consequences of our new rules, and maybe (hopefully) even think to ourselves, "wow."[1]

In the second lecture, we explored the beautiful generalization of the simple fact about slicing ham sandwiches (and, in fact, we started even further back with slicing cake and frosting). Starting from simple constructions that we can draw nicely on papers and blackboards, we formalized what we were doing using new tools like hyperplanes and hypersurfaces in hopes of making a more generalized statement. Using these tools and some pretty functions, we went from our simple statements in $\mathbb{R}^2$ and $\mathbb{R}^3$ to making claims about higher dimensions, reaching powerful facts about how we can split objects we cannot picture, in

[1] I'm glad I chose this topic to share because it has made for lot's conversations, which may (I hope) suggest that these concepts/theorems resonated for others like they did for me.

spaces we cannot envision, using cuts we cannot imagine. I hope that you, like I, were moved by the dramatic conclusion that is at the same time so abstract (in that it's hard to imagine) and exact (a highly precise statement), and maybe that even made you go "wow."

What ties these lectures together, something stronger than that they are about geometry, is that these concepts are meant to make your brain tingle in a new way. And, on this note of asking you to think about new possibilities and surprises, let's check out how duplication of sets through rigid motion manipulation is real.

Abstract

Join us for an intriguing exploration of set-theoretic geometry where we take a countable set of points in R^2, split it into two pieces, and through only rigid motions (translation and rotation), end up with two sets identical to the original. Want more than two? Just repeat the process. This is the Sierpiński-Mazurkiewicz Paradox. No, this isn't magic or reliant on controversial principles like the Axiom of Choice. The Banach-Tarski paradox is a similar paradox in some ways that we will also explore and compare to the Sierpiński-Mazurkiewicz Paradox. The Banach-Tarski paradox a relatively famous result that uses the Axiom of Choice to showcase a similar idea about duplicating sets by manipulating them using rigid motions. Those who do not accept the Axiom of Choice do not accept the Banach-Tarski paradox. However, being a consequence of only simple principles, the Sierpiński-Mazurkiewicz Paradox is an unquestionable result of set manipulation that allows us to duplicate some special sets regardless of your stance on the Axiom of Choice, showcasing the surprising and beautiful nature of mathematics. Learn about the Sierpiński-Mazurkiewicz Paradox and how it relates to other duplication paradoxes, as well as the Axiom of Choice. No prior knowledge of any of these concepts is necessary; we will define every notion along the way in detail.

Comparing the Paradoxes

Both the Banach-Tarski Paradox and the Sierpinski-Mazurkiewicz Paradox enable us to "duplicate" sets by splitting them into pieces and manipulating the pieces using only rigid motions. However, the constructions of the two paradoxes are highly different, so it is important to highlight the differences. The following chart emphasizes the differences for duplicating a set X in each paradox.

Sierpinski-Mazurkiewicz Paradox	**Banach-Tarski Paradox**
No Axiom of Choice	Uses Axiom of Choice
X is **unbounded**	X is **bounded** (sphere)
X is **countable**	X is **uncountable**
Only known to work in $\mathbb{R}^2$	Only works in $\mathbb{R}^3$
Break X into measurable parts (measure 0)	Break X into nonmeasurable parts

The Axiom of Choice

Intuitively, the Axiom of Choice says that given any collection of non-empty sets, it is possible to create a function that selects one element from each set. For example, if I have a collection of bags where each bag contains candy of a different color, the axiom of choice says it is possible to create a function that chooses one candy from each bag. So, by the Axiom of Choice, it is possible for me to end up with one candy of each color. More formally,

Axiom of Choice

For any set X of non-empty sets, there exists a choice function f defined on X such that for every set A in X, f(A) is an element of A.

The Axiom of Choice has been a historically controversial topic in mathematics. While many mathematicians accept it because it is essential for various results in set theory and other areas, it has also faced criticism and skepticism due to its non-constructive nature. The debate over its acceptance has led to significant discussions and developments in the field, making it a notable exception to the general consensus found in most other mathematical areas.

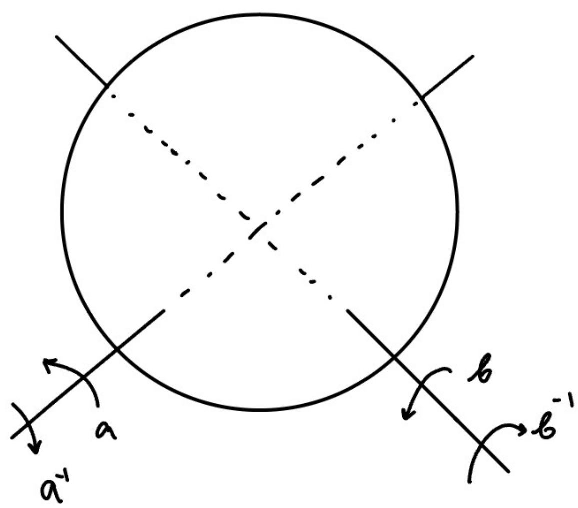
There are many paradoxical results that result from accepting the Axiom of Choice, and the Banach-Tarski paradox is one of them.

The Banach-Tarski Paradox

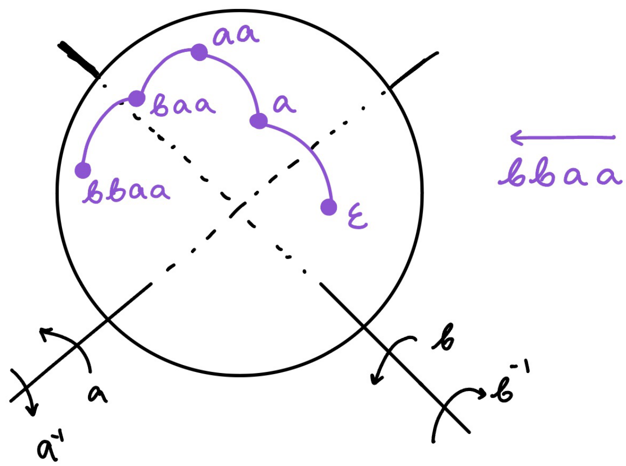
The Banach-Tarski Paradox is a statement in set-theoretic geometry that says that a solid ball in 3-dimensional space can be divided into a finite number of disjoint subsets, which can then be reassembled using only rotations and translations to form two identical copies of the original ball. This paradox relies on the Axiom of Choice.

The Construction

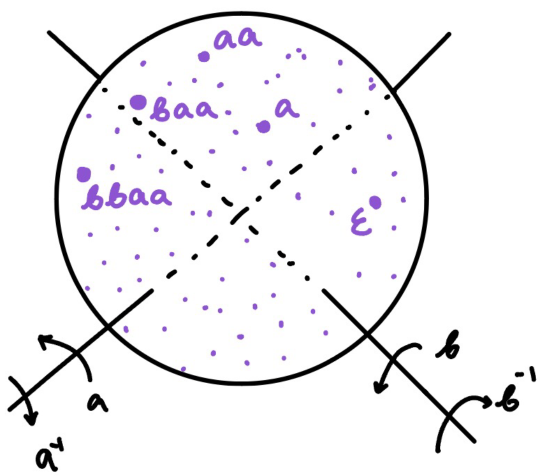
We start with a sphere and define 4 rotations on it: a, a^-1, b, b^{-1}. We define these rotations so that the magnitude of the angle of rotation of a and a^-1 are the same, but the directions and different We define b and b^{-1} analogously.



We can create strings out of our rotations, a, a^-1, b, and b^{-1}, to encode sequences of rotations. Let's make the rule that we will not write sequences adjacent "a" and "a^{-1} or b and "b^{-1}. If we chose the angles of our rotations carefully, we will never get back to the same spot by doing a sequence of rotations without adjacent a and a^{-1} or adjacent b and b^{-1}. It is a nice exercise to think about what types of angles these would have to be. Hint: most angles you are probably thinking of would have this property. Also, we will read our strings from right to left for convenience later in the proof. It is difficult to illustrate rotating the whole sphere, so we can chose one point, epsilon, and track how it moves as we apply strings (sequences of rotations) to the sphere. The image below illustrates is an example that shows how we would interpret the string *bbaa*.



The number of strings made of $a, a^{-1}1, b$, and b^{-1} is countably infinite. (You should convince yourself of the previous fact if it is not immediately clear.[2]). So, there is a countably infinite number of points on the sphere that we can get to from ϵ by moving the sphere according to a sequence of rotations. If we were to mark all of the points we can get to like this, we would have countably infinitely many points marked, as shown below. From now on, I am going to refer to the collection of points we can get to from ϵ by applying a string of rotations as the "ϵ-orbit." The image below illustrates the countably infinitely many points in the ϵ-orbit using purple dots.

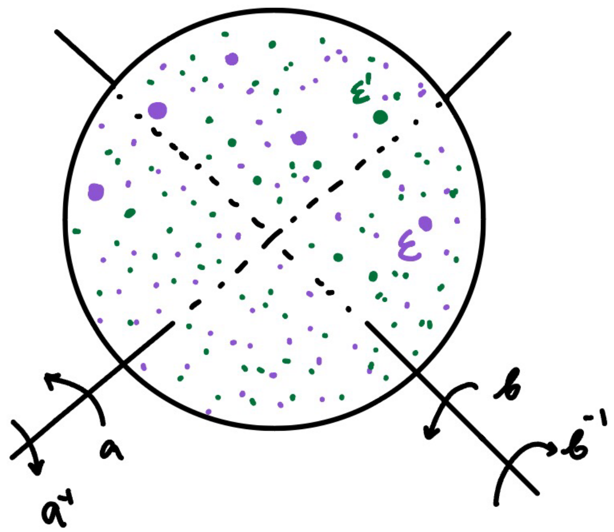


The number of points on a sphere is uncountably infinite (if this is not clear you should stop and convince yourself), so even if we were to apply every single possible sequence of rotations on epsilon and mark a countably infinite number

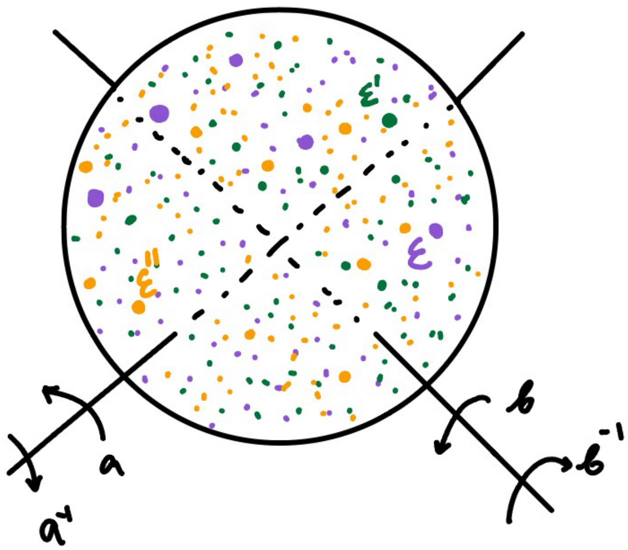
[2]Hint: Consider how many variations there are for strings of different lengths

of points, we would not reach every single point on the sphere.

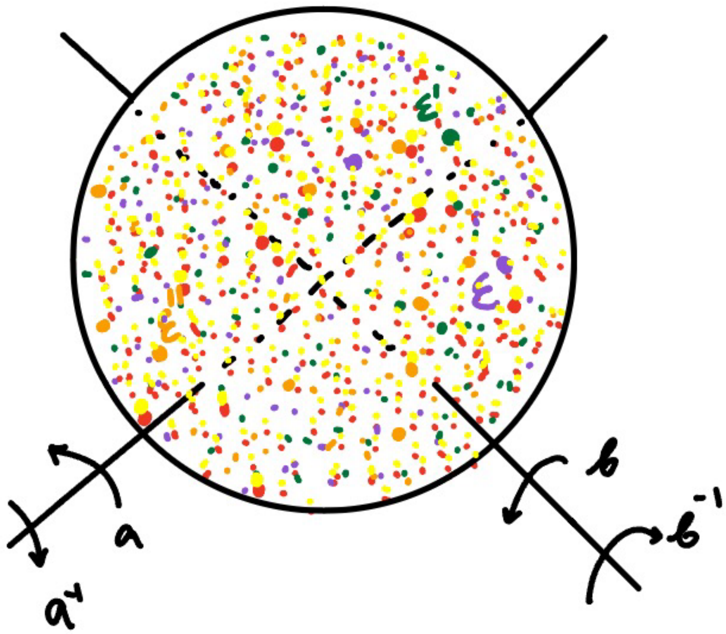
Since the ϵ-orbit does not cover all of the the points on our sphere, we can define a new point, call it ϵ', that is not in the ϵ-orbit. Now we mark all of the points that we can get to by applying a sequence of rotations on ϵ' and call the collection of all these points the "ϵ'-orbit. We can similarly mark all of the points in the ϵ'-orbit using a different color as illustrated below.



We repeat this process again and again, creating an analogous "ϵ''-orbit, "ϵ'''-orbit, "ϵ''''-orbit, and so on...The image below uses different colors to denote different orbits, illustrating the process of having some starting points and reaching a lot of other points by rotating the sphere using different sequences of rotations.



We next create more and more orbits until there is no point not already marked. We will have uncountably infinitely many orbits since there are uncountably many points on the sphere and countably infinitely many points in each orbit. Do not let my notation for the first couple of orbits confuse you into thinking there are countably infinitely many orbits. After doing this process of creating all of the orbits, every point on the sphere should be reached and colored based on its orbit.



Let's call every orbit an equivalence class $(C_\alpha)_{\alpha \in I}$ where I is a set of indices. We can then chose 1 point out of each equivalence class. In other words, we chose one point of every color. *Note: choosing one element out of each equivalence class is only possible if we accept the Axiom of Choice.* These chosen elements serve as representatives for their equivalence class. Next, we take all of these representatives and put them into a set, call it A. The next step is to create a new set, A', which will be the collection of all the points you can get to by starting at the points in A and applying a series of rotations. A' contains all the points on the sphere.

Let $S(a)$ be the set of strings that start with "a." Similarly, define $S(a^{-1})$ to be the set of strings that start with "a^{-1}," $S(b)$ to be the set of strings that start with b, $S(b^{-1})$ to be the set of strings that start with b^{-1}, and ϵ to be the empty string.

The set of all strings composed of a, a^{-1}, b, and b^{-1}, called the "free group on a and b," can be broken into 5 pieces by classifying all the strings into sets based on their first letter. So, the set of all strings, F(a,b), is

$$F(a, b) = S(a) \cup S(a^{-1}) \cup S(b) \cup S(b^{-1}) \cup \epsilon.$$

Let's take a closer look at the set $S(a)$. $S(a)$ is the set of strings starting with an "a" followed by a word starting with an "a" union the set of strings starting with an "a" followed by a word starting with "b" union the set of strings starting with an "a" followed by a words starting with "b^{-1} union the one-letter string "a." Writing this using our notation we get

$$S(a) = aS(a) \cup aS(b) \cup aS(b^{-1}) \cup a$$

There are a lot of a's, so let's append a^{-1} to all of the strings to create a new set of strings, $a^{-1}S(a)$, which can be expanded as follows.

$$a^{-1}S(a) = S(a) \cup S(b) \cup S(b^{-1}) \cup \epsilon$$

Notice, the expansion of $a^{-1}S(a)$ looks quite similar to the set of all strings made of a and b (we called this $F(a, b)$). In fact, the only set that $a^{-1}S(a)$ is missing is $S(a^{-1})$. If we union $S(a)$ with $S(a^{-1})$, we get the full set of strings.

$$F(a, b) = S(a^{-1}) \cup a^{-1}S(a) = S(a^{-1}) \cup S(a) \cup S(b) \cup S(b^{-1}) \cup \epsilon$$

So, we can get the full set of strings using only two and a little manipulation. We can create an identical argument to get

$$F(a, b) = S(b^{-1}) \cup b^{-1}S(b) = S(b^{-1}) \cup S(a) \cup S(b) \cup S(b^{-1}) \cup \epsilon$$

We just made two copies of F(a,b) by taking disjoint subsets of it and manipulating them slightly.

Previously, we thought of $S(a), S(a^{-1}), S(b), S(b^{-1})$, and ϵ as sets of strings, but we can also think of these as sets of points from our collections of all points, A', based on the last rotation we took to get there. For example, $S(a)$ would be the set of points in A' where the last rotation we take to get there is an a. In other words, $S(a)$ is the set of points on our sphere that we reach by applying a sequence of rotations that has "a" at the start of the string (recall, we read our strings backwards when doing the rotations). $S(a^{-1})$, $S(b)$, and $S(b^{-1})$ are defined analogously. $\{\epsilon\}$ is now interpreted as the set of points in A' that were in our set of representatives A, since these are the points on the sphere that we get by not doing any rotations.

Using our new definitions for $S(a), S(a^{-1}), S(b), S(b^{-1})$, and ϵ, we break all the points on the sphere into 5 disjoint sets.

$$A' = S(a) \cup S(a^{-1}) \cup S(b) \cup S(b^{-1}) \cup \epsilon$$

The argument for duplicating points on a sphere is identical to the one we used to duplicate all the set of all strings. Applying that argument here, we conclude

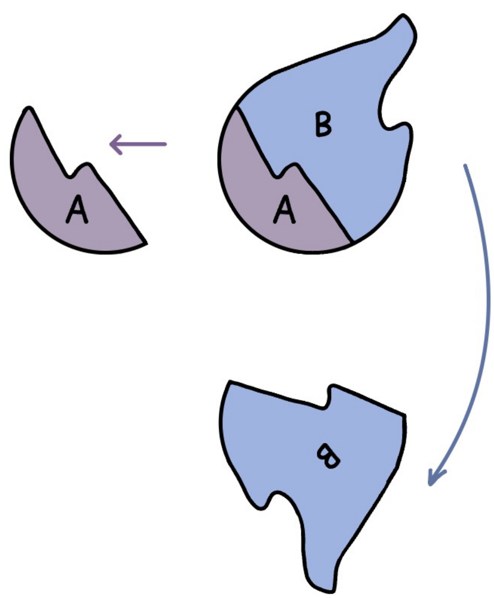
$$A' = S(a^{-1}) \cup a^{-1}S(a)$$

$$A' = S(b^{-1}) \cup b^{-1}S(b).$$

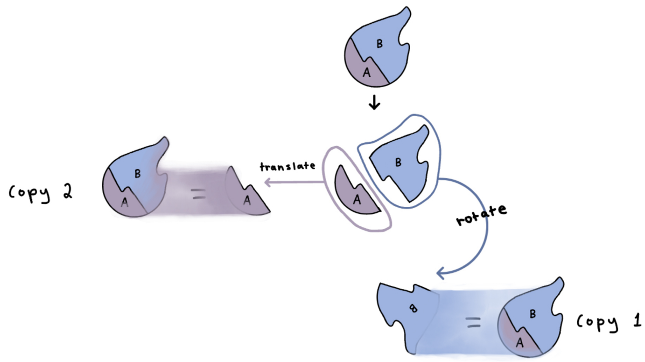
By appending a^{-1} to our set of points S(a), we rotate that piece by a^{-1}. Similarly, appending b^{-1} to $S(b)$ means that we take our set of points S(b) and rotate all of them by b^{-1}. So, by taking two pieces of our sphere at a time and rotating one of the pieces, we are able to reconstruct the whole thing. We have enough pieces to reconstruct the whole sphere twice. Now we know how to duplicate a sphere!

The Sierpinski-Mazurkiewicz Paradox

I have a set X that I partition using some algorithm into two parts: A and B. By taking A and translating it by a vector, will we get the full original set X? Well, if we chose the set X to be the one depicted in the image below, obviously not. If were to take take B, and rotate it by one radian clockwise, would we end up with the original set X. Again, obviosly not for the set depicted...



The goal of the The Sierpinski-Mazurkiewicz Paradox is to construct a set for which the answers to the previous questions are "yes." In other words, we will show a construction for some special set X where splitting it into two parts, A and B, translating A, and rotating B will somehow make each part equal to the original, becoming two full copies of the set X. If we are able to do this, then we know it is possible to duplicate a set by splitting it into two pieces using applying rigid motions.



The main difference between this paradox and the Banach-Tarski paradox is that the Banach-Tarski paradox uses the Axiom of Choice to make the construction, whereas the Sierpinski-Mazurkiewicz Paradox uses a precise algorithm to create the two sets from the original and we can even measure all of the sets involved in the construction (they all have measure 0).

Part 1 (Duplicating an abstract (helping) set $\tilde{X}$)

Take the helping set

$$\tilde{X} = \mathbb{Z}_{\geq 0}[t] = \{f = a_0 + a_1 t + a_2 t^2 + ...|a_0, a_1, ... \geq 0\}.$$

This is the set of polynomials with nonnegative coefficients. We split functions belonging to $\tilde{X}$ into two sets based on their constant term.

$$\tilde{A} = \{f \in \tilde{X}|f(0) > 0\} \subset \tilde{X}, \tilde{A} \neq \emptyset$$

$$\tilde{B} = \{f \in \tilde{X}|f(0) = 0\} \subset \tilde{X}, \tilde{B} \neq \emptyset$$

$\tilde{A}$ is the set of polynomials with nonnegative coefficients and a positive constant term, whereas $\tilde{B}$ is the set of polynomials with nonnegative coefficients and no constant term. Every polynomial $f \in \tilde{X}$, belongs to exactly one of the sets $\tilde{A}$ and $\tilde{B}$ because the constant term of a polynomial $f \in \tilde{X}$ is either zero (in which case it belongs to $\tilde{A}$) or positive (in which case it belongs to $\tilde{B}$). *Notice, we have a concrete way of splitting $\tilde{X}$ into 2 parts, no Axiom of Choice necessary.* Now, let's manipulate our set $\tilde{A}$ to form a new set

$$\tilde{A} - 1 = \{f - 1|f \in \tilde{A}\}$$

What type of elements belong to this set $\tilde{A}-1$? Recall that $a_0 > 0$ in $\tilde{A}$ by construction. From this set of polynomials $\tilde{A}$, we subtracted 1 from each polynomial. Subtracting 1 from a polynomial only affects the first term, and so after subtracting 1 from the polynomials in $\tilde{A}$ we get the set of polynomials with constant term $(a_0 - 1) \geq 0$ and the other terms are the same as before. Hence, $\tilde{A}-1$ is the set of polynomials with nonnegative coefficients. Recall that $\tilde{X}$ is also the set of polynomials with nonnegative coefficients (by definition), so $\tilde{A}-1 = \tilde{X}$. These two sets, $\tilde{X}$ and $\tilde{A}-1$, are actually the same set.

Recall that $f \in \tilde{B}$ are polynomials with nonnegative coefficients and no constant terms. Let's manipulate $\tilde{B}$ to make a new set.

$$\frac{1}{t}\tilde{B} = \{t^{-1}f | f = a_1 t + a_2 t^2 + ... \text{ where } \forall a_i \ \ a_i \geq 0\}$$

Expanding out the form of the polynomials in this set, it is evident that $\frac{1}{\tilde{B}} = \tilde{X}$.

$$\frac{1}{t}\tilde{B} = \{a_1 + a_2 t + a_3 t^2 + ... | a_i \geq 0 \ \forall i\} = \tilde{X}$$

So, in conclusion, we first split $\tilde{X}$ into two nonempty, not intersecting, sets that union to the full set $\tilde{X}$. We then called these sets $\tilde{A}$ and $\tilde{B}$. Finally, we subtracted 1 from all the elements in $\tilde{A}$ and multiplied all the elements of $\tilde{B}$ by something, which turned them both into copies of $\tilde{X}$. In this way, we duplicated $\tilde{X}$ algebraically.

Part 2 (Geometric analogy)

Now, we will take our algebraic constructions in the previous part, and transform them into something geometric and beautiful.

We use the sets of polynomials with the variable t called $\tilde{X}$, $\tilde{A}$ and $\tilde{B}$ that we defined earlier, and substitute e^i for t to get three new sets we call X, A, and B.

$$\tilde{X} \longrightarrow X \subset \mathbb{C}$$

$$\tilde{A} \longrightarrow A \subset \mathbb{C}$$

$$\tilde{B} \longrightarrow B \subset \mathbb{C}$$

Mapping from the set of polynomials to the value of the polynomials at e^i is an injective mapping, meaning that different polynomials map to different complex numbers. The following lemma proves this claim.

Lemma: If $f(t) \neq g(t)$, then $f(e^i) \neq g(e^i)$

Let $f(t) \neq g(t)$, and suppose, for the sake of contradiction, that $f(e^i) = g(e^i)$. Consider $h = f - g$. Since $f(t) \neq g(t)$, h is not the zero polynomial. We also know that $h(e^i) = 0$. However, this is a contradiction since e^i is transcendental and so it should not be the root of a nonzero polynomial h.

□

It is important that we have injective mappings because, as a result, we know that when we plug e^i into $\tilde{A}$ and $\tilde{B}$ to create A and B, the resulting sets do not overlap. So, by substituting e^i, we get a new set X, and a partition of X into two *disjoint* subsets A and B.

$$\tilde{A} \cap \tilde{B} = \emptyset \implies A \cap B = \emptyset$$

$$\tilde{X} = \tilde{A} \cup \tilde{B} \implies X = A \cup B$$

Using the set we constructed in the previous part called $\tilde{A} - 1$, we can make an analogous construction called A-1, where now we substract 1 from the polynomails evaluated at e^i rather than just the polynomials. For the same reasons as in the previous part, this should be equal to the full set of polynomials evaluated at e^i.

$$\tilde{A} - 1 = \tilde{X} \implies A - 1 = X.$$

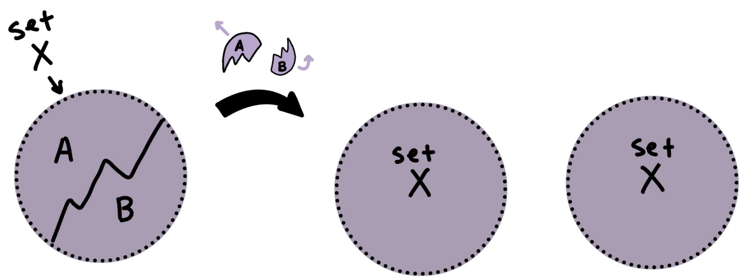
Since we are working in the complex plane, the geometric representation of taking everything in A and subtracting 1 from the real part is shifting the whole set to the left by 1 unit. So, we created the full set X out of A by shifting A by 1 unit to the left.

Similarly,

$$t^-1\tilde{B} = \tilde{X} \implies e^{-i}B = X$$

In the complex plane, multiplying a complex number z by e^{-i} results in rotating z by -1 radian counterclockwise (or equivalently, 1 radian clockwise). So, we created the full set X out of B by rotating B 1 radian clockwise.

In summary, partitioning our constructed set X into two disjoint subsets, A and B, translating one part, and rotating the other part results in two copies of the original. By following these steps, we are able to duplicate the set X by breaking it into two parts and manipulating each slightly using only rigid motions.



Did you say ”wow”?