The Banach-Tarski Theorem

Rich Schwartz

January 26, 2010

1 The Main Result

The purpose of these notes is to prove the famous Banach-Tarski Theorem. We say that $A, B \subset \boldsymbol{R}^3$ are *equivalent* if there are finite partitions into disjoint pieces,

$$A = A_1 \cup \ldots \cup A_n; \qquad B = B_1 \cup \ldots \cup B_n,$$

and isometries $I_1, ..., I_n$ such that $I_j(A_j) = B_j$ for all j. In this case, we write $A \sim B$. When $A \sim B$ it means, informally, that one can cut A into n pieces, like a puzzle, and reassemble those pieces into B. The implied map $A \to B$ is a *piecewise isometric map*.

The Banach-Tarski Theorem requires the Axiom of Choice. I'll state the precise version that is needed here.

Real Axiom of Choice (R.A.C.): Let $\{X_\alpha\}$ be a disjoint union of subsets of $\boldsymbol{R}^3$. Then there exists a set $S \subset \bigcup X_\alpha$ such that S contains exactly one element of X_α for each α.

Say that A is a *fat set* if A is bounded and A contains a ball.

Theorem 1.1 (Banach-Tarski) *Assume the R.A.C. If A and B are arbitrary fat sets, then $A \sim B$.*

What makes this theorem amazing is that A could be a tiny ball and B could be an enormous ball.

2 Reduction to a Simpler Result

The goal in this section is to reduce the Banach-Tarski Theorem to a simpler result that has a more direct proof.

We write $A \prec B$ if $A \sim B'$ for some $B' \subset B$. We write $B \succ A$ if there is a partition $B = B_1 \cup ... \cup B_n$ and isometries $I_1, ..., I_n$ such that $A \subset \bigcup I_j(B_j)$. The sets $I_1(B_1), ..., I_n(B_n)$ need not be disjoint from each other. Below we will prove the following three results.

1. If $A \prec B$ and $B \prec A$ then $A \sim B$.
2. If $B \succ A$ then $A \prec B$.
3. If $A \sim B$ and $B \sim C$ then $A \sim C$.

Before proving these results, we use them to reduce the Banach-Tarski Theorem to the following simpler result.

Theorem 2.1 (Doubling) *Assume the R.A.C. Then there exist* 3 *disjoint unit balls* A, B_1, B_2 *such that* $A \succ B$ *where* $B = B_1 \cup B_2$.

Let B_r denote the unit ball of radius r centered at the origin.

Corollary 2.2 *Assume the R.A.C. Then* $B_r \sim B_s$ *for any* $r, s > 0$.

Proof: By scaling, we can assume that $1 = r < s$. Clearly $B_1 \prec B_s$. In light of Statements 1 and 2 above, it suffices to prove that $B_1 \succ B_s$. There is some n such that B_s can be covered by 2^n translates of B_1. Iterating the Doubling Theorem n times, we see that B_1 is equivalent to 2^n disjoint translates of B_1. But then $B_1 \succ B_s$. ♠

By Statement 3, the relation $\sim$ is an equivalence relation. Hence, it suffices to prove the Banach-Tarski Theorem when $B = B_1$, the unit ball. But $B_r \subset A \subset B_s$ for some pair of balls B_r and B_s. Since $B_r \sim B_s$ and $A \subset B_s$, we have $B_r \succ A$. By Statement 2, we have $A \prec B_r$. But $B_r \prec A$. Hence $A \sim B_r$. But $B_r \sim B_1$. Hence $A \sim B_1$. This finishes the reduction.

Proof of Statement 1: This is basically the Schroeder-Bernstein Theorem. We have injective and piecewise isometric maps $f : A \to B$ and $g : B \to A$. Say that an n-chain is a sequence of the form $x_n \to ... \to x_0 \in A$, where

- $x_k \in A$ if $k > 0$ is even. In this case $f(x_k) = x_{k-1}$
- $x_k \in B$ if k is odd. In this case $g(x_k) = x_{k-1}$.

For each $a \in A$ let $n(a)$ denote the length of the longest n-chain that ends in $a = x_0$. It might be that $n(a) = \infty$. Let $A_n = \{a \in A | \ n(a) = n\}$. Swapping the roles of A and B, define B_n similarly.

Now observe that

- $f(A_k) = B_{k+1}$ for $k = 0, 2, 4, ...$
- $g^{-1}(A_k) = B_{k-1}$ for $k = 1, 3, 5$.
- $f(A_\infty) = B_\infty$.

The restriction of f to

$$A' = A_0 \cup A_2 \cup ... \cup A_\infty$$

is an injective piecewise isometry and the restriction of g^{-1} to

$$A'' = A - A' = A_1 \cup A_3 \cup A_5...$$

is also an injective piecewise isometry. (Note that A'' does not include A_∞.) Define $h(a) = f(a)$ if $a \in A'$ and $h(a) = g^{-1}(a)$ if $a \in A''$. By construction $f(A') \cap g^{-1}(A'') = \emptyset$. Hence h is an injection. Also, $B = f(A') \cup g^{-1}(A'')$. Hence h is a surjection. Hence h is a bijection. By construction h is a piecewise isometric map. ♠

Proof of Statement 2: Assume $B \succ A$. Define

- $A_1 = A \cap I_1(B_1)$.
- $A_2 = A \cap I_2(B_2) - A_1$.
- $A_3 = A \cap I_3(B_3) - A_1 - A_2$, etc.

Then $A = A_1 \cup ... \cup A_n$ is a partition. Let $B'_j = I_j^{-1} A_j$ and let $B' = \bigcup B'_j$. Then $B'_1 \cup ... \cup B'_n$ is a partition of B'. By construction $A \sim B' \subset B$. Hence $A \prec B$. ♠

Proof of Statement 3: We have partitions

$$A = A_1 \cup ... \cup A_n; \qquad B = B_1 \cup ... \cup B_n = B_1' \cup ... \cup B_m'; \qquad C = C_1' \cup ... \cup C_m'$$

and isometries $I_1, ..., I_n$ and $J_1, ..., J_m$ such that $I_j(A_j) = B_j$ and $J_j(B_j') = C_j'$. Define

$$B_{ij} = B_i \cap B_j'; \qquad A_{ij} = I_i^{-1}(B_{ij}); \qquad C_{ij} = J_j(B_{ij}),$$

and isometries $K_{ij} = J_j \circ I_i$. By construction $\{A_{ij}\}$ partitions A and $\{C_{ij}\}$ partitions C and K_{ij} isometrically carries A_{ij} to C_{ij}. Hence $A \sim C$. ♠

3 Depleted Balls

We are left to prove the Doubling Theorem. Here we will reduce the Doubling Theorem to another related result. Say that a *depleted ball* is a set of the form $B - X$, where B is a unit ball and X is a countable union of lines through the center of B.

Theorem 3.1 (Depleted Ball) *Assume the R.A.C. Then there exists a depleted ball Σ and a partition $\Sigma = \Sigma_1 \cup \Sigma_2 \cup \Sigma_3$ such that*

- *Σ_i and Σ_j are isometric for all pairs i, j.*
- *$\Sigma_3 \succ \Sigma_1 \cup \Sigma_2$.*

Corollary 3.2 *Assume the R.A.C. Then there are 9 disjoint depleted balls A, $B_1, ..., B_8$ such that $A \sim B$ where $B = B_1 \cup ... \cup B_8$.*

Proof: Iterating the conclusion of the Depleted Ball Theorem, we see that $\Sigma_1 \succ Y$, where Y is any finite union of isometric copies of Σ_1. Our Corollary follows almost immediate from this. ♠

Lemma 3.3 *Any unit ball can be covered by 4 isometric copies of any depleted ball.*

Proof: Say that a *punctured ball* is a ball minus its center. It suffices to prove that any punctured ball can be covered by two isometric copies of any depleted ball. Applying isometries, it suffices to consider the case when all the objects are subsets of the usual unit ball B. Let $C_j = B - X_j$ for $j = 1, 2$. Here X_j is a countable union of line through the origin. Let $SO(3)$ denote the rotation group of B. Let $A(i,j) \subset SO(3)$ denote the set of rotations which do not carry the ith line of X_1 to the jth line of X_2. Then $A(i,j)$ is open dense. By the Baire Category Theorem, the intersection $\bigcap A(i,j)$ is nonempty. Let I be some element of this intersection. We have $X_2 \subset I(C_1)$. Hence

$$B = C_2 \cup X_2 \subset C_2 \cup I(C_1) \subset B.$$

Hence $B = I(C_1) \cup C_2$. ♠

The last two results combine in an obvious way to imply the Doubling Theorem.

4 The Depleted Ball Theorem

It remains to prove the Depleted Ball Theorem. This is the interesting part of the proof. Consider the countable group

$$\Gamma = \langle A, B | A^3 = B^2 = \text{Identity} \rangle.$$

In other words, Γ is the group of all words in A and B subject to the relations that A^3 and B^2 are the identity word. Abstractly, Γ is the modular group. We have a partition $\Gamma = \Gamma_1 \cup \Gamma_2 \cup \Gamma_3$, where

- Γ_1 consists of those words starting with A.
- Γ_2 consists of those words starting with A^2.
- Γ_3 consists of the empty word and also those words starting with B.

We have the following structure:

$$A\Gamma_k = \Gamma_{k+1}; \qquad \Gamma_1 \cup \Gamma_2 \subset B\Gamma_3.$$

Indices are taken mod 3 for the first equation. These two algebraic facts are quite close to the conclusion of the Depleted Ball Theorem. The trick is to convert the algebra into geometry.

Let B denote the unit ball in $\boldsymbol{R}^3$ and let $SO(3)$ denote the group of rotations of B. Below we will prove the following technical lemma.

Lemma 4.1 *There exists an injective homomorphism $\rho : \Gamma \to SO(3)$.*

Essentially, a random homomorphism from Γ to $SO(3)$ will be injective. We identify A and B with their images under ρ. So, A is an order 3 rotation of B and B is an order 2 rotation of B. In general, we identify elements of Γ with their images under ρ.

A nontrivial element of $SO(3)$ is a rotation about some line through the origin. We say that a line in $\boldsymbol{R}^3$ is *bad* if it is the line fixed by some element of Γ. Since Γ is a countable group, there are only countable many bad lines. We let X denote the union of these bad lines. We let $\Sigma = B - X$. Then Σ is a depleted ball. Moreover, the group Γ acts *freely* on Σ in the following sense. If $\gamma(p) = p$ for some $\gamma \in \Gamma$ and some $p \in \Sigma$, then γ is the identity element.

We have an relation on Σ. We write $p_1 \sim p_2$ if and only if $p_1 = \gamma(p_2)$ for some $\gamma \in \Gamma$. The fact that Γ is a group implies easily that $\sim$ is an equivalence relation. This gives us an uncountable partition

$$\Sigma = \bigcup \Sigma_\alpha$$

into the equivalence classes. By the Real Axiom of Choice, we can find a new set $S \subset \Sigma$ such that S has one member in common with each S_α.

Lemma 4.2 *Let $\gamma_1, \gamma_2 \in \Gamma$ be distinct elements. Then $\gamma_1(S) \cap \gamma_2(S) = \emptyset$.*

Proof: We argue by contradiction. Suppose that $p \in \gamma_1(S) \cap \gamma_2(S)$. Then there is a point $q \in S$ such that $\gamma_1(q) = \gamma_2(q)$. But then $\gamma_2^{-1}\gamma_1(q) = q$. But Γ acts freely on Σ, giving a contradiction. ♠

Lemma 4.3

$$\Sigma = \bigcup_{\gamma \in \Gamma} \gamma(S).$$

Proof: choose $p \in \Sigma$. By construction, there is some $q \in S$ such that $p \sim q$. This means that $p = \gamma(q)$ for some $\gamma \in \Gamma$. Hence $p \in \gamma(S)$. ♠

Now define

$$\Sigma_k = \Gamma_k(S) := \bigcup_{\gamma \in \Gamma_k} (S).$$

The previous two results show that $\Sigma = \Sigma_1 \cup \Sigma_2 \cup \Sigma_3$ is a partition of Σ. At the same time

$$A(\Sigma_k) = \Sigma_{k+1}; \qquad B(\Sigma_3) = B\Gamma_3(S) \supset (\Gamma_1 \cup \Gamma_2)(S) = \Sigma_1 \cup \Sigma_2.$$

The first part of this equation shows that Σ_i and Σ_j are isometric for all i, j. The second part shows that $\Sigma_1 \cup \Sigma_2$ is isometric to a subset of Σ_3. Hence $\Sigma_3 \succ (\Sigma_1 \cup \Sigma_2)$. This proves the Depleted Ball Theorem.

5 The Injective Homomorphism

It remains to produce an injective homomorphism $\rho : \Gamma \to SO(3)$. Let X denote the set of all homomorphisms ρ. By stringing out the coordinates of $\rho(A)$ and $\rho(B)$, we can identify X with a subset of $\boldsymbol{R}^{18}$. The conditions on $\rho(A)$ and $\rho(B)$ give polynomial conditions on the coordinates of points in X. Hence X is an algebraic subvariety of $\boldsymbol{R}^{18}$.

Let $X(\gamma) \subset X$ denote the set of ρ such that $\rho(\gamma)$ is trivial. Assume for the moment that $X(\gamma)$ is always nonempty. The condition that $X(\gamma) = 0$ gives a finite list of polynomial functions on X. Any polynomial function on an algebraic variety either vanishes identically or vanishes on a nowhere dense set. From this we see that $X(\gamma)$ is open dense in X. But the countable intersection of open dense subsets of an algebraic variety is non-empty, by the Baire Category Theorem, and we choose ρ from this intersection.

It only remains to show that $X(\gamma)$ is always nonempty. Our proof seems more painful than necessary.

A *Mobius transformation* is a linear fractional transformation of $\boldsymbol{C} \cup \infty$. Each Mobius transformation has two possible matrix representatives. Given distinct $z_1, z_2 \in \boldsymbol{C}$ there is a unique Mobius transformation $T(z_1, z_2)$ such that $T(z_1, z_2)$ has order 3 and $T(z_1, z_2)$ rotates clockwise by $2\pi/3$ about z_1. Of the two possibilities, we can continuously choose a matrix representative $M(z_1, z_2)$ for $T(z_1, z_2)$.

Let N be a matrix representing the map $z \to -z$. Assuming that we have fixed the word γ, we let $F_{ij}(z_1, z_2)$ denote the (ij)th matrix entry of the matrix obtained by substituting $M(z_1, z_2)$ for A and N for B in the word γ. Then F_{ij} is a well-defined rational function on $\boldsymbol{C}^2$.

Lemma 5.1 *F_{ij} is nontrivial for some i, j.*

Proof: We can choose (z_1, z_2) so that the matrices $M(z_1, z_2)$ and N generate a group that is conjugate to the famous modular group from complex analysis. The modular group is the image of an injective homomorphism $\rho : \Gamma \to PSL_2(\boldsymbol{R})$, the group of isometries of the hyperbolic plane. But then $\rho(\gamma)$ is not the identity matrix. Hence, the matrix coefficients of $M(z_1, z_2)$ are not all constant functions. ♠

We let $F = F_{ij}$ for the indices guaranteed by the previous result. Let

$$R_\Delta = \{(z, -1/\overline{z})|\ z \in \boldsymbol{C} - \{0\}\}.$$

Lemma 5.2 *F is nonconstant on R_Δ.*

Proof: We will suppose not and derive a contradiction Consider the following rational map on $\boldsymbol{C}^2$.

$$\theta(z_1, z_2) = \Big(z_1 + 1/z_2, i(z_1 - 1/z_2)\Big).$$

By construction $\theta(R_\Delta)$ is open in $\boldsymbol{R}^2$. The function $\theta \circ F \circ \theta^{-1}$ is a rational function on $\boldsymbol{C}^2$ that is constant on an open subset of $\boldsymbol{R}^2$. This forces $\theta \circ F \circ \theta^{-1}$ to be globally constant. But then F is globally constant as well. ♠

Now we finish our proof. We are trying to show that the set $X(\gamma)$ is nonempty. We conjugate our whole picture by stereographic projection

$$s : S^2 \to \boldsymbol{C} \cup \infty.$$

The maps $\rho(A)$ and $\rho(B)$ are then conjugated to Mobius transformations. Choosing $\rho(B)$ so that it fixes the vertical axes in $\boldsymbol{R}^3$, we see that the Mobius transformation $s\rho(B)s^{-1}$ is precisely the map $z \to -z$. The map s carries antipodal points to points of the form $(z, -1/\overline{z})$. Hence $s\rho(A)s^{-1}$ is a map of the form $T(z, 1/\overline{z})$. By the previous lemma, we can choose z such that $s\rho(\gamma)s^{-1}$ is nontrivial. But then $\rho(\gamma)$ is nontrivial for the homomorphism ρ corresponding to z.