A Complete, Mechanically-Verified Proof of the Banach-Tarski Theorem in ACL2(R)

Jagadish Bapanapally
Department of Computer Science, University of Wyoming, Laramie, WY, USA

Ruben Gamboa
Department of Computer Science, University of Wyoming, Laramie, WY, USA

Abstract

This paper presents a formal proof of the Banach-Tarski theorem in ACL2(r). The Banach-Tarski theorem states that a unit ball can be partitioned into a finite number of pieces that can be rotated to form two identical copies of the ball. We have formalized 3D rotations and generated a free group of 3D rotations of rank 2. In prior work, the non-denumerability of the reals was proved in ACL2 (r), and a version of the Axiom of Choice that can consistently select a representative element from an equivalence class was introduced in ACL2 version 3.1. Using the free group of rotations, and this prior work, we show that the unit sphere can be decomposed into two sets, each equivalent to the original sphere. Then we show that the unit ball except for the origin can be decomposed into two sets each equivalent to the original ball by mapping the points of the unit ball to the points on the sphere. Finally, we handle the origin by rotating the unit ball around an axis such that the origin falls inside the sphere. Seemingly paradoxically, the construction results in two copies of the unit ball.

2012 ACM Subject Classification Theory of computation → Logic and verification

Keywords and phrases ACL2(r), Banach-Tarski, Rotations

Digital Object Identifier 10.4230/LIPIcs.ITP.2022.5

Supplementary Material *Software (Source Code)*:
`https://github.com/acl2/acl2/tree/master/books/nonstd/nsa/Banach-Tarski`
archived at `swh:1:dir:964ff589fe9739a46201bbeac71a6733aa4f274b`

Funding *Jagadish Bapanapally*: The research presented in this paper was partially supported by a grant from IOG Singapore Pte. Ltd.

Acknowledgements We want to thank professor John Cowles at the University of Wyoming for assisting us in the proof verifying the denumerability of the poles.

1 Introduction

The Banach-Tarski theorem [11] states that we can break the unit ball into a finite number of sets, then rotate the sets to form two identical copies of the unit ball. This seems impossible because it breaks our intuition that when we partition the ball into finite sets, the total volume of the pieces must be the same as the volume of the original ball. This would be the case if all the pieces had a well-defined volume. The Banach-Tarski theorem is possible because the construction breaks the ball into non-measurable sets [7], which means they don't have a well-defined volume. Such a partition of the unit ball is obviously subtle, and the entire construction depends on the Axiom of Choice [7] and the non-denumerability of reals [4]. Many properties of matrix algebra [5], modular arithmetic [1] and trigonometric functions [3] that are needed for the proof have already been formalized in ACL2(r).

In this paper we present a complete proof of the Banach-Tarski theorem in ACL2(r) [6], a variant of ACL2 that offers support for the real numbers by the way of non-standard analysis. The ACL2(r) source files for this proof are in the ACL2 community books under the directory nonstd/nsa/Banach-Tarski/. We begin in Section 2, with a free group of reduced words using

lists. The lists in this group are decomposed in various ways, resulting in multiple ways of reconstructing the free group. This is, in fact, a key step in the Banach-Tarski paradox. Then, in Section 3, we present a free group of rotations that correspond to the free group of reduced words, as shown in Section 4. We also formalize 3D rotations and prove some crucial properties of rotations.

Using the free group of rotations, in Section 5 we show how almost all of the unit sphere can be partitioned into sets, such that the sets can be rotated and rearranged to form two copies of the unit sphere, almost. This works for all points on the sphere, except the points that lie on the axis of rotation of one of the rotations in the free group. The set of such points is countable, and this is shown by proving some basic facts about constructing countable sets. This establishes the Hausdorff's Paradox. The proof then proceeds by adding some extra rotations that essentially wipe the poles of the rotations. This proves the Banach-Tarski paradox on the unit sphere.

Finally, in Section 6 the construction is extended to the unit ball. Except for the origin, this is done by projecting each point inside the unit ball onto the unit sphere, and using the decomposition of the unit sphere constructed in Section 5. The final step is to account for the origin by introducing a rotation around a point close enough to the origin that the origin is always mapped to a point inside the unit sphere.

Throughout, the proof of the Banach-Tarski theorem involves proving a lot of equivalences between various sets. In our proof of the Banach-Tarski theorem we use predicates and Skolem functions to represent various sets, then prove equivalence between these predicates.

2 A Free Group of Reduced Words

In this section, we introduce the free group over the letters a and b. This group contains all words that can be formed from a, b, a^{-1}, and b^{-1} such that no letter and its inverse appear together. For example, $abba$ is a member of this free group but $abb^{-1}a$ is not.

We use lists in ACL2(r) to represent words. A weak word is an empty list or a list that has characters a or a^{-1} or b or b^{-1}. e.g., $'(a\ b\ b^{-1}\ a^{-1})$ is a weak word. The single quote in the example means that it is a list, which would otherwise become a function call. In the ACL2(r) source files, we have defined the functions `wa`, `wa-inv`, `wb` and `wb-inv` which return the ACL2(r) characters $\#\backslash a$, $\#\backslash b$, $\#\backslash c$, and $\#\backslash d$ respectively. e.g., (`wa`)=$\#\backslash a$. We use the ACL2(r) characters $\#\backslash a$, $\#\backslash b$, $\#\backslash c$, and $\#\backslash d$ to represent a, a^{-1}, b, and b^{-1} respectively, but in this paper we will simply refer to a, a^{-1}, b, and b^{-1} to avoid confusion. The predicate *weak-wordp* recognizes elements of the set of weak words, as shown in Listing 1. Since ACL2(r) does not have support for infinite sets, such as the set of weak words, we represent these sets implicitly using recognizers for their elements.

A reduced word is a weak word such that character a^{-1} does not appear beside the character a and character b^{-1} does not appear beside the character b in the list. e.g., $'(a\ b\ a^{-1})$ is a reduced word and $'(a\ a^{-1}\ b)$ is not a reduced word. The predicates *a-wordp*, *a-inv-wordp*, *b-wordp*, and *b-inv-wordp* represent the set of reduced words that start with characters a, a^{-1}, b, and b^{-1} respectively. The predicate *reducedwordp*, shown in Listing 2, represents the set of all reduced words. *reducedwordp* returns true if the argument belongs to the set *a-wordp* or *a-inv-wordp* or *b-wordp* or *b-inv-wordp* or if it is an empty list.

The function *word-inverse* finds the inverse of a reduced word. If the argument is a weak word, *word-inverse* flips each character in the list to its inverse and then reverses the list. e.g., *word-inverse*$('(a\ a^{-1}\ b^{-1})) = '(b\ a\ a^{-1})$. Listing 3 shows the definition of the flip function and the inverse function.

Listing 1 Definition of the set of weak words.

```
(defun weak-wordp (w)
  (cond ((atom w) (equal w nil))
        (t (and (or (equal (first w) (wa))
                    (equal (first w) (wa-inv))
                    (equal (first w) (wb))
                    (equal (first w) (wb-inv)))
                (weak-wordp (rest w))))))
```

Listing 2 Definition of the set of reduced words.

```
(defun reducedwordp (x)
  (or (a-wordp x)
      (a-inv-wordp x)
      (b-wordp x)
      (b-inv-wordp x)
      (equal x '())))
```

The group operation *compose* takes two arguments. If the arguments are weak words, then the *compose* function first appends the two lists and then "fixes" the result by deleting any letter and its inverse that appear beside each other. Thus, the final result of *compose* is always a reduced word. e.g., *compose*('(*a b b*), '(b^{-1})) = '(*a b*). Listing 4 shows the definition of the fixing function and the group operation *compose*.

If $w1$ and $w2$ are reduced words, then (append $w1$ $w2$) is a weak word. If x is a weak word, then *word-fix*(x) returns a reduced word. So, *compose*($w1, w2$) = *word-fix*(append $w1$ $w2$) is a reduced word. This establishes that *compose* is closed over the set of reduced words. In fact, *compose* is a group operator over reduced words, as suggested earlier. A key lemma required to prove that it satisfies the associative property and the inverse property is that if x is a reduced word, then *word-fix*(rev(x)) = (rev(*word-fix*(x))), which we proved by induction on x. This proves that, with the group operation *compose* and considering the empty list as the identity element, the set of reduced words is a free group. Listing 5 shows the group properties of this set.

Listing 3 Definition of the Inverse operation.

```
;; Definition of the flip function
(defun word-flip (x)
  (cond ((atom x) nil)
        ((equal (car x) (wa)) (cons (wa-inv) (word-flip (cdr x))))
        ((equal (car x) (wa-inv)) (cons (wa) (word-flip (cdr x))))
        ((equal (car x) (wb)) (cons (wb-inv) (word-flip (cdr x))))
        ((equal (car x) (wb-inv)) (cons (wb) (word-flip (cdr x))))))

;; Definition of the Inverse operation
(defun word-inverse (x)
  (rev (word-flip x)))
```

Listing 4 Definition of the group operation *compose*.

```
;; Definition of the fixing function
(defun word-fix (w)
  (if (atom w)
      nil
    (let ((fixword (word-fix (cdr w))))
      (let ((w (cons (car w) fixword)))
        (cond ((equal fixword nil)
               (list (car w)))
              ((equal (car (cdr w)) (wa))
               (if (equal (car w) (wa-inv))
                   (cdr (cdr w))
                   w))
              ((equal (car (cdr w)) (wa-inv))
               (if (equal (car w) (wa))
                   (cdr (cdr w))
                   w))
              ((equal (car (cdr w)) (wb))
               (if (equal (car w) (wb-inv))
                   (cdr (cdr w))
                   w))
              ((equal (car (cdr w)) (wb-inv))
               (if (equal (car w) (wb))
                   (cdr (cdr w))
                   w)))))))

;; Definition of the group operation
(defun compose (x y)
  (word-fix (append x y)))
```

Listing 5 Group properties of the set of reduced words.

```
;; Closure property
(defthmd closure-prop
  (implies (and (reducedwordp x)
                (reducedwordp y))
           (reducedwordp (compose x y)))
  :hints ...)

;; Associative property
(defthmd assoc-prop
  (implies (and (reducedwordp x)
                (reducedwordp y)
                (reducedwordp z))
           (equal (compose (compose x y) z)
                  (compose x (compose y z))))
  :hints ...)

;; Inverse property
(defthmd reduced-inverse
  (implies (reducedwordp x)
           (equal (compose x (word-inverse x))
                  '()))
  :hints ...)
```

Denote the set of reduced words by $W(a,b)$, the set of reduced words starting with character a by $W(a)$, and similarly for $W(a^{-1})$, $W(b)$, and $W(b^{-1})$. Then clearly $W(a,b) = () \bigsqcup W(a) \bigsqcup W(a^{-1}) \bigsqcup W(b) \bigsqcup W(b^{-1})$, where $\bigsqcup$ denotes the union of *disjoint* sets. In addition to this we can show two other equivalences of the set of reduced words:

- $W(a,b) = a^{-1}W(a) \bigsqcup W(a^{-1})$ and
- $W(a,b) = b^{-1}W(b) \bigsqcup W(b^{-1})$,

where $xW(y) = \{xw \mid w \in W(y)\}$.

As we mentioned previously, we use recognizers to represent sets, and since ACL2(r) supports quantifiers via Skolem functions, we represent the set $a^{-1}W(a) = \{x \mid \exists w \in W(a) \text{ s.t. } x = compose('(a^{-1}), w)\}$ and $b^{-1}W(b) = \{x \mid \exists w \in W(b) \text{ s.t. } x = compose('(b^{-1}), w)\}$. The formal proof follows the proof of the two equivalences given in [11]. If an element x belongs to $W(a)$ then $compose('(a^{-1}), x) \subset () \bigsqcup W(a) \bigsqcup W(b) \bigsqcup W(b^{-1})$, because the *compose* function appends $'(a^{-1})$ and x and then deletes the first two characters a^{-1} and a in the appended list as they are inverses of each other. Moreover, the first character of $a^{-1}x$ is the second character of x, so it cannot be a^{-1}, since x is a reduced word that starts with a.

Likewise, if an element x belongs to $W(a)$ or $W(b)$ or $W(b^{-1})$ then there exists an element *word-a* that belongs to the set $W(a)$ such that x equals to $(compose('(a^{-1}), word\text{-}a)$, namely the element $a^{-1}x$. So, we have $a^{-1}W(a) = () \bigsqcup W(a) \bigsqcup W(b) \bigsqcup W(b^{-1})$. The same way we prove $b^{-1}W(b) = () \bigsqcup W(a) \bigsqcup W(a^{-1}) \bigsqcup W(b)$. With these two equivalences of the sets $a^{-1}W(a)$ and $b^{-1}W(b)$ we get two corollaries Corollary 2 and Corollary 3 which we use to prove the Banach-Tarski theorem on S^2.

▶ **Corollary 1.** $W(a,b) = () \bigsqcup W(a) \bigsqcup W(a^{-1}) \bigsqcup W(b) \bigsqcup W(b^{-1})$

▶ **Corollary 2.** $W(a,b) = a^{-1}W(a) \bigsqcup W(a^{-1})$

▶ **Corollary 3.** $W(a,b) = b^{-1}W(b) \bigsqcup W(b^{-1})$

Notice that these corollaries, while being about lists, already contain the key to the Banach-Tarski paradox. The set $W(a,b)$ is decomposed into five disjoint subsets, then it can be reconstructed in two different ways by taking the union of two of those subsets after prepending a letter to one of the subsets. In the same way, the sphere can be deconstructed into a number of sets, which can then be rotated and reassembled in two different ways to reconstruct a unit sphere.

3 A Free Group of 3D Matrices

Matrices in ACL2 are represented with the data structure *array2p*. We define a predicate *r3-matrixp* that recognizes the set of 3D matrices: *r3-matrixp* returns true if the argument is of type *array2p* and if its dimensions are 3×3, and if each element of the matrix is a real number.

We define now the four matrices A^+, A^-, B^+, and B^- as

$$A^{\pm} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{3} & \mp\frac{2\sqrt{2}}{3} \\ 0 & \pm\frac{2\sqrt{2}}{3} & \frac{1}{3} \end{bmatrix} \quad B^{\pm} = \begin{bmatrix} \frac{1}{3} & \mp\frac{2\sqrt{2}}{3} & 0 \\ \pm\frac{2\sqrt{2}}{3} & \frac{1}{3} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and we associate these matrices with the letters a, a^{-1}, b, and b^{-1} from the free group respectively. Moreover, we associate a list $(x_1, x_2, \ldots, x_n) \in W(a,b)$ with the matrix $X_1 \times X_2 \times \cdots \times X_n$, where $\times$ denotes matrix multiplication, and X_i is the matrix associated with letter x_i. We have defined a recursive function *rotation* that performs this mapping

from words in the free group to rotations. Denote the resulting set $R(a,b)$. i.e., $R(a,b) = \{\mathit{rotation}(w) \mid w \in W(a,b)\}$. By induction, it is easy to verify that every element of the set $R(a,b)$ belongs to *r3-matrixp*.

To show the set $R(a,b)$ is a free group isomorphic to $W(a,b)$, we show that if $w \in W(a,b)$ and w is not the empty list, then $\mathit{rotation}(w)$ is not equal to I, the identity matrix. Equivalently, we show that $(\mathit{rotation}(w))(0,1,0) \neq (0,1,0)$ unless w is the empty list.

To do this, suppose that $w \in R(a,b)$, and consider the rotation $R(w)$. In particular, suppose that $R(w)$ rotates the point $(0,1,0)$ to (x',y',z'). Define (x,y,z) as

$$(x,y,z) = 3^n \left(\frac{x'}{\sqrt{2}}, y', \frac{z'}{\sqrt{2}}\right)$$

where $n = |w|$. Using induction, we showed that x, y, and z are integers.

So now, suppose that $(\mathit{rotation}(w))(0,1,0) = (0,1,0)$ for some non-empty word w. It follows that $(x,y,z) = (0,3^n,0)$, where $n = |w| > 0$, thus $x \equiv y \equiv z \equiv 0 \pmod 3$. But this cannot be the case. If $|w| = 1$, then $\mathit{rotation}(w)$ is one of $A^{\pm}$ or $B^{\pm}$, and considering each of the four cases by brute force, it is clear that $(x,y,z) \not\equiv (0,0,0) \pmod 3$. Using induction, there are 16 cases to consider, but in all of these cases we again conclude that $(x,y,z) \not\equiv (0,0,0) \pmod 3$. This shows that if $|w| > 0$, then $\mathit{rotation}(w)$ is not the identity matrix.

Using this fact, the group properties of $W(a,b)$, and the associativity of the matrix multiplication, we then showed that there is a one-to-one relation between the set $R(a,b)$ and the set $W(a,b)$. So defining $R(a) = \{\mathit{rotation}(w) \mid w \in W(a)\}$, $R(a^{-1}) = \{\mathit{rotation}(w) \mid w \in W(a^{-1})\}$, $R(b) = \{\mathit{rotation}(w) \mid w \in W(b)\}$, and $R(b^{-1}) = \{\mathit{rotation}(w) \mid w \in W(b^{-1})\}$, then the set of rotations $R(a,b)$ can be partitioned as

$$R(a,b) = I \bigsqcup R(a) \bigsqcup R(a^{-1}) \bigsqcup R(b) \bigsqcup R(b^{-1}).$$

That is, the paradoxical partition of the free words $W(a,b)$ from Section 2 can be reproduced in the set of rotations $R(a,b)$.

4 A Free Group of Rotations of Rank 2

Before proceeeding directly into the Banach-Tarski construction, we need to prove some basic facts from matrix algebra. As discussed previously, the matrix transpose operation (*m-trans*) was formalized in prior work [5], and as part of that, it was shown that $(A \times B)^T = B^T \times A^T$.

We extended that formalization by introducing the function *r3-m-determinant* that computes the determinant of a matrix, the function *r3-m-inverse* that computes the inverse of a 3D matrix (when possible). Using these functions, we defined the predicate *r3-rotationp* that recognizes rotations in $\mathbb{R}^3$.

▶ **Definition 4.** *A matrix M is a rotation matrix if it satisfies these conditions [10]:*

- *M is a 3D matrix,*
- *$M^{-1} = M^T$, and*
- $\det(M) = 1$.

Another important detail is that every element of $R(a,b)$ must be a rotation of $\mathbb{R}^3$. Given the correspondence between $R(a,b)$ and $W(a,b)$ established in Section 3, what we need to show is that for any $w \in W(a,b)$, $\mathit{rotation}(w)$ satisfies the axioms in Definition 4. This was done using induction on the list w. It is easy to verify that the base cases are rotations; i.e., I, A^+, A^-, B and B^- are all rotation matrices. For the induction to go through, the lemma we need to prove $\mathit{rotation}(xw)$ is a rotation in $\mathbb{R}^3$ given that $\mathit{rotation}(w)$ is a rotation, is that the product of two rotation matrices M_1 and M_2 is also a rotation matrix.

The final lemma from matrix algebra that we needed was to show that every rotation matrix preserves distances [9]; i.e., that $||Mx|| = ||x||$ whenever M is a rotation matrix and x is a vector. Since the focus of this project was on the Banach-Tarski paradox and not matrix algebra, we proceeded to prove these results as directly as possible, without using deeper results from linear algebra, such as the geometric meaning of determinants. In the end, we proceeded using the roadmap suggested by the following lemmas, all proved in ACL2(r):

▶ **Lemma 5.** $r3\text{-}matrixp(m_1) \wedge r3\text{-}matrixp(m_2) \implies r3\text{-}matrixp(m_1 \times m_2)$

▶ **Lemma 6.** $r3\text{-}matrixp(m_1) \wedge r3\text{-}matrixp(m_2) \implies \det(m_1 \times m_2) = \det(m_1) \cdot det(m_2)$

▶ **Lemma 7.** $r3\text{-}matrixp(m) \implies m \times I = I \times m = m$

▶ **Lemma 8.** $r3\text{-}matrixp(m) \wedge \det(m) \neq 0 \implies m \times m^{-1} = m^{-1} \times m = I$

▶ **Lemma 9.** $r3\text{-}matrixp(m_1) \wedge \det(m_1) \neq 0 \wedge r3\text{-}matrixp(m_2) \wedge \det(m_2) \neq 0$
$\implies (m_1 \times m_2)^{-1} = m_2^{-1} \times m_1^{-1}$

▶ **Lemma 10.** $r3\text{-}rotationp(m_1) \wedge r3\text{-}rotationp(m_2) \implies r3\text{-}rotationp(m_1 \times m_2)$

▶ **Lemma 11.** $r3\text{-}rotationp(m) \implies r3\text{-}rotationp(m^{-1})$

▶ **Lemma 12.** *Rotations preserve distances.*

Proof. Let $p_1 = (x_1, y_1, z_1)$ and R be a rotation matrix, and consider $p_2 = Rp_1 = (x_2, y_2, z_2)$. Using the previous lemmas,

$$\begin{aligned} x_1^2 + y_1^2 + z_1^2 &= p_1^T \times p_1 \\ &= p_1^T \times (I \times p_1) \\ &= p_1^T \times ((R^{-1} \times R) \times p_1) \\ &= p_1^T \times ((R^T \times R) \times p_1) \\ &= (p_1^T \times R^T) \times (R \times p_1) \\ &= (R \times p_1)^T \times (R \times p_1) \\ &= p_2^T \times p_2 \\ &= x_2^2 + y_2^2 + z_2^2. \end{aligned}$$

◀

5 Banach-Tarski Theorem on the Unit Sphere

Before finishing the proof of the Banach-Tarski theorem for the unit sphere, we want to mention two key lemmas needed to carry out the proof. First, if $w_1, w_2 \in W(a, b)$, then by the definition of *rotation* and *compose*, $rotation(w_1) \times rotation(w_2) = rotation(compose(w_1, w_2))$. Second, if $r \in R(a, b)$, then $\exists w \in W(a, b)$ such that $r = rotation(w)$, and by the previous lemma $r^{-1} = rotation(w^{-1})$. Moreover, since $w^{-1} \in W(a, b)$, $r^{-1} \in R(a, b)$.

Returning to the main proof, let D be the set of poles of all of the rotations belonging to the set $R(a, b) - I$; i.e., $D = \{p \in S^2 \mid \exists r.\, r \in R(a, b) \wedge r \neq I \wedge r(p) = p\}$.

Now, consider a point $p \in S^2 - D$ and $r \in R(a, b)$. It follows that $r(p) \in S^2 - D$ as well. Otherwise, $r(p) \in D$ and by the definition of D there exists a witness $r_w \in R(a, b) - I$ such that $r_w(r(p)) = r(p)$. But then $r^{-1}(r_w(r(p))) = p$. By the previous lemmas, $r^{-1} r_w r \neq I \implies p \in D$. This proves if $r \in R(a, b)$ and $p \in S^2 - D$, then $r(p) \in S^2 - D$.

Define the orbit of a point $p \in S^2 - D$ as $\{r(p) \mid r \in R(a,b)\}$. Using the Axiom of Choice, implemented as `defchoose` in ACL2, we can choose one representative of each of these orbits. Let M be the set of all of the chosen points from each of the orbits. In Section 5.1 we will show how we used the Axiom of Choice in our proof and how we decomposed $S^2 - D$ into two sets each equivalent to $S^2 - D$. Then in Section 5.2 we'll show the set D is countable; i.e., we will show all the poles of rotations belonging to $R(a,b)$ can be enumerated. Since S^2 is not countable, there exists a point $P_{s2} \in S^2 - D$. Then in Section 5.3 we find an angle $a_{s2} \in [0, 2\pi)$ such that the rotation of any point in D around the axis from the origin to P_{s2} by an angle that is a multiple of a_{s2}, the resulting point does not lie in the set D. The remainder of the proof decomposes S^2 into two sets each equivalent to S^2 by proving equivalences between different sets as suggested in Section 1.

5.1 Decomposing the Unit Sphere minus the Set of Poles

ACL2 supports existential quantification by the way of the `defun-sk` event [8]. We have defined the orbit of a point $point = \{o\text{-}point \mid \exists w.\, w \in W(a,b) \wedge o\text{-}point = rotation(w) \times point\}$ as a Skolem function using defun-sk as shown below.

```
(defun-sk orbit-point-p-q (o-point point)
  (exists w
          (and (reducedwordp w)
               (m-= (m-* (rotation w (acl2-sqrt 2))
                         point)
                    o-point))))
```

The function *orbit-point-p-q* returns true if the point *o-point* belongs to the orbit of *point* and it chooses a witness reduced word w such that $o\text{-}point = rotation(w) \times point$.

Now using the Axiom of Choice we want to choose one representative from each of the orbits of the points in the set $S^2 - D$. The Axiom of Choice in ACL2 is implemented using `defchoose` which was previously used in the proof of the Vitali's theorem [2]. So the choice set M is defined as follows:

```
(defchoose choice-set-s2-d-p (c-point) (p)
  (and (point-in-r3 c-point)
       (orbit-point-p-q c-point p))
  :strengthen t)
```

In the definition of *choice-set-s2-d-p*, *point-in-r3* is the predicate that recognizes points in $\mathbb{R}^3$. If $p \in S^2$, then *choice-set-s2-d-p*(p) picks a point *c-point* in $\mathbb{R}^3$ that is in the orbit of the point p. The *strengthen* option in the choice function ensures that the same canonical witness is chosen for any other point *p1* in the same equivalence class as p.

Since M contains one representative from each of the orbits of the points belonging to the set $S^2 - D$, $S^2 - D = R(a,b)M$. For example, below is how we define the set $R(a,b)M = \{p \mid \exists p_1.\, p_1 \in S^2 - D \wedge c_{p1}$ is the chosen point from the orbit of $p_1 \wedge \exists r \in R(a,b).\, r \times c_{p1} = p\}$. Similarly, we define the sets M, $R(a)M$, $R(a^{-1})M$, $R(b)M$, $R(b^{-1})M$, $a^{-1}R(a)M$, and $b^{-1}R(b)M$.

```
(defun-sk diff-s2-d-p-q-1 (cp1 p)
  (exists w
          (and (reducedwordp w)
               (m-= (m-* (rotation w (acl2-sqrt 2)) cp1) p))))

(defun-sk diff-s2-d-p-q (p)
  (exists p1
          (and (s2-d-p p1)
               (diff-s2-d-p-q-1 (choice-set-s2-d-p p1)
                                p))))

;; Definition of the set R(a,b)M
(defun diff-s2-d-p (p)
  (and (point-in-r3 p)
       (diff-s2-d-p-q p)))
```

If the sets M, $R(a)M$, $R(a^{-1})M$, $R(b)M$, $R(b^{-1})M$ are disjoint, and the sets $a^{-1}R(a)M$, $R(a^{-1})M$ are disjoint, and $b^{-1}R(b)M$, $R(b^{-1})M$ are disjoint, then we have our decomposition of $S^2 - D$. Suppose that $R(a)M$ and $R(b)M$ are not disjoint. To simplify the discussion, define $\hat{p}$ as the point chosen for the orbit of p, and R_w as the rotation matrix of the word w.

Now, let p be a point in the intersection, i.e., $p \in R(a)M$ and $p \in R(b)M$. Then $\exists p_a \in S^2 - D$ and $\exists w_a \in W(a)$ such that $R_{w_a} \times \hat{p_a} = p$ and $\exists p_b \in S^2 - D$ and $\exists w_b \in W(b)$ such that $R_{w_b} \times \hat{p_b} = p$. Since $\hat{p_a}$ lies in the orbit of p_a, $\exists w_{pa} \in W(a, b)$ such that $R_{w_{pa}} \times p_a = \hat{p_a}$ and $\exists w_{pb} \in W(a, b)$ such that $R_{w_{pb}} \times p_b = \hat{p_b}$. So, $R_{w_a} \times R_{w_{pa}} \times p_a = p = R_{w_b} \times R_{w_{pb}} \times p_b$, which implies p_a and p_b belong to the same orbit. In other words, $\hat{p_a} = \hat{p_b}$, since those are the representatives points for their (one) orbit. Since, $R_{w_a} \times \hat{p_a} = R_{w_b} \times \hat{p_b}$, we have that $R_{w_b^{-1} w_a} \times \hat{p_a} = R_{w_b^{-1}} \times R_{w_a} \times \hat{p_a} = \hat{p_b} = \hat{p_a}$. Notice that $compose(w_b^{-1}, w_a) \neq$ '(), since w_b^{-1} ends with b^{-1} and w_a starts with a. Thus, $R_{w_b^{-1} w_a} \neq I$ which implies that $\hat{p_a} \in D$. But this is a contradiction since $\hat{p_a}$ is in the orbit of p_a. So, the sets $R(a)M$ and $R(b)M$ must be disjoint. Similar arguments show that the other sets are also disjoint. By the definition of $R(a, b)$ and by Corollary 2, Corollary 3 we can transfer the decomposition of $W(a, b)$ into the following decompositions of the set $S^2 - D$. Thus, the set $S^2 - D$ can be decomposed into two disjoint copies of itself. Listing 6 shows the proof of these decompositions of $S^2 - D$ in ACL2(r) where *s2-d-p* is the recognizer for the set $S^2 - D$, *diff-n-s2-d-p* is the recognizer for the set M, *diff-a-s2-d-p* is the recognizer for the set $R(a)M$, *diff-a-inv–s2-d-p* is the recognizer for the set $R(a^{-1})M$, *diff-b-s2-d-p* is the recognizer for the set $R(b)M$, *diff-b-inv-s2-d-p* is the recognizer for the set $R(b^{-1})M$, *a-inv-diff-a-s2-d-p* is the recognizer for the set $a^{-1}R(a)M$, and *b-inv-diff-b-s2-d-p* is the recognizer for the set $b^{-1}R(b)M$.

$$S^2 - D = R(a, b)M = M \bigsqcup R(a)M \bigsqcup R(a^{-1})M \bigsqcup R(b)M \bigsqcup R(b^{-1})M$$
$$S^2 - D = a^{-1}R(a)M \bigsqcup R(a^{-1})M$$
$$S^2 - D = b^{-1}R(b)M \bigsqcup R(b^{-1})M$$

5.2 The Set of Poles is Countable

We have seen that $S^2 - D$ can be decomposed, and that the pieces can be recombined in two different ways to create two copies of $S^2 - D$. We now want to show set D is countable.

Listing 6 Decompositions of the set $S^2 - D$ in ACL2(r).

```
;; Unit sphere minus the set of poles broken down into 5 sets
(defthmd s2-d-p-equivalence-1
  (iff (s2-d-p p)
       (or (diff-n-s2-d-p p)
           (diff-a-s2-d-p p)
           (diff-a-inv-s2-d-p p)
           (diff-b-s2-d-p p)
           (diff-b-inv-s2-d-p p)))
  :hints ...)

;; A copy of the unit sphere minus the set of poles
(defthmd s2-d-p-equivalence-2
  (iff (s2-d-p p)
       (or (a-inv-diff-a-s2-d-p p)
           (diff-a-inv-s2-d-p p)))
  :hints ...)

;; Another copy of the unit sphere minus the set of poles
(defthmd s2-d-p-equivalence-3
  (iff (s2-d-p p)
       (or (b-inv-diff-b-s2-d-p p)
           (diff-b-inv-s2-d-p p)))
  :hints ...)
```

Let p be a point in D. Then, there exists a non-empty word $w \in W(a,b)$ such that $R_w p = p$. R_w is a rotation matrix, so it has the form

$$R_w = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix}.$$

If R_w is symmetric, then $R_w = R_w^T = R_w^{-1}$. But since $R_w^{-1} = R_{w^{-1}}$, we have that $R_{w^{-1}} = R_w$. But this is not possible, since we have also shown that the mapping from $W(a,b)$ to $R(a,b)$ is one-to-one. So R_w can not be symmetric, thus at least one of $m_{32} \neq m_{23}$, $m_{13} \neq m_{31}$, or $m_{21} \neq m_{12}$ must be true.

Let $K = \sqrt{(m_{32} - m_{23})^2 + (m_{13} - m_{31})^2 + (m_{21} - m_{12})^2}$. Since R_w is not symmetric, $K \neq 0$. So consider the point

$$f_p = \frac{1}{K}(m_{32} - m_{23}, m_{13} - m_{31}, m_{21} - m_{12}).$$

By computation, it is easy to verify that $R_w f_p = f_p$ and $R_w(-f_p) = -f_p$. So indeed, f_p and $-f_p$ are poles of R_w. Now we show these are the only poles of the rotation R_w; i.e., we show the original point $p \in D$ is either equal to f_p or $-f_p$. By construction, $R_w p = p$, and this implies that $p = R_w^{-1} p$, hence $p = R_w^T p$. This means that p satisfies the equations $R_w x = R_w^T x$ and $||x|| = 1$. Geometrically, the solutions to the first equation lie on a line through the origin, and the solutions to the second equation lie on the unit sphere, so the intersection of these results in two points. Algebraically, we proved that the only solutions to these equations are $x = f_p$ and $x = -f_p$.

Using this fact, we can now define an enumeration of all the poles; i.e., we define a sequence that contains all the poles of any rotation R_w corresponding to a non-empty reduced word $w \in W(a,b)$.

The first step is to enumerate all the words in $W(a, b)$. We do this by defining the function *generate-words-main* that returns all the possible words (including weak words, like '$(a\ a^{-1}\ b)$) up to a given input length. It is straightforward to prove that all words in $W(a, b)$ eventually appear in this sequence. Using this enumeration, we then enumerate all the poles by replacing a word w in the sequence with its corresponding pair of poles. The function *poles-list* returns the n^{th} pole, and we proved that all poles appear somewhere in this sequence.

This establishes that the poles are countable, and since the points on the sphere are not, there is at least one point p on the sphere that is not a pole. In the next section, we will use this point to decompose the entire unit sphere.

5.3 Decomposing the Unit Sphere

Up to this point, we have been working with rotations of the form R_w where w is a reduced word. Now, we consider general rotations around a line that passes through the origin and an arbitrary point in the sphere S^2. We defined the function *rotation-3d* that takes an angle $\theta \in [0, 2\pi)$ and a point p in S^2 and returns the matrix corresponding to that general rotation. We will use $R_{p,\theta}$ to denote this general rotation matrix.

Recall from Section 5.2 that there is a point p that lies on the unit sphere but is not one of the poles; i.e., there is a point p such that $p \in S^2 - D$. We would like to choose $\theta \in [0, 2\pi)$ such that for any $p' \in D$, $R_{p,\theta}p' \notin D$; i.e., the rotation $R_{p,\theta}$ rotates D away from D. More than that, we want to choose θ so that even if we rotate multiple times by θ the result is still not in D; i.e., for any $p' \in D$, $R^n_{p,\theta}p' \notin D$ for any $n > 1$. Before finding this θ, we observe (and proved formally in ACL2(r)) that $R_{p,\theta_1} \times R_{p,\theta_2} = R_{p,\theta_1+\theta_2}$ and by induction $R^n_{p,\theta} = R_{p,n\theta}$.

So consider the set of all angles α such that they rotate some element of D to an element of D, perhaps by rotating multiple times; i.e., consider the set

$$A = \left\{\alpha \mid \alpha \in [0, 2\pi) \wedge n \in \mathbb{Z}^+ \wedge \exists p'.\, p' \in D \wedge R_{p,n\alpha}p' \in D\right\}.$$

Now, any angle γ can be written uniquely as $\gamma = 2\pi k + \beta$, where k is an integer and $\beta \in [0, 2\pi)$. In particular, we showed in ACL2(r) that for the positive angle $n\alpha$, there is a unique non-negative integer k and an angle $\beta \in [0, 2\pi)$ such that $n\alpha = 2\pi k + \beta$.

Moreover, suppose p_1 and p_2 are in D, and that there is an angle $\alpha \in [0, 2\pi)$ and a positive integer n such that $R_{p,n\alpha}p_1 = p_2$. As observed, $n\alpha$ can be written uniquely as $n\alpha = 2\pi k + \beta$, which means that the angle α itself can be written as $\frac{2\pi k+\beta}{n}$. Moreover, the angle β is uniquely determined by $n\alpha$, and α is uniquely determined by the choice of p_1, p_2, and n. So enumerating the possible values of n (positive integer) and k (non-negative integer) will also enumerate all the possible values of $\alpha \in A$.

We formalized the proof in ACL2(r) that the Cartesian product of two countable sets is also countable, and we used this result to show that the set A is countable since the sets of possible n and k values as well as pairs (p_1, p_2) are countable. As before, since A is a countable set of angles in $[0, 2\pi)$ there must be some angle $\theta \in [0, 2\pi)$ that is not in A. This angle θ satisfies the desired condition, namely that for any $p' \in D$ and $n \geq 1$, $R^n_{p,\theta}p' = R_{p,n\theta}p' \notin D$.

What we have at this point is a rotation matrix $R_{p,\theta}$ that maps the set of poles P to somewhere in $S^2 - D$. It is easy to verify that if $m \neq n$, then $R_{p,n\theta} \neq R_{p,m\theta}$. Now, consider the set $E = D \bigsqcup R_{p,\theta}D \bigsqcup R_{p,2\theta}D \bigsqcup R_{p,3\theta}D \bigsqcup \cdots$. From the definition of E, it follows easily that $R_{p,\theta}E = E - D$. Thus, the set $S^2 - D$ can be decomposed as

$$S^2 - D = (S^2 - E) \bigsqcup (E - D) = (S^2 - E) \bigsqcup R_{p,\theta}E.$$

With a bit of tedious algebraic manipulation, this formula can be used to find a disjoint decomposition of the entire surface S^2:

$$S^2 = \big((S^2 - D) \cap (S^2 - E)\big) \bigsqcup R_{p,-\theta}\big((S^2 - D) \cap E\big).$$

$S^2 - E$ is just a set, so this equality has the form

$$S^2 = \big((S^2 - D) \cap F\big) \bigsqcup R_{p,-\theta}\big((S^2 - D) \cap E\big).$$

In Section 5.1, we showed how $S^2 - D$ could be decomposed into disjoint sets such that the pieces could be rotated and recombined to create two copies of $S^2 - D$. Replacing $(S^2 - D)$ in the equality above with those two decompositions of $S^2 - D$ results in a similar decomposition of S^2 (but with many more terms). This establishes the Banach-Tarski theorem for the entire sphere S^2. Listing 7 shows the proof of the Banach-Tarski theorem on S^2 in ACL2(r) where *s2-def-p* is the recognizer for the set S^2.

Listing 7 Decompositions of S^2 in ACL2(r).

```
;; Unit sphere broken down into 14 sets
(defthmd s2-equiv-1
  (iff (s2-def-p p)
       (or (set-a1 p)
           (set-a2 p)
           (set-a3 p)
           (set-a4 p)
           (set-a5 p)
           (set-a6 p)
           (set-a7 p)
           (set-a8 p)
           (set-a9 p)
           (set-a10 p)
           (set-a11 p)
           (set-a12 p)
           (set-a13 p)
           (set-a14 p)))
  :hints ...)

;; A copy of the unit sphere
(defthmd s2-equiv-2
  (iff (s2-def-p p)
       (or (set-a-inv-a3 p)
           (set-a-inv-r-a4 p)
           (set-r-1-a-inv-a5 p)
           (set-r-1-a-inv-r-a6 p)
           (set-a7 p)
           (set-a8 p)))
  :hints ...)

;; Another copy of the unit sphere
(defthmd s2-equiv-3
  (iff (s2-def-p p)
       (or (set-b-inv-a9 p)
           (set-b-inv-r-a10 p)
           (set-r-1-b-inv-a11 p)
           (set-r-1-b-inv-r-a12 p)
           (set-a13 p)
           (set-a14 p)))
  :hints ...)
```

6 Banach-Tarski Theorem on the Unit Ball

In Section 5, we showed how the unit sphere S^2 can be decomposed into a finite collection of disjoint sets such that the subsets can be rotated and recombined to construct two copies of S^2. In this section, we use that fact to define a similar construction for the unit ball B^3.

First, we will decompose the unit ball except the origin. Suppose that $p \in B^3 - \{0\}$, and let $r = ||p||$. Define the point $p' = p/r$. It is easy to show that $p' \in S^2$. Geometrically, it is obvious that if we rotate a point on S^2, then all the points along the line from the origin to that point will be rotated by the same angle and direction. We proved this fact algebraically in ACL2(r). Using this fact, it is trivial to generalize the Banach-Tarski decomposition of S^2 into a similar decomposition of $B^3 - \{0\}$.

Generalizing the decomposition to the entire unit ball B^3 is conceptually similar to the way the decomposition of $S^2 - D$ was extended to cover all of S^2. The trick, then, was to find a rotation that would essentially erase the points in D, and this was possible because D is countable. The origin is just a single point, so the same strategy of rotating the origin away should work. The major complication is that any rotation with an axis that passes through the origin will map the origin to itself. So we need to consider rotations along arbitrary axes, and these are not linear transformations, so they cannot be simply encoded as matrices.

Besides the linear transformation of rotation, we also need translation. We defined the function, *rotation-about-arbitrary-line* that accepts an arbitrary point p, an angle θ, and an axis of rotation l (defined using two points), and returns the result of rotating p around l by θ. We proved that this operation satisfies the expected properties of rotations in 3D.

Four of these properties were needed to complete the proof. First, the result of rotating a point p by an angle θ around an axis l is always a point in $\mathbb{R}^3$. Second, if $\theta = 0$, rotation around any axis l by θ is the identity transformation. Third, rotating a point p by an angle θ_1 about an axis l and then rotating the result by an angle θ_2 about the same axis l, is the same as rotating the point p by $\theta_1 + \theta_2$ around l. Finally, the result of rotating the origin around a specific axis l that is close to the origin results in a point that is inside the unit ball B^3. The last two properties combine to show that repeatedly rotating the origin around this specific axis by θ will always yield a result that is inside B^3.

The rest follows the same strategy presented in Section 5.3. Fix the axis of rotation l as above, so that origin is always mapped to some point inside B^3. Now we want to find an angle $\alpha \in [0, 2\pi)$ such that if we rotate the origin by an angle $n\alpha$ around l, the result is never the origin; i.e., let $R_{l,\theta}p$ be the result of rotating p around l by θ. Then $R_{l,\theta}0, R_{l,2\theta}0, R_{l,3\theta}0, \ldots$ is a countably infinite sequence of points that are all inside B^3.

We find a suitable α by partially solving the equation $R_{l,n\theta}0 = 0$. In particular, we showed that this requires that $\cos(n\theta) = 1$, and this means that θ must have the form $\theta = \frac{2\pi k}{n}$, where n is a positive integer and $k \in \mathbb{Z}$. Similar to the construction in Section 5.3, the set of possible θ can be enumerated, so there must be at least one angle $\alpha \in [0, 2\pi)$ that is *not* one of the θ. Thus, $R_{l,n\alpha}0 \neq 0$ for any positive integer n.

Exactly as before, let $Z = \{R_{l,n\alpha}0 \mid n \in \mathbb{N}\}$. Then $B^3 - \{0\} = (B^3 - Z) \bigsqcup R_{l,\alpha}Z$. This is then used to show that

$$B^3 = \big((B^3 - \{0\}) \cap (B^3 - Z)\big) \bigsqcup R_{l,-\alpha}\big((B^3 - \{0\}) \cap Z\big).$$

And just as before, replacing $(B^3 - \{0\})$ with the decompositions found above yields a decomposition of all of B^3 that satisfies the Banach-Tarski paradox. Listing 8 shows the decompositions of B^3 in ACL2(r) where *b3* is the recognizer of the set B^3.

Listing 8 Decompositions of B^3 in ACL2(r).

```
;; Unit ball broken down into 52 sets
(defthmd b3-equiv-1
  (iff (b3 p)
       (or (b3-00 p)
           (b3-01 p)
           ...
           ...
           (b14-00 p)
           (b14-01 p)
           (set-b20 p)
           (set-b10 p)
           (rota-1-b3-10 p)
           ...
           ...
           (rota-1-b14-11 p)
           (rota-1-b21 p)
           (rota-1-b11 p)))
  :hints ...)

;; A copy of the unit ball
(defthmd b3-equiv-2
  (iff (b3 p)
       (or (rot-3-b3-00 p)
           (rot-3-b3-10 p)
           ...
           ...
           (rot-8-b8-00 p)
           (rot-8-b8-10 p)
           (rota-1-rot-3-b3-11 p)
           (rota-1-rot-3-b3-01 p)
           ...
           ...
           (rota-1-rot-8-b8-11 p)
           (rota-1-rot-8-b8-01 p)))
  :hints ...)

;; Another copy of the unit ball
(defthmd b3-equiv-3
  (iff (b3 p)
       (or (rot-9-b9-00 p)
           (rot-9-b9-10 p)
           ...
           ...
           (rot-14-b14-00 p)
           (rot-14-b14-10 p)
           (rota-1-rot-9-b9-11 p)
           (rota-1-rot-9-b9-01 p)
           ...
           ...
           (rota-1-rot-14-b14-11 p)
           (rota-1-rot-14-b14-01 p)))
  :hints ...)
```

7 Conclusion

In this paper we have presented a formalization of the Banach-Tarski theorem in ACL2(r). Although ACL2(r) may not be the obvious choice to formalize such an abstract theorem, it turns out that the key step in the proof is reasoning about free groups, and since this is tantamount to reasoning about lists, it is perfectly natural for theorem provers in the Boyer-Moore family of provers, like ACL2(r). Moreover, even though there is very limited support for quantification in ACL2(r), we have shown that we can define complex structures and prove properties about them. The proof also makes use of 3D rotations, and we formalized these rotations and proved many key properties about them. We have formalized the proof for the Cartesian product of two countable sets is countable, and used this proof in the decomposition of S^2 and B^3. The proven properties of 3D rotations from section 4 and countable sets are readily available to use by anyone who chooses to do so, and these are available in rotations.lisp and countable-sets.lisp in the ACL2(r) source files. We have used many properties of modular arithmetic and trigonometric functions, and these were previously formalized in ACL2(r). Also critical in a few steps was the fact that certain sets can be enumerated, but that no non-trivial interval of reals can be – and this had also been proved in prior work. The end result is a proof of the Banach-Tarski paradox: The unit ball B^3 in $\mathbb{R}^3$ can be decomposed into finitely many pieces that can be rotated and reassembled to form two copies of B^3.

References

1 Piergiorgio Bertoli and Paolo Traverso. *Design Verification of a Safety-Critical Embedded Verifier*, pages 233–245. Kluwer Academic Publishers, USA, 2000.

2 John Cowles and Ruben Gamboa. Using a first order logic to verify that some set of reals has no lesbegue measure. In *International Conference on Interactive Theorem Proving*, pages 25–34. Springer, 2010.

3 Ruben Gamboa. *Mechanically Verifying Real-Valued Algorithms in ACL2*. PhD thesis, Citeseer, 1999.

4 Ruben Gamboa and John Cowles. A Cantor Trio: Denumerability, the Reals, and the Real Algebraic Numbers. In *International Conference on Interactive Theorem Proving*, pages 51–66. Springer, 2012.

5 Ruben Gamboa, John Cowles, and JV Baalen. Using ACL2 Arrays to Formalize Matrix Algebra. In *Fourth International Workshop on the ACL2 Theorem Prover and Its Applications (ACL2'03)*, volume 1, 2003.

6 Ruben A. Gamboa and Matt Kaufmann. Nonstandard Analysis in ACL2. *J. Autom. Reason.*, 27(4):323–351, November 2001. doi:10.1023/A:1011908113514.

7 Thomas J Jech. *The Axiom of Choice*. Courier Corporation, 2008.

8 J. Strother Moore. Milestones from the pure lisp theorem prover to acl2. *Form. Asp. Comput.*, 31(6):699–732, December 2019. doi:10.1007/s00165-019-00490-3.

9 Rotation matrix. Rotation matrix – Wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/Rotation_matrix, 2021. Online; Accessed: 2022-02-04.

10 Madeline Tremblay. The Banach-Tarski Paradox. unpublished, 2017.

11 Tom Weston. The Banach-Tarski Paradox. *Citado*, 2:15, 2016.