

$(n + 3)$ -Coloring the n -Sphere: An Impossibility Theorem

Joel Clarke Gibbons Ph.D.

Logistic Research & Trading Co.

4052 Niles Road

Saint Joseph, Michigan, 49085

Abstract

We address a fundamental combinatorial proposition for the n -sphere and a similarly fundamental proposition in inversive geometry on the n -sphere, and demonstrate the intimate connection between them. Specifically, in terms of combinatorial geometry, we show that any coloring of the n -sphere by $n + 3$ colors must $(n + 2)$ -color some $(n - 1)$ -sphere. In regard to inversive geometry, we characterize the structure of the class of smallest subsets of the n -sphere that has the property that if T is a well-defined function of the n -sphere that preserves $(n - 1)$ -spheres and if the image of T contains a member of this class, T must be an inversive transformation.

Keywords: circle-preserving map, sphere-preserving map, inversive transformation, coloring the n -sphere.

1. Introduction

Conventionally a Coloring of a manifold is a partition into well-defined, non-empty sets. The cells of the partition—the colors—are not restricted in any other way. The cardinality of the coloring is the index set of the partition, and is often appended to the description of the coloring, so an m -coloring of the n -sphere is a partition of S^n into m sets. I will prove the following, for all $n \geq 2$.

Theorem 1.n. Any $(n + 3)$ -coloring of S^n must $(n + 2)$ -(or $n + 3$)-color some $(n - 1)$ -sphere. It is impossible to $(n + 3)$ -color the n -sphere while ensuring that no $(n - 1)$ -sphere meets more than $(n + 1)$ colors.

Corollary 1. Under the same conditions, there is a nested tower of spheres such that the sphere of dimension m is $(m + 3)$ -colored, for $m = 2, 3, \dots, (n - 1)$. In particular, there is a 4-colored circle. This follows trivially. Conversely, Theorem 1.n is equivalent to the proposition that an $(n + 3)$ -coloring must four-color some circle because starting from a four-colored circle one can simply build up to an $(m + 3)$ -coloring of an m -sphere that contains the circle. The process starts from the four-colored circle, and chooses a point off the circle representing a fifth color. The two-sphere spanned by the circle and the point is a five-colored two sphere. Proceed to higher dimensions by iterating this construction.

These propositions address the Combinatorial Geometry of the sphere, and are related to well known results of [1]. The logic of this theorem and of Erdős' findings in what he called Euclidean Ramsey Theory are more in the nature of dual than parallel. Nonetheless, it would be fair to characterize Theorem 1.n as being a proposition in Spherical Ramsey Theory. The Euclidean analog of this theorem was proved in [2]. Our main result can however be stated in a seemingly different way, as a proposition in Inversive Geometry. Specifically,

Definition 1.1. A subset M of the m -sphere lies in *General Position* if for any $(n - 1)$ -sphere S , the Complement of S contains at least two points of M . Obviously, M must contain at least $n + 3$ points. If M lies in general position, every set containing M lies in general position.

Definition 1.2. A map T of the n -sphere to itself is *Circle-Preserving* if for every circle C and set X of four points on C , if $T(X)$ contains four points they must lie on the circle $T(C)$. If $T(X)$ contains three or fewer points, this condition is met for that four-tuple trivially. It is not assumed that the

image of the four points on C be four distinct points, or that T is necessarily a continuous function.

Definition 1.3. A map T of the n -sphere is *Sphere-Preserving* if for every $(n-1)$ -sphere S in S^n if X is a set of $(n+2)$ points on S , and if $T(X)$ contains $n+2$ points, they must lie on a $(n-1)$ -sphere. The same qualification applies to T with regard to continuity.

Sphere-preserving implies circle-preserving trivially, but the converse is by no means trivial. The Inversive Group of transformations of S^n consists of all the $(n-1)$ -sphere-preserving bijections. Inversive transformations are of two kinds. One kind are the Linear Fractional Transformations and the other kind are the composition of Linear Fractional Transformations with the reflection of the sphere through the equator. The former kind are orientation preserving and the latter kind are orientation reversing, but ignoring the direction of angles, all are conformal. The term “inversive” derives from the fact that they are factorable into a composition of inversions through $(n-1)$ -spheres. The Linear Fractional Transformations consist of four general types: parallel translations, rotations about an $(n-2)$ -sphere, dilations, and a compound of translation and dilation that expands from one point and contracts on another point.

Theorem 2.n. Let M be a subset of S^n in general position, and let $T : S^n \rightarrow S^n$ be a sphere-preserving map. If the image of T contains M , T must be an inversive transformation.

Theorem 2.n is a statement about how many points are needed to “determine” a sphere-preserving map. In [4] we proved that if M contains at least $(2n+2)$ points, then the map T must be an inversive transformation, so the practical effect of this paper is to reduce from $2n+2$ to $n+3$ the assumed number of points in M . Every Inversive Transformation is determined by its values on $n+2$ points in the domain. If however we do not assume that T is a bijection, and in fact is we do not even assume that it is a continuous function, more information is needed. Surprisingly, only one more pair $(z, T(z))$ is needed, as long as the set of $n+3$ points lie in general position.

Theorem 2.n admits the following corollary.

Corollary 2. Let M be a subset of S^n consisting of $n+3$ points lying in general position and let T be a sphere-preserving map of the n -sphere. T is an inversive transformation if and only if $T(M)$ lies in general position, so given that T is sphere-preserving, it is sufficient to check its values on a set

of $n + 3$ points. I will defer the proof of this equivalence until after proving Theorem 2.n.

In [3] and [4] we recounted some history of the problem of characterizing subsets of the sphere which have the property that every circle-preserving map of S^2 need only be evaluated on that set. To recapitulate briefly, this problem was first addressed by Carathéodory in the 1930s. The six point characterization in [4] was really the next attempt at this question. Theorem 2.n above completes this account. There is no broader class of subsets of the n -sphere than the class of all $(n + 3)$ points sets in general position. We will present examples to demonstrate this.

The proof of Theorem 1.n will occupy most of this paper because the statement in terms of coloring seems to be the more tractable statement to deal with. I will take up the equivalence of Theorems 1.n and 2.n at the end of this paper. Theorem 1.n is true for all $n \geq 2$, but the proof for $n = 2$ and for $n > 2$ are different, so it will be necessary to present them separately. We cannot simply proceed from the $n = 2$ case by induction on n , but once we have proved the case $n = 3$, we can then generalize by induction from there. Accordingly, we must treat the case $n = 2$ and the case $n = 3$ individually.

2. Preliminaries

Notation. I will denote the circle through points x, y , and z by (xyz) , the two-sphere through x, y, z , and w by $(xyzw)$, and so on for all higher dimensions. Denote the line through x and y by $[xy]$.

Proposition 2.1. There is an $(n + 2)$ -coloring of S^n which has no $(n + 2)$ -colored $n - 1$ sphere, so the cardinality $(n + 3)$ in Theorem 1.n is the minimum necessary.

PROOF. Viewing the n -sphere as the extended Euclidean space, let the origin $\underline{0}$ use one of the colors and ∞ use another one. Denoting the axes X_1, X_3, \dots, X_n , apply the third color to $X_1 - \{\underline{0}, \infty\}$, the fourth color to the remainder of the $X_1 - X_2$ plane, the fifth color to the remainder of the $X_1 - X_2 - X_3$ three-space, and so on. Any $(n - 1)$ -sphere that does not contain the origin or the point at infinity lies in a space that is covered by n colors. Any sphere that contains either the origin or the point at infinity, but not both, still lies in a space covered by $n + 1$ colors. The remaining $(n - 1)$ -spheres to check are the hyperplanes through the origin. They meet

both of the first two colors, but they cannot meet all n of the colors, so they are at most $(n + 1)$ -colored. We demonstrate this fact in the following way.

Consider the $(n - 1)$ -dimensional hyperplane through the origin $\{(x_1, \dots, x_n) | \sum b_i x_i = 0\}$ for some vector b of coefficients. If $b_1 \neq 0$, the hyperplane cannot contain the X_1 axis except for the origin, because on that axis $x_2 = \dots = x_n = 0$, so $b_1 x_1 = 0$. If the X_1 axis does not lie in the hyperplane, we are done; the hyperplane contains no points of color three. Otherwise, we conclude that $b_1 = 0$. If $b_2 \neq 0$, the hyperplane cannot meet the $X_1 - X_2$ plane, minus the X_1 axis, because $x_3 = \dots = x_n = 0$, so $b_2 x_2 = 0$. Proceeding inductively, we find either that we identify the color that does not meet the hyperplane—the color corresponding to the first non-zero b —or the b 's are all zero. If all the b 's are zero the “hyperplane” is the entire n -dimensional space, which is a contradiction. It is possible to $(n + 2)$ -color S^n without $(n + 2)$ -coloring any $(n - 1)$ -sphere. QED

Definition 2.1. In dimension n , the *singular coloring* consists of a unique $(n - 1)$ -sphere S which is $(n + 2)$ -colored and the complement of S , which is monochromatic. Thus while the color number result holds, it does not imply that the coloring is in any way regular. In terms of inversive transformations, this corresponds to a discontinuous sphere-preserving map, i.e., T is the identity on the unique sphere S and constant on the complement of S .

3. Proof of Theorem 1.2

The method of proof will be to start by supposing that a coloring exists in violation of the theorem, to then characterize such a coloring, and from the characterization to adduce a contradiction which implies that the assumed coloring is itself impossible. For definiteness we will refer to a five-coloring that has no four-(or five)-colored circles as a *Wild Coloring*. Throughout we will use the symbols Γ and Γ_i to denote the coloring and its cells, $i = 1, \dots, 5$. First however,

The following argument hinges on a single construction. Choose points x_i in Γ_i , $i = 1, \dots, 5$. The circles $(x_1 x_2 x_3)$ and $(x_1 x_4 x_5)$ intersect at two points, x_1 and another point x'_1 . Let T be the Linear Fractional Transformation.

$$\begin{aligned} T(x_1) &= \underline{0}, \text{ the origin,} \\ T(x'_1) &= \infty, \text{ and} \\ T(x_5) &= \underline{1}. \end{aligned}$$

Clearly, Γ is a wild coloring if and only if $T(\Gamma)$ is wild. T maps the circle $(x_1x_4x_5)$ to the x -axis and maps $(x_1x_2x_3)$ to a line through the origin that I will refer to as the v -axis. Since the circles were not necessarily orthogonal, these axes are not generally orthogonal. (Note the T is conformal.) Luckily, the angle of elevation of the v -axis drops out of the calculations entirely, and presents no problem.

The x -axis meets colors Γ_1, Γ_4 , and Γ_5 . The v -axis meets Γ_1, Γ_2 , and Γ_3 . I will designate points on the x -(v -) axis by x (resp v), and use subscripts to designate colors where they are known. The construction is illustrated in the figure below.

For an arbitrary point v on the v -axis, the circle (vx_4x_5) meets the v -axis at a second point $h(v|x_4, x_5)$. If the circle is tangent to the v -axis, $h(v) = v$, but that presents no problem. We need to have the exact formula for the function h . The circle through $v, h(v), x_4$, and x_5 has its center at (a, b) and radius r .

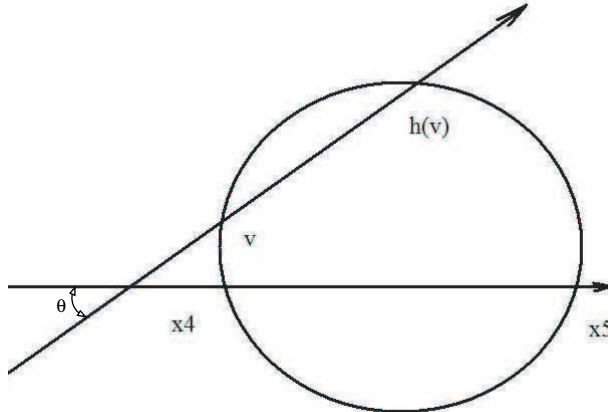


Figure 1: Circle meets the v -axis.

1. $(x_4 - a)^2 + b^2 = r^2$,
 2. $(x_5 - a)^2 + b^2 = r^2$,
 3. $(v \cos \theta - a)^2 + (v \sin \theta - b)^2 = r^2$, and
 4. $(h(v) \cos \theta - a)^2 + (h(v) \sin \theta - b)^2 = r^2$.
- From Equations 1 and 2,
5. $a = (x_4 + x_5)/2$.
- From Equations 1 and 3,

$$6. v^2 \cos^2 \theta - 2av \cos \theta - x_4^2 + 2ax_4 + v^2 \sin^2 \theta - 2bv \sin \theta = 0.$$

From 1 and 4,

$$7. h^2(v) \cos^2 \theta - 2ah(v) \cos \theta - x_4^2 + 2ax_4 + h^2(v) \sin^2 \theta - 2bh(v) \sin \theta = 0$$

Equations 6 and 7 simplify to

$$8. v^2 - 2av \cos \theta - x_4^2 + 2ax_4 = 2bv \sin \theta \text{ and}$$

$$9. h^2(v) - 2ah(v) \cos \theta - x_4^2 + 2ax_4 = 2bh(v) \sin \theta.$$

Equations 8 and 9 imply that

$$(v^2 - 2av \cos \theta - x_4^2 + 2ax_4)/v = (h^2(v) - 2ah(v) \cos \theta - x_4^2 + 2ax_4)/h(v).$$

Substituting for a and collecting terms,

$$10. v^2 h(v) + x_4 x_5 h(v) = v h^2(v) + x_4 x_5 v.$$

If $h(v) = v$, we have our formula. Otherwise, the solution of 10 is

$$11. h(v|x_4, x_5) = x_4 x_5 / v.$$

Lemma 3.1. This calculation can be generalized in the following way. Given lines X and V , parameterized in polar coordinates so that the two have the same scale, and points x and x' on X and v and v' on V , and a circle that meets both lines at these four points, the following relationship must hold:

$$12. xx' = vv'.$$

Specifically, in terms of the coloring of S^2 , if $x \in \Gamma_4$ and $x' \in \Gamma_5$, v and v' must always be the same color. Equally, if $v \in \Gamma_2$ and $v' \in \Gamma_3$, x and x' must have the same color. Since Γ_1 appears on both lines, if either x or v is in Γ_1 the coloring of the points on the other line is somewhat indefinite. If for instance x is in Γ_1 , then either v or v' , or both of them, can be color in Γ_1 also. In that case however it is still impossible for $v \in \Gamma_2$ and $v' \in \Gamma_3$. At most one of those colors is possible. In what follows, let $V_i = \Gamma_i \cap V$, $i = 1, 2, 3$, and $X_i = \Gamma_i \cap X$, $i = 1, 4, 5$.

Equation 12 yields both the function of V to itself cited above and a function k from X to itself:

$$13. k(x|v_2, v_3) = v_2 v_3 / x, \text{ where } x \text{ and } k(x) \text{ have the same color. Note that}$$

Lemma 3.2. Equation 12 and 13 imply that X_5 is closed under reciprocals and multiplication. X_5 is a multiplicative group, that we will hereafter denote G .

PROOF. Define functions m and n as follows:

$$14.a \quad m(v|x_5, x'_5) = h(h(v|x_4, x_5)|x_4, x'_5) = (x'_5/x_5)v$$

$$14.b \quad m(v|x_4, x'_4) = h(h(v|x_4, x_5)|x'_4, x_5) = (x'_4/x_4)v$$

$$15.a \quad n(x|v_2, v'_2) = k(k(x|v_2, v_3)|v'_2, v_3) = (v'_2/v_2)x$$

$$15.b \quad n(x|v_3, v'_3) = k(k(x|v_2, v_3)|v_2, v'_3) = (v'_3/v_3)x.$$

Since $1 \in X_5 = G$,

$m(v|1, g) = gv$ for all v , but from 14 we have

$n(1|v_2, v'_2) = v'_2/v_2$, so the ratio of any two numbers in V_2 is in G .

Specifically,

$1/g = (v/gv) \in G$. So G is closed under reciprocals.

It is also closed under multiplication. Choose g and g' in G . QED

$$m(gv_2|1, g') = gg'v_2 = g''v_2 \in V_2.$$

This implies that $gg' = g''$ is in G .

For any v_2 in V_2 , $V_2 = \{gv_2|g \in G\}$. Equally, $V_3 = \{gv_3|g \in G\}$ for v_3 in V_3 . Fix an element of V_2 , which we will denote v_2^0 .

$$k(1|v_2^0, v_3) = v_2^0v_3 = g. \text{ Let } v_3^0 = v_3/g.$$

$$v_3^0 = 1/v_2^0.$$

Furthermore, the reciprocal of every element in V_2 is an element of V_3 . The converse is also true of course. Now, from 14,

$$k(x_4|v_2^0, v_3^0) = 1/x_4 \in X_4,$$

so X_4 is closed under reciprocals. From 14.b,

$$m(v_2^0|x_4, x'_4) = (x'_4/x_4)gv_2.$$

Thus, $(x'_4/x_4) = ((x'_4x_4)gv_2)/(gv_2) \in G$. There is an element, x_4^0 of X_4 such that $X_4 = \{gx_4^0|g \in G\}$.

$k(x_4^0|v_2^0, v_3^0) = 1/x_4^0 = gx_4^0$, so x_4^0 is a real square root of $1/g$ for some g . Now consider the quotient group $H = (R^1/G)$. X_4 is a coset of G , and therefore it is a single element of H . (More precisely, X_4/G is an element of H), and the element in question is the square root of a number in G . Call this, as an element of H , χ_4 . Since $(x_4^0)^2 = g$, which is in G , $(\chi_4)^2 = 1$.

x_4^0 cannot equal 1 because $1 \in X_5$. Therefore, we can take $x_4^0 = -1$, and $X_4 = \{-g|g \in G\}$. Now, from Equation 12,

$$h(v|x_4^0, 1) = h(v|-1, 1) = -1/v, \text{ so if } v \in V_i, (-1/v) \in V_i.$$

For $v_2 \in V_2, 1/v_2 \in V_3$, as we demonstrated previously. Therefore for every v_2 in V_2 , $-v_2$ is in V_3 . Lastly however, this implies that $k(1/g|-v_3, v_3) = -v_3^2g = g'$. Therefore, v_3^0 is the square root of some number x^0 in X_4 . Say $v_3^0 = \sqrt{x^0}$. (i.e., $x^0 = -g'/g$).

Let $G^* = G \cup X_4$. For $g, g' \in G$ and $x, x' \in X_4, xg \in X_4$. Furthermore, $xx' \in G$ and $gg' \in G$. Together, this implies that G^* is a multiplicative group. Since every element of V_2 is a G multiple of v_2^0 and every element of V_3 is a G multiple of $v_3^0 = 1/v_2^0$, the product of two numbers in V_i is a g multiple of x^0 , for $i = 1, 2$, and any cross product $v_2v_3 \in G$. Therefore $V_2 \cup V_3$ is isomorphic to G^* , and is therefore a group. It must contain the number 1. Assume that 1 is in V_2 . Then $1/1 = 1$ is in V_3 , but this leads to a contradiction, because it violates the condition that V_2 and V_3 be disjoint. QED

4. Proof of Theorem 1.3

Now taking up the case of $n = 3$, we begin with a small lemma.

Lemma 4.1. Let $(x_i x_j x_k)$ be a circle that meets colors i, j , and k , and suppose x'_i is a fourth point on the circle which also lies in Γ_i . Neither of the arcs $(x_i x'_i)$ nor $(x'_i x_i)$ —roughly speaking, the two open semi-circles—can be monochromatic.

PROOF. Assume the contrary. For definiteness, let $i = 1, j = 2$, and $k = 3$. As usual, where the color of a point is known, I will identify it with the subscript. Up to an inversive transformation, we can arrange that the circle is a line through the origin and that $x_1 = \underline{0}$ and $x'_1 = \infty$. Call this line L . The semicircles then correspond to the upper and lower rays from $\underline{0}$. Assume that x_2 and x_3 both lie in the lower ray; i.e. assume that the positive ray is monochromatic.

Choose x_4 and consider the plane spanned by the lines L and $[\underline{0}x_4]$. For any point x in the upper half plane—where now “upper” and “lower” are defined by taking the line $[\underline{0}x_4]$ to be the other axis. The circle through x, x_2 , and x_4 meets the upper ray, which lies in Γ_1 . Hence it lies in the union

$\Gamma_1 \cup \Gamma_2 \cup \Gamma_4$. The same x also lies on the circle $(xx_3x_4) \subset \Gamma_1 \cup \Gamma_2 \cup \Gamma_4$. Hence x must lie in the intersection of these sets, which is $\Gamma_1 \cup \Gamma_4$. Thus the upper half-plane lies in $\Gamma_1 \cup \Gamma_4$. If there is an $x^* \in \Gamma_4$ on the upper half-plane, then the circle $(x^*x_2x_3)$ intersects the upper half-plane on a monochromatic arc in Γ_4 , because that circle cannot meet Γ_1 . If any x in the interior of either the first or second quadrant lies in Γ_1 , the arc of (xx_2x_3) in the upper half-plane is a subset of Γ_1 . The upper half-plane is therefore covered by the pencil of circles $\{(xx_2x_3) \mid x \text{ in the upper half-plane}\}$, where $(xx_2x_3) \cap (\text{upper half-plane})$ is a monochromatic arc. There would have to be an x^{**} in the upper half-plane with the property that it has one color—assume it is in Γ_4 for definiteness—and also lies in the closure of Γ_1 . This is necessary since the quadrants are connected sets. (Alternatively, there is an x^{**} in Γ_1 and also in the closure of Γ_4 .) The arc through x^{**} , the arc $(x^{**}x_2x_3) \cap (\text{upper half-plane})$, would then constitute a part of the boundary of Γ_4 , and the boundary of Γ_1 . Any circle that cuts this arc at x^{**} would also meet arcs that lie in Γ_1 . In other words, any circle except $(x^{**}x_2x_3)$ itself would meet all four colors, which is impossible by assumption. It follows that x^{**} cannot exist, implying that the upper half-plane is monochromatic.

Choose x_5 not lying in this plane, and the third axis $[\underline{0}x_5]$. By repeating the above argument, we can show that the upper half space is monochromatic. A 2-sphere cannot be tangent to the $x_4 - x_5$ plane at both x_4 and x_5 , so if x is any point in the lower half-space, the 2-sphere $(xx_2x_4x_5)$ meets the upper half-space somewhere. Thus, the lower half space lies in $\Gamma_1 \cup \Gamma_2 \cup \Gamma_4 \cup \Gamma_5$. Therefore, the whole space is covered by the union of these four colors, contradicting the assumption that Γ is a six-coloring. QED

PROOF OF THEOREM 1.3. Consider a circle $(x_1x_2x_3)$ and a 2-sphere $(x_1x_4x_5x_6)$ which meet at x_1 and x'_1 . By applying an inversive transformation we can arrange that $x_1 = \underline{0}$ and $x'_1 = \infty$, that the sphere is rotated to the $x - y$ plane, and that the circle is rotated to a line through the origin. Call the line X . $X \subset \Gamma_1 \cup \Gamma_2 \cup \Gamma_3$. The plane lies in the union $\Gamma_1 \cup \Gamma_4 \cup \Gamma_5 \cup \Gamma_6$. From the previous lemma, we can assume that the ray of X lying above the plane contains a point in Γ_2 and the ray below the plane contains a point of Γ_3 . (Otherwise, x_2 lies in the ray below the origin and x_3 lies above the origin.) The plane contains points x_4, x_5 , and x_6 , and corresponding rays $[\underline{0}x_4]$, $[\underline{0}x_5]$, and $[\underline{0}x_6]$. Note that the ray through $x_j, j = 4, 5, 6$, must be bi-chromatic because the two lines through the origin, X and $[\underline{0}x_i]$ span a plane (a 2-sphere) which by assumption can only meet four colors: colors 1, 2, 3, and i .

Let $(x_2x_3x_4x_5)$ be chosen so that the origin lies in its interior, which is guaranteed by the fact that the origin lies between x_2 and x_3 . It intersects the plane on a circle that contains the origin in its interior, and therefore it must meet the ray $[0x_6]$. It is therefore a five-colored sphere. QED

5. Proof of Theorem 1.n

The two pieces of the argument—lemmas 4.1 and the theorem itself—generalize directly, as I will show. All the work is involved in generalizing the lemma.

Lemma 5.1. Let x_1, x'_1, x_2 , and x_3 be points on a circle. Neither arc connecting x_1 and x'_1 can be monochromatic.

PROOF. For simplicity, arrange as before that x_1 be the origin, and x'_1 be the point at infinity, so the circle is a line through the origin. The proposition amounts to the statement that if x_2 lies on one side of the origin, there must be an x_3 on the other side. We proceed sequentially choosing points $x_4, x_5, \dots, x_{(n+2)}$, arranged so that all of them lie on the boundary of the upper half-space. If x_2 and x_3 must both lie on the lower ray, we can demonstrate that the upper half-space is monochromatic, and that the whole n -space is covered by the union $\Gamma_1 \cup \Gamma_2 \cup \Gamma_4 \cup \dots \cup \Gamma_{(n+2)}$. This contradicts the assumption that Γ is an $(n + 3)$ -coloring.

Theorem 1.n, for $n > 3$.

PROOF. The construction for all n simply duplicates the construction for $n = 3$. The only thing we need to verify is that the rays $[0x_j], j = 4, \dots, (n + 3)$, must be bi-chromatic. Any such ray together with the line called X above spans a 2-plane, which can contain at most four colors. Otherwise, if this plane is five-colored there must be a four-colored circle or line, as was demonstrated in §2. Starting with a four-colored circle, select a point in Γ_5 and generate the 2-sphere spanned by the circle and the point. It is a 5-colored 2-sphere. Choose an x_6 not on this sphere and generate the 3-sphere spanned by the two-sphere and x_6 . Continue out to $(n + 1)$ and a point $x_{(n+2)}$ to obtain an $(n + 2)$ -colored $(n - 1)$ -sphere. This proves that the rays are bi-chromatic or a contradiction ensues.

Now consider the $(n - 1)$ -sphere containing $(x_2x_3 \dots x_{(n+2)})$. It must meet the ray $[0x_{(n+3)}]$ since the origin, 0 , lies in the interior. It is therefore an $(n - 1)$ -sphere containing $(n + 2)$ colors.

Proof of Theorem 2.n.

PROOF OF THEOREM 2.2. Let T be the circle-preserving map in question and let M be a five point set in general position contained in the image of T . Define a partition $\Gamma = \{\Gamma_1, \Gamma_2, \Gamma_3, \Gamma_4, \Gamma_5, \Gamma_6\}$ by

$$\Gamma_i = T^{-1}(m_i) \text{ for } m_i \in M, i = 1, 2, 3, 4, 5,$$

and

$$\Gamma_6 = S^2 - \cup M_i, i = 1, \dots, 5.$$

If Γ_6 is empty we are done, because the first five sets are then a five-coloring. There would be a four-colored circle—i.e. a circle containing points in four of the five colors—so the image of that circle would contain four of the m 's. This violates the condition that M lie in general position. If therefore Γ_6 is not empty, the image of T contains a sixth point m_6 . We have demonstrated in [4] that T must be an inversive transformation.

PROOF OF THEOREM 2.3. Let T be a map of S^3 to itself that preserves 2-spheres, and per the statement of the theorem, suppose the image of T contains a six point set in general position. I will show that T is an inversive transformation.

Let $\Gamma_i = T^{-1}(x'_i), i = 1, \dots, 6$, where $X' = \{x'_1, \dots, x'_6\}$ is a set in general position lying in the image of T . If the union of these sets covers the domain, we have a contradiction because there would be a five-colored sphere. That would imply that five of the six image points lie on a single 2-sphere, the image of the five-colored sphere. Thus we conclude that T has other image points. Let Γ_7 denote the complement of the union of the first six gammas. There must be a 2-sphere, S , that meets five of the seven gammas, and Γ_7 must be one of them. Assume for definiteness that this sphere is $(x_4x_5x_6x_7)$, and that x_3 is also on S . From Theorem 2.2 we know that the restriction of T to this 2-sphere is an inversive transformation, so $T(S)$ is the entire 2-sphere $(x'_4x'_5x'_6x'_7)$. Any 2-sphere through x_1, x_2 , and S (meeting S on a circle) maps to a sphere S' , where $S' \cap \text{Im}(T)$ is a set in general position, so T is an inversive transformation on any such 2-sphere. If M denotes the domain swept out by these spheres, it is easy to show that three-space is covered by the family of 2-spheres intersecting M and containing a point x not in M . T must therefore be an inversive transformation. QED

Proof of Theorem 2.n, for $n > 3$.

PROOF. We proceed by induction. Let T be a sphere-preserving map of the n -sphere and assume that the image of T contains a set of $(n + 3)$ points in general position. If this set comprised the entire image, the coloring with $\Gamma_i = T^{-1}(x'_i), i = 1, \dots, n + 3$, where $X' = \{x'_1, \dots, x'_{(n+3)}\}$ is a set in general position lying in the image of T . From Theorem 1.n, there would be an $(n + 2)$ -colored $(n - 1)$ -sphere in the domain of T . That would imply that $n + 2$ of the $n + 3$ image points lie on an $(n - 1)$ -sphere, contradicting the assumption that the image lies in general position. It follows that there must be an $(n + 4)^{\text{th}}$ image point $x'_{(n+4)}$. As in the previous proof, let $\Gamma_{(n+4)}$ be the complement of the union of the $n + 3$ colors already defined. There must be an $(n - 1)$ -sphere, $S^{(n-1)}$, in the domain colored by $n + 2$ colors, one of which is $\Gamma_{(n+4)}$. From Theorem 1.($n - 1$), T must be an inversive transformation on this sphere, and in particular it must be a bijection. We can then repeat the construction in the previous proof to arrive at the conclusion that T must itself be an inversive transformation.

PROOF OF COROLLARY 2. It remains to prove that we only need to evaluate a sphere-preserving map on a set $M = \{x_i | i = 1, \dots, n + 3\}$ consisting of $n + 3$ points in general position. If $T(M)$ lies in general position, we are done because the image of T then contains this set. Suppose however, that $T(M)$ does not lie in general position, and specifically, that there is an $(n - 1)$ -sphere S whose complement meets $T(M)$ on only one point, $x'_1 = Tx_1$. Consider the $(n - 2)$ -sphere $(x_4x_5 \dots x_{(n+3)})$ and the three $(n - 1)$ -spheres $(x_1x_4x_5 \dots x_{(n+3)})$, $(x_2x_4x_5 \dots x_{(n+3)})$, and $(x_3x_4x_5 \dots x_{(n+3)})$. The point $T(x_2)$ must lie on $T((x_3x_4x_5 \dots x_{(n+3)})) = S$. So T maps both $(x_2x_4x_5 \dots x_{(n+3)})$ and $(x_3x_4x_5 \dots x_{(n+3)})$ into S . However, T is an inversive transformation if and only if T^{-1} exists and is also an inversive transformation. Far from being inversive however, this would imply that T^{-1} is not even sphere-preserving. QED

Conclusion

It remains an open question whether in Theorem 2.n it is sufficient to assume merely that T preserves circles. Specifically,

Query: if T is a circle-preserving transformation of S^n whose image contains a set in general position (i.e. the complement of every $(n - 1)$ -sphere contains at least two points in the image of T), is it necessarily true that T preserves spheres of all dimensions from $2, \dots, (n - 1)$?

References

- [1] P. Erdős et. al., Euclidean Ramsey Theory, 1, Journal of Combinatorial Theory Series A, **14** (1973), pp. 341–353.
- [2] D. Carter and A. Vogt, Collinearity-preserving Functions between Desarguesian Planes. Memoirs of the A.M.S. **27**(235), Sept. 1980.
- [3] J. Gibbons, Finite Partitions of Spheres, Journal of Combinatorial Theory Series A, **33**(2), Sept 1982, pp 220–222.
- [4] J. Gibbons and C. Webb, Circle-Preserving Functions of Spheres, Trans. Am. Math. Soc., **248**(1), Feb 1979, pp. 67–83.