$(n+3)$-Coloring the n-Sphere

Joel Clarke Gibbons

Abstract. We address a combinatorial proposition for the n-sphere and a corresponding 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. Lastly, we demonstrate that the combinatorial theorem is equivalent to the theorem that defines this class of sets.

Mathematics Subject Classification (2000). Primary 51B10.

Keywords. Circle-preserving map, sphere-preserving map, inversive transformation, coloring the n-sphere, coloring n-space.

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. The following is true for all $n \geq 2$.

Theorem 1. Any $(n+3)$-coloring of S^n must at least $(n+2)$-color some $(n-1)$-sphere.

Under the same conditions, there is a nested tower of spheres such that the sphere of dimension k is $(k+3)$-colored, for $k = 2, 3, ..., (n-1)$. In particular, there is a 4-colored circle. This follows trivially. Conversely, Theorem 1 is equivalent to the proposition that an $(n+3)$-coloring must four-color some circle because starting from a four-colored circle one "can for any k build up to a $(k+3)$-coloring of a k-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 Erdös [3]. The logic of this theorem and of Erdös' findings in Euclidean Ramsey Theory are more in the nature of dual than parallel. Nonetheless, it would be fair to characterize Theorem 1 as being a proposition in Geometrical Ramsey Theory. Euclidean Ramsey Theory addresses the following problem. Consider a four-tuple (H, X, m, n), where H is a subgroup of the non-singular affine group of R^n, X is a finite subset of R^n, and m is the cardinality of a coloring of R^n. The four-tuple is said to have the Ramsey property if the orbit of X under H necessarily contains a monochromatic set, for every coloring of cardinality m. Euclidean Ramsey Theory consists of discovering and characterizing such four-tuples. To translate this framework onto the sphere, we simply replace the affine group by the inversive group. Now Theorem 1 can be expressed in analogous terms in the following way.

Let H be a subgroup of the inversive group of S^n, let X be a subset of the sphere, and let m be an integer which is the cardinality of a coloring of the sphere. An m-coloring Γ colors X and every set $h(X)$ in the orbit of X under H, for $h \in H$. Define the incidence, k, of (Γ, H, X, m, n) as the largest cardinality of the derived colorings of the sets $h(X)$, and define the incidence of the four-tuple (H, X, m, n) is the minimum of the incidence of (Γ, H, X, m, n), taking the minimum over all m-colorings Γ. In these terms, Theorem 1 asserts that when H is the entire inversive group, X is any $(n-1)$-sphere in S^n, and $m = n+3$, the incidence $k = n+2$. This result improves the results announced in [5], where it was proved: that for $n = 2$ and $m = 6, 4 = k$.Our main result can however be stated in a seemingly different way, as a proposition in Inversive Geometry. Specifically,

Definition 1.1. a subset B of the n-sphere lies in *general position* if for any $(n-1)$-sphere S, the Complement of S contains at least two points of B.

Obviously, B must contain at least $n+3$ points. If B lies in general position, every set containing B lies in general position.

Definition 1.2. a map T of the n-sphere to itself is *circle-preserving* if for every circle $C, T(C)$ is contained in a circle. If $T(C)$ is a set of three or fewer points, this condition is met 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, $T(S)$ lies in an $(n-1)$-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 consists of the Möbius Transformations and the other kind are the composition of a Möbius Transformation with the reflection of the sphere through the equator. These transformations are also referred to, respectively, as homographies and anti-homographies. 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 (but not uniquely) into a composition of inversions through $(n-1)$-spheres. The Möbius 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. Let M be a subset of S^n in general position, and let $T : S^n \to S^n$ be a sphere-preserving map. If the image of T contains M, T must be an inversive transformation.

The problem of characterizing conditions on a map T that imply that it is an inversive transformation dates from a 1937 paper of C. Caratheodory [2]. He demonstrated that if T is a function defined on an open region U in S^2, into S^2, and if it is $1-1$ and preserves circles in U, then T is the restriction of an inversive map of the 2-sphere. This result was broadened in [1], where it is shown that it is enough to assume that T be $1-1$ on a set consisting of three intersecting circles. Since in both cases, the region U can be the entire sphere, the parallel with Theorem 2 is clear. In [5] this proposition was extended in two ways. First, the result was generalized to all dimensions, and secondly, the counterpart of the set of three rings was replaced, in dimension n, by a finite set of $2n+2$ points in general position.

Every inversive transformation is determined by its values on $n+2$ points in the domain in the sense that if it is given that T in an inversive transformation, it is enough to evaluate T on a set of $n+1$ points on an $(n-1)$-sphere along with an $(n+2)^{\text{nd}}$ point not on that sphere. 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 lies in general position.

Theorem 2 admits the following corollary.

Corollary 1. 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.The proof of the proposition will be deferred until after proving Theorem 2.

Lastly, we prove

Theorem 3. Theorems 1 and 2 are equivalent.

This proposition highlights the role of pure geometry in the combinatorial proposition. The proof of Theorem 1 will occupy most of this paper because the statement in terms of coloring seems to be the more tractable statement to deal with. Theorem 1 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 $n = 3$, the proof generalizes easily to all higher dimensions. Accordingly, we must treat the case $n = 2$ and the case $n = 3$ individually.

2. Preliminaries

Notation: 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 $[xy]$, the plane through x, y, and z by $[xyz]$, and so on.

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 is the minimum necessary.

Proof. Viewing the n-sphere as the extended Euclidean space, let the origin $\mathbf{0}$ use one of the colors and ∞ use another one. Denoting the axes $X_1, X_3, ..., X_n$, apply the third color to $X_1 - \{\mathbf{0}, \infty\}$, 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 $\{(x_1, ..., x_n) \mid \Sigma b_i x_i = 0\}$ 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 $x_2 = ... = x_n = 0$, so $b_1x_1 = 0$. 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 $x_3 = ... = x_n = 0$, so $b_2x_2 = 0$. 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. □

3. Proof of Theorem 1 for the Case $n = 2$

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, $i = 1, ..., 5$. First however,

The following argument hinges on a single construction. Choose points x_i in $\Gamma_i, i = 1, ..., 5$. The circles $(x_1x_2x_3)$ and $(x_1x_4x_5)$ intersect at two points, x_1 and another point x_1'. For some choice of $\{x_1, x_2, x_3, x_4, x_5\}$ it cannot be ruled out a priori that the circles are tangent at x_1, and that no x_1' is available. If however this is the case for that five-tuple, choose x_2' not on $(x_1x_2x_3)$. If there is no such point, $\Gamma_2 \subset (x_1x_2x_3)$. Similarly for x_3 of course, so either x_2' is available or x_3' is, or $\Gamma_2 \cup \Gamma_3 \subset (x_1x_2x_3)$. If either x_2' or x_3' is available, the circle $(x_1x_2'x_3)$ or $(x_1x_2x_3')$, resp., is transverse to $(x_1x_4x_5)$. Lastly, we can repeat this argument for Γ_4 and Γ_5, but then either there are circles $(x_1x_2x_3)$ and $(x_1x_4x_5)$ which are transverse at x_1 or the complement of two such circles is monochromatic, lying in Γ_1. In that case, choose any triple $\{x_2, x_3, x_4\}$. The circle (x_2, x_3, x_4) must also meet Γ_1, and we have found a four-colored circle. Thus if we assume that no circle is four-colored, there must be a point x_1 at which $(x_1x_2x_3)$ and $(x_1x_4x_5)$ are transverse. Now, let T be the Linear Fractional Transformation.

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

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, denoted X, and maps $(x_1x_2x_3)$ to a line through the origin that we may refer to as the v-axis, V. 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. 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 formula we seek is a consequence of a well-known fact:

Lemma 3.1. Let K and L be lines that cross at a point P, and let C be a circle which meets line K, (resp. L) and let x, x' (resp. v, v') denote the distances from P to the points where K (resp. L) meets C. If K (resp. L) is tangent to $C, x = x'$ (resp. $v = v'$). Orient L and M, so x, x', v, and v' are signed. Then $xx' = vv'$.

Proof. This is an exercise in analytic geometry. The formula for the circle gives simultaneous equations for the coordinates of the four points. The conclusion follows from the system of equations. Refer to the following figure 1.

The lemma implies that, in terms of the parameters defined above,

$$h(v|x_4, x_5) = x_4x_5/v. \tag{1}$$

For points v_2 and v_3 on V and x and $k(x)$ on X, which are points on a circle C,

$$k(x|v_2, v_3) = v_2, v_3/x. \tag{2}$$

Furthermore, in terms of the coloring of S^2, since $x_4 \in \Gamma_4$ and $x_5 \in \Gamma_5$, if Γ is a wild 5-coloring, v and $h(v)$ must always be the same color. Equally, if $v_2 \in \Gamma_2$ and $v_3 \in \Gamma_3, x$ and $k(x)$ 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 $V_i = \Gamma_i \cap V, 1 = 1, 2, 3$, and $X_i = \Gamma_i \cap X, i = 1, 4, 5$. □

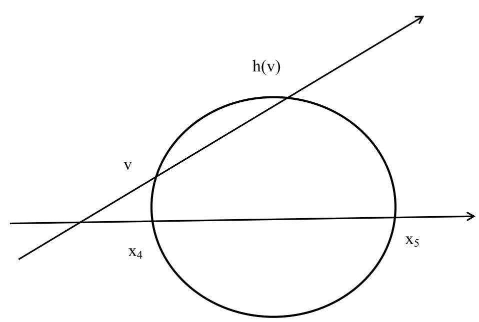

FIGURE 1

Lemma 3.2. Equations 1 and 2 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:

$$m(v|x_5, x_5') = h(h(v|x_4, x_5)|x_4, x_5') = (x_5'/x_5)v \tag{3a}$$

$$m(v|x_4, x_4') = h(h(v|x_4, x_5)|x_4', x_5) = (x_4'/x_4)v \tag{3b}$$

$$n(x|v_2, v_2') = k(k(x|v_2, v_3)|v_2', v_3) = (v_2'/v_2)x \tag{4a}$$

$$n(x|v_3, v_3') = k(k(x|v_2, v_3)|v_2, v_3') = (v_3'/v_3)x. \tag{4b}$$

Since $T(x_5) = 1, 1 \in X_5$.

$m(v|1, g) = gv$ for all $v,$ but from (4a) 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.

(Note that $n(1)$ and 1 have the same color.) Therefore

$1/g = (v_2/gv_2) \in G$. So G is closed under reciprocals.

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

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

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

So the set of 5-colored points on X is a multiplicative group. The preceding calculations imply a simple structure on V_2 and V_3. 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^o.

$k(1|v_2^o, v_3) = v_2^o v_3 = g$. Let $v_3^o = v_3/g$.

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

Furthermore, the reciprocal of every element in V_2 is an element of V_3. The converse is also true of course. Now, from equation 2,

$k(x_4|v_2^o, v_3^o) = 1/x_4 \in X_4,$ so X_4 is closed under reciprocals. From (3b),

$m(v_2^o|x_4, x_4') = (x_4'/x_4)gv_2$.

Thus, $(x_4'/x_4) = ((x_4'/x_4)gv_2)/(gv_2) \in G$. There is an element, x_4^o of X_4 such that $X_4 = \{gx_4^o|g \in G\}$. To summarize, $X_4, V_2,$ and V_3 each have a single generator over G. Now by equation 2,

$k(x_4^o|v_2^o, v_3^o) = 1/x_4^o = gx_4^o,$ so x_4^o 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 $(x_4^o)^2 = g$, which is in $G, (\chi_4)^2 = 1$. Therefore, we can take $x_4^o = -1$, and $X_4 = \{-g|g \in G\}$. Now, from equation 1,

$h(v|x_4^o, 1) = h(v|-1, 1) = -1/v,$ so if $v \in V_i, (-1/v) \in V_i$ for $i = 1$ or 2.

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

$$k(1/g| - v_3, v_3) = v_3^2 g = g'.$$

Therefore, v_3^o is the square root of some number x^o in X_4. Say $v_3^o = \sqrt{x^o}$. (i.e. $x^o = -g'/g$).

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^o and every element of V_3 is a G multiple of $v_3^o = 1/v_2^o$, the product of two numbers in V_i is a g-multiple of x^o, for $i = 1, 2$, and any cross product $v_2 v_3 \in G$. Therefore the product of any two numbers in $V_2 \cup V_3$ is in G^*.

Let H^* be the set of products of numbers in $V_2 \cup V_3$. H^* is a group. It is clearly closed under multiplication and reciprocals, since the reciprocal of a product is the product of the reciprocals. Furthermore, it contains the number 1 because $v_2^o v_3^o = 1$. $H^* v_2^o = V_2 \cup V_3$, so the union is a coset of H^* in the multiplicative reals. We claim that the union is H^*. In the quotient space, denote the coset $\{V_2 \cup V_3\}$ by ϕ. Since $(v_2^o)^2$ is in $H^*, \phi^2 = 1$, so $\{V_2 \cup V_3\} = -H^* = \{-h | h \in H^*\}$. For $v_2 \in V_2$, we know that $-v_2 \in V_3$, and vice versa. Thus, $-H^* = H^*$. Therefore, $H^* = V_2 \cup V_3$. Since H^* is a group however, either $1 \in V_2$ or $1 \in V_3$. 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. Q.e.d.

Note: We could draw the same conclusion in a slightly different way. Since $1 \in G, -1 \in X_4$, it follows that $1 \in V_2$, and $-1 \in V_3$. Thus the unit circle has all four colors on it. The conclusion is in any case that the Wild 5-coloring of the 2-sphere is impossible, because the derived coloring of the lines X and V is impossible.

4. Proof of Theorem 1 for the case $n = 3$

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

Lemma 4.1. Under the assumptions of Theorem 1, for all $n \geq 3$ let $C = (x_i x_j x_k)$ be an arbitrary circle meeting three colors. C cannot have a monochromatic arc.

Proof. First for $n = 3$. Assume the contrary. For definiteness, let $C = (x_1 x_2 x_3)$, and suppose that the monochromatic arc lies in Γ_1. We need to treat two cases. The first case is that C contains at least one other point of Γ_2 or Γ_3. Obviously, the alternative is that x_2 and x_3 are, respectively, the only points of C lying in Γ_2 and Γ_3. We begin with the first case. Assume there is an x_2' in Γ_2. Up to an inversive transformation T we can arrange that $x_2 = \mathbf{0}$, that $x_3 = \infty$, and that C is the vertical axis. Let $(a - b)$ denote a monochromatic interval on C. Finally, select a point x_4, necessarily not lying on C.

The pencil of line through x_4 and through a point x in the interval $(a-b)$ (see Figure 2) covers two connected regions which cannot meet Γ_2 because they meet Γ_1, Γ_3, and Γ_4; i.e. meeting the interval $(a - b), \infty$, and x_4. Similarly, the pencil of circles through the origin, x_4, and a point in the interval $(a - b)$ cannot meet Γ_3. The intersection of these regions therefore lies entirely within $\Gamma_1 \cup \Gamma_4$. The shaded region shown lies within one of these regions. Now, fixing on the shaded region, every line through the origin that crosses this region intersects it on a

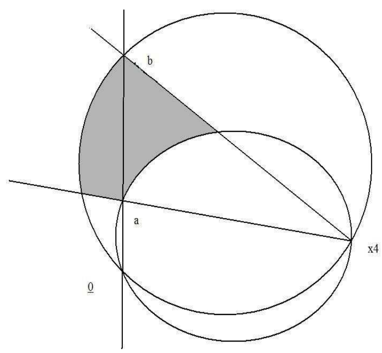

FIGURE 2

monochromatic arc which lies either in Γ_1 or Γ_4. Since the region is connected, there is an x' in the region which is in the closure of both Γ_1 and Γ_4. Consider the line $[x_2 x']$. It is transverse to $[\mathbf{0}x']$ at x', so it must meet both Γ_1 and Γ_4. By construction however it also meets Γ_2 and Γ_3. So either this construction is impossible or we have found a four-colored circle.

Now we can take up the second case, in which—continuing the previous notation—the vertical axis lies entirely in Γ_1 except for the origin and the point at infinity. We see immediately that the origin can be the only Γ_2 point on the plane spanned by C and x_4, because if there was an x_2', we could use it to obtain the same contradiction as before. Similarly, there cannot be any point in Γ_3 on this plane. The entire punctured plane, omitting the origin, lies in $\Gamma_1 \cup \Gamma_4$. Now select x_5. The construction can be repeated on the plane spanned by C and x_5, and on the plane $[\mathbf{0}x_4x_5]$, which in the latter case implies that this plane lies in $\Gamma_4 \cup \Gamma_5$. The latter result implies that the entire punctured three-space, excepting the origin, lies in $\Gamma_1 \cup \Gamma_4 \cup \Gamma_5$. This is a contradiction however since it implies that Γ_6 is empty. The Lemma follows for $n = 3$.

For general $n > 3$ the same cases obtain. The proof in the first case is unchanged because it relies on a contradiction that occurs with the extension of C to a plane containing C. The proof of the second case requires some further comment, but the argument is essentially unchanged. In n dimensions, the coloring has $(n+3)$ colors. Choose a sequence $\{x_4, x_5, ..., x_{n+2}\}$ as before, That is, x_4 is any point of color 4, the plane spanned by C and x_4 is four-colored. Then x_5 lies off this plane, and the three-space spanned by C, x_4, and x_5 is five-colored. So x_6 lies off this space, and the four dimensional space spanned by C, x_4, x_5, and x_6 is six-colored. It follows that the S^n would be $(n+2)$-colored, which is a contradiction. □

Corollary 2. Given circle C and points x and x' on C, $C - \{x, x'\}$ consists of two arcs. If these arcs meets colors i and j, different from the color of point x, we can choose points x_i and x_j on C to lie in opposite arcs.

Proof. Choose any x_i in one of the arcs. If the opposite arc does not meet Γ_j, Γ_i must be dense in it, because it is bi-chromatic. So either the opposite arc contains many j-colored points or it contains many i-colored points. If it is the latter case, we simply choose x_j lying in the first arc. □

Proof. Proof of Theorem 1. For $n = 3$, suppose the contrary, which is that no two sphere in S^3 is 5-colored. Select a circle $(x_1 x_2 x_3)$. Up to an inversive transformation, we can take it to be a line through the origin, with $x_1 = \mathbf{0}$. The corollary guarantees that if x_2 lies on one side of the origin, there must be an x_3 on the other side. Now consider plane $(\mathbf{0} x_4 x_5 x_6)$ through points x_4, x_5, and x_6. If it also contains the line $(\mathbf{0} x_2 x_3)$, it meets all six colors and we are done. Actually, the plane cannot contain any points of Γ_2 or Γ_3; it must lie 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 $(0 x_2 x_3)$ lying above the plane contains a point in Γ_2 and the ray below the plane contains a point of Γ_3 (or vice versa). The plane contains points x_4, x_5, and x_6, and corresponding rays $[\mathbf{0} x_4], [\mathbf{0} x_5]$, and $[\mathbf{0} x_6]$. Note that the ray through $x_j, j = 4, 5, 6$, must be bi-chromatic because the two lines through the origin, $(0, x_2, x_3)$ and $[\mathbf{0} x_i]$ span a plane (a 2-sphere) which by assumption can only meet four colors: colors $1, 2, 3$, and i.

Let $(x_2 x_3 x_4 x_5)$ 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 $[\mathbf{0} x_6]$. It is therefore a five-colored sphere, since either a Γ_1 or a Γ_6 must also meet $(x_2 x_3 x_4 x_5)$.

For $n > 3$, the construction simply duplicates the construction for $n = 3$. The only thing we need to verify is that the rays $[\mathbf{0} x_j], j = 4, ..., (n+3)$, must be bi-chromatic. Any such ray together with the line called $(x_1 x_2 x_3)$ 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 §3. 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 either the rays are bi-chromatic or a contradiction ensues.

Now consider the $(n-1)$-sphere containing $(x_2 x_3 ... x_{(n+2)})$. It must meet the ray $[\mathbf{0} x_{(n+3)}]$ since the origin, $\mathbf{0}$, lies in the interior. It is therefore an $(n-1)$-sphere containing $(n+2)$ colors. □

5. Proof of Theorem 2

Proof of Theorem 2 for the case $n = 2$: 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

$$\Gamma_i = T^{-1}(m_i) \text{ for } m_i \in M, i = 1, 2, 3, 4, 5, \text{ and}$$
$$\Gamma_6 = S^2 - \cup M_i.$$

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 [5] that T must be an inversive transformation.

Proof. Proof of Theorem 2 for the case $n = 3$. Let T be a sphere-preserving map of S^3 to itself, and per the statement of the theorem, suppose the image of T contains a six point set in general position. Let $\Gamma_i = T^{-1}(x'_i), i = 1, ..., 6$, where $X' = \{x'_1, ..., x'_6\}$ 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 ($n = 2$) 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. □

For general n, 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 $\Gamma_i = T^{-1}(x'_i), i = 1, ..., n+3$, where $X' = \{x'_1, ..., x'_{n+3}\}$ is a set in general position lying in the image of T. From Theorem 1, 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, 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 the Corollary which states that the sphere-preserving transformation T is determined on a set of $(n+3)$ points: it remains to prove that we only need to evaluate a sphere-preserving map on a set $M = \{x_i | i = 1, ..., n+3\}$ 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 $(x_4 x_5 ... x_{n+3})$ and the three $(n-1)$-spheres $(x_1 x_4 x_5 ... x_{n+3})$, $(x_2 x_4 x_5 x_{n+3})$, and $(x_3 x_4 x_5 ... x_{n+3})$. The point $T(x_2)$ must lie on $T((x_3 x_4 x_5 ... x_{n+3})) = S$. So T maps both $(x_2 x_4 x_5 ... x_{(n+3)})$ and $(x_3 x_4 x_5 ... x_{(n+3)})$ 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 even if T^{-1} exists, it is not even sphere-preserving. □

Up to this point we have proved that Theorem 1 implies Theorem 2. This proposition and its converse we state as theorem 3.

Theorem 4. Theorems 1 and 2 are equivalent.

Proof. All that is needed here is to demonstrate that Theorem 2 implies Theorem 1 since we deduced Theorem 2 from Theorem 1 already. Let Γ be an $(n+3)$-coloring of S^n. Select a set $\{x_1, ..., x_{n+3}\} \subset S^n$ in general position and define a function $T : S^n \to S^n$ by $T(\Gamma_i) = x_i$. If the coloring has no $(n+2)$-colored $(n-1)$-sphere, T is sphere-preserving trivially because the image of an $(n-1)$-sphere lies in a set that must lie in an $(n-1)$-sphere. Since such a function T is not an inversive transformation however—its image lies in a discrete set—a contradiction of Theorem 2 ensues. □

6. Coloring R^n

The general case of coloring n-space seems not to have been covered in the literature. The theorem corresponding to Theorem 1 is

Theorem 5. Given an $(n+2)$-coloring of R^n, some $(n-1)$-hyperplane is $(n+1)$-colored.

Proof. The proof is very easy, given Theorem 1. For $n = 2$, there are points x_1, x_2, x_3, and x_4 lying in four different colors. If three of them are collinear, we are done. If not, they form a quadrilateral. Assume that the vertices are numbered clockwise, so $[x_1 x_3]$ and $[x_2 x_4]$ are diagonals. They meet at a point x which lies in one of the four colors, and which therefore 3-colors one of the diagonals.

For $n > 2$, map R^n into S^n by stereographic projection, and color the North Pole by an $(n+3)^{\text{rd}}$ color. There is an $(n+2)$-colored $(n-1)$-sphere somewhere. If it passes through the North Pole, it projects in n-space to an $(n+1)$-colored hyperplane. If it does not, it projects to an $(n+2)$-colored $(n-1)$-sphere, $S^{(n-1)}$, in R^n. This sphere contains an $(n+1)$-colored $(n-2)$-sphere which is the locus of intersection of $S^{(n-1)}$ with a hyperplane. The hyperplane therefore meets these $(n+1)$ colors. □

The equivalents of Theorems 2 and 3 also follow trivially.

7. Conclusion

It remains an open question whether in Theorem 2 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 $2, ..., (n-1)$?

References

[1] J.D. Aczel and M.A. McKiernan, On the Characterization of Plane Projective and Complex Möbius Transformations, *Math. Nach.* **33** (1967), pp. 315–337.

[2] C. Caratheodory, The Most General Transformation of Plane Regions Which Transform Circles into Circles. *Bull. Am. Math. Soc.* **43** (1937), pp. 573–579.

[3] P. Erdös et. al., Euclidean Ramsey Theory, 1, *Journal of Combinatorial Theory Series A*, **14** (1973), pp. 341–353.

[4] J. Gibbons, Finite Partitions of Spheres, *Journal of Combinatorial Theory Series A*, **33**(2), Sept 1982, pp. 220–222.

[5] J. Gibbons and C. Webb, Circle-Preserving Functions of Spheres, *Trans. Am. Math. Soc.*, **24**(1), Feb 1979, pp. 67–83.

This paper has been submitted to the AMS regional meeting at Cornell University, Ithaca, NY. Meeting number 1072, submission #1072-05-5. Meeting date: September 10–11, 2011.

Joel Clarke Gibbons
Logistic Research & Trading Co.
4052 Niles Road
Saint Joseph, Michigan, 49085
e-mail: jgibbons@logisticresearch.com