

Classifying $\overline{\mathbb{C}}$ Automorphisms and the Hyperbolic Plane

Rohan Mukherjee

August 15, 2023

Abstract

In this paper we look at the Riemann Sphere, it's incredible equivalent forms, where they each aid intuition, classify it's automorphisms. From here we discuss the hyperbolic plane \mathbb{H}^2 , and classify a large class of it's self-homeomorphisms.

1 Arithmetic in $\overline{\mathbb{C}}$

Before we look at the equivalent structures of the extended complex plane, we will first look at arithmetic on it. We define $\overline{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$. Define the following:

$$z + \infty = \infty \text{ for all } z \in \mathbb{C}$$

$$z \times \infty = \infty \text{ for all } z \in \mathbb{C}$$

$$\frac{z}{0} = \infty \text{ and } \frac{z}{\infty} = 0 \text{ for all } z \in \mathbb{C} \setminus 0$$

with $\infty \times \infty = \infty$, while $\infty - \infty$ and $0 \times \infty$ are left undefined. Similarly, let $\infty/0 = \infty$, $0/\infty = 0$, and leave ∞/∞ and $0/0$ undefined. With these definitions $\overline{\mathbb{C}}$ does not define a field, as ∞ does not have an additive inverse.

2 Topology

Now, we must define a topology on $\overline{\mathbb{C}}$. Being the 1-point compactification of \mathbb{C} . First,

Definition 1. A space X is said to be locally compact at x if there exists a compact subspace C of X that contains a neighborhood of x . If X is locally compact at each of its points, X is said to be locally compact.

We guide the reader to section 29 of Munkres for the proof of the following theorem:

Theorem 1. Let X be a space. Then X is locally compact Hausdorff if and only if there exists a space Y satisfying the following conditions:

1. X is a subspace of Y .
2. The set $Y - X$ consists of a single point.
3. Y is a compact Hausdorff space.

If Y and Y' are two spaces satisfying these conditions, then there is a homeomorphism of Y with Y' that equals the identity map on X .

The topology Y is constructed as follows:

$$\mathcal{T} = \mathcal{T}_X \cup \{Y - C \mid C \subset X \text{ compact} \}$$

For example, $\overline{\mathbb{C}} \setminus B_1(0)$ is an open subset of $\overline{\mathbb{C}}$. Next, we prove exercise 5 in the aforementioned section of Munkres, that is

Theorem 2. If $f : X_1 \rightarrow X_2$ is a homeomorphism of locally compact Hausdorff spaces, then f extends to a homeomorphism of their one-point compactifications.

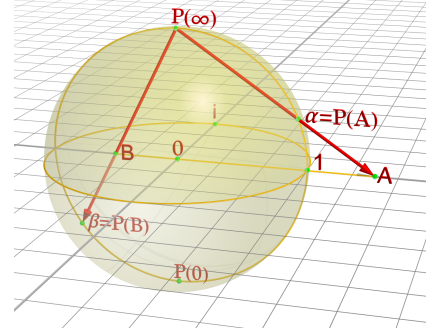
Proof. Let $Y_1 = X_1 \cup \{p\}$, $Y_2 = X_2 \cup \{q\}$ denote the one-point compactification of X_1 , X_2 respectively. Define $\tilde{f}(x) = f(x)$ for $x \in X_1$, and $\tilde{f}(p) = q$. For any open set $U \subset Y_2$, if $q \notin U$ then $\tilde{f}^{-1}(U) = f^{-1}(U)$ which is an open subset of X_1 , and by how we defined the topology on Y_1 , it too is an open subset of Y_1 . If q is in U , then $Y_2 \setminus U = C \subset X_2$ is compact. Thus $\tilde{f}^{-1}(C) = f^{-1}(C)$ is compact (since f is a homeomorphism, its inverse is continuous). Thus $\tilde{f}^{-1}(U) = Y_1 \setminus \tilde{f}^{-1}(C)$ is open in Y_1 , since it is the complement of a compact set in X_1 . By running the exact same argument for the inverse map, we have proven the theorem. \square

3 Equivalent Forms of $\overline{\mathbb{C}}$

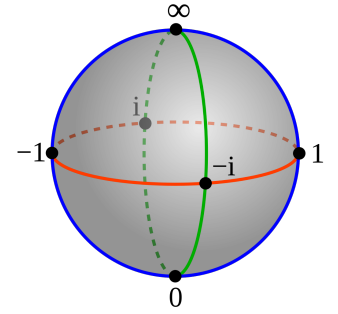
We have seen the homeomorphism from $\mathbb{S}^1 \setminus \{N\}$ to \mathbb{R} previously on homework 4. The incredibly idea of stereographic projection works here as well. Define $f : \mathbb{S}^2 \setminus (0, 0, 1) \rightarrow \mathbb{C}$ by

$$f(x, y, z) = \frac{x + iy}{1 - z}$$

Thought of as a map from $\mathbb{R}^3 \rightarrow \mathbb{R}^2$, i.e. having $f(x, y, z) = (x/(1 - z), y/(1 - z))$, we see that each coordinate is a rational function in x, y, z , whose denominator is never 0, and thus is continuous. An equivalent formulation for this map is that if $(\theta, \varphi) \in \mathbb{S}^2$ is the spherical coordinates for a point on the sphere, then its image will be $\cot(\frac{1}{2}\theta)e^{i\varphi}$. From here, since $\cot(\frac{1}{2}\theta)$ is surjective to $[0, \infty)$ for $\theta \in [0, \pi]$, any point $re^{i\alpha}$ will be the image of some point on the \mathbb{S}^2 . Similarly, since $\cot(\frac{1}{2}\theta)$ is one-to-one and strictly positive for $\theta \in [0, \pi]$, if $\cot(\frac{1}{2}\theta_1)e^{i\varphi_1} = \cot(\frac{1}{2}\theta_2)e^{i\varphi_2}$, taking absolute values gives that $\theta_1 = \theta_2$, and canceling the \cot from each side (unless $\theta_1 = \theta_2 = \pi$, in which case the point in question is the south pole), yields that $\varphi_1 = \varphi_2$, so we have guaranteed injectivity. This map also happens to have continuous inverse. Now, \mathbb{S}^2 is a compact Hausdorff space with $\mathbb{S}^2 \setminus \{(0, 0, 1)\} \subset \mathbb{S}^2$ and also $\overline{\mathbb{S}^2 \setminus \{(0, 0, 1)\}} = \mathbb{S}^2$, thus \mathbb{S}^2 is the one-point-compactification of $\mathbb{S}^2 \setminus \{(0, 0, 1)\}$. As we defined $\overline{\mathbb{C}}$ to be the one-point-compactification of \mathbb{C} , using the previous theorem lets us extend this homeomorphism to a homeomorphism of the sphere with the extended complex plane. For this reason the extended complex plane is sometimes called the Riemann sphere.



This form is especially important for visualizing the extended complex plane. For example, since \mathbb{S}^2 is sequentially compact, every sequence has a convergent subsequence. We propose a question to the reader: what is a convergent subsequence of $(n)_{n \in \mathbb{N}} \subset \overline{\mathbb{C}}$? Itself! $(n)_{n \in \mathbb{N}} \rightarrow \infty$, which may be visualized as wrapping along the right side of the Riemann sphere.



Now we will get to the form that will help us the most during this paper. We have seen real projective space \mathbb{RP} before. It turns out that the extended complex plane is also homeomorphic to the complex projective space. We give the following definition:

Definition 2. We define $\mathbb{CP} = \left(\mathbb{C}^2 \setminus (0, 0) \right) / \sim$ where $(z, w) \sim (\lambda z, \lambda w)$ for every $\lambda \in \mathbb{C} \setminus 0$.

The topology on this space is of course the quotient topology. Showing that this is homeomorphic to $\overline{\mathbb{C}}$ goes as follows: Define the map $q : \mathbb{C}^2 \setminus \{0\} \rightarrow \overline{\mathbb{C}}$ by $q(z, v) = z/v$. We refer the reader to the MSE post in the bibliography to see the proof that this map is open and continuous (<https://math.stackexchange.com/questions/410929/why-is-the-complex-projective-line-homeomorphic-to-a-2-sphere>). We must now show that this map is constant on the equivalence classes. $[z, w] = \{(\lambda z, \lambda w) \mid \lambda \in \mathbb{C}\}$, thus this claim follows immediately since $\lambda z / \lambda w = z/w$. We thus have the following commutative diagram:

$$\begin{array}{ccc} \mathbb{C}^2 - (0, 0) & & \\ \downarrow \pi & \searrow q & \\ \mathbb{CP} & \xrightarrow{\tilde{q}} & \overline{\mathbb{C}} \end{array}$$

where \tilde{q} is a quotient map since we just verified that q is. We need to check injectivity. If nothing is 0, $z_1/w_1 = z_2/w_2$, then $z_1/z_2 = w_1/w_2$. This says precisely that $(z_1, w_1) \in [z_2, w_2]$, hence we are done. if $w_1 = 0$, then w_2 is necessarily 0 (note that $z_2 \in \mathbb{C}$, not $\overline{\mathbb{C}}$). Thus they are in the same equivalence class. If $z_1 = 0$, then $z_2 = 0$ by the same reasoning. Symmetry covers the rest of the cases, thus we have checked injectivity. A bijective quotient map is a homeomorphism, thus we are done. This form of the extended complex plane will be especially helpful—via this homeomorphism, any rational function translates immediately into a continuous function on the complex projective plane by $f/g \leftrightarrow [f(x), g(x)]$.

4 Automorphisms of $\overline{\mathbb{C}}$

For all our purposes, an automorphism from $\overline{\mathbb{C}}$ to $\overline{\mathbb{C}}$ is a bijective holomorphic function (this automatically guarantees holomorphic inverse). Holomorphic in this case just means C^∞ , we refer the reader to Stein & Shakarchi Book 2 for reference. We introduce the holy grail of all holomorphic maps, so important that they are given a special name: conformal mappings AKA Möbius transformations, maps of the form

$$f(z) = \frac{az + b}{cz + d}, \quad a, b, c, d \in \mathbb{C} \text{ with } ad - bc \neq 0$$

The condition $ad - bc \neq 0$ should be familiar to the reader—it is equivalent to the determinant of the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ having nonzero determinant. Recall from the previous section that the rational function f can be expressed as $[az + b, cz + d]$ in projective coordinates, which may be rewritten as $[z, 1] \cdot \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Now the condition shouldn't come as such a surprise! Long, tedious calculations show that composing Möbius transformations is the exact same as composing these matrices—in particular, the composition of two Möbius transformations is another Möbius transformation. Also, given $ad - bc \neq 0$, the inverse Möbius transformation would be

$$f^{-1}(z) = \frac{dz - b}{-cz + a}$$

which should not come as a surprise. One can extend a function $f : \mathbb{C} \rightarrow \mathbb{C}$ to include ∞ in its domain by defining $f(\infty) = \lim_{z \rightarrow \infty} f(z)$ when the limit exists. f is necessarily continuous at ∞ as given any $w_n \rightarrow \infty$, $\lim_{n \rightarrow \infty} f(w_n) = \lim_{z \rightarrow \infty} f(z) = f(\infty)$ (We may use this equivalent form of continuity since $\overline{\mathbb{C}}$ is first-countable, being homeomorphic to \mathbb{S}^2 .) f is said to be meromorphic at ∞ if $f(1/z)$ is meromorphic in 0. We may also include ∞ in the range: if $f(z)^{-1}$ has a zero at $a \in \overline{\mathbb{C}}$, then we say that $f(a) = \infty$. One calls f meromorphic if it is meromorphic at each $a \in \overline{\mathbb{C}}$. I claim that meromorphic implies continuity. Indeed, letting $a \in \overline{\mathbb{C}}$ be arbitrary, if $a \neq \infty$, and $f(a) \neq \infty$, then $f(a)$ is analytic (as a complex function) and hence obviously continuous. If $f(a) = \infty$, $f(z)^{-1}$ has a 0 at a , and is analytic there (by definition of a pole), so is continuous. Finally, for any $M \in \mathbb{N}$, we need to find a neighborhood $U = \overline{\mathbb{C}} \setminus K$ (K compact) of ∞ so that $|f(z)| \geq M$ for every $z \in U$. This follows directly from the definition of $f(\infty)$. Now, we note the following incredibly surprising fact:

Theorem 3. *If $f : \overline{\mathbb{C}} \rightarrow \overline{\mathbb{C}}$ is an automorphism (bijective meromorphic function), then f is a Möbius transformation of the form $\frac{az+b}{cz+d}$ with $ad - bc \neq 0$.*

This theorem is laughably false for real-valued functions. For example, x^n for any odd n is an automorphism of \mathbb{R} . The proof is very long and very hard, so we will just prove this lemma:

Lemma 1. *If $f : \mathbb{C} \rightarrow \mathbb{C}$ is holomorphic everywhere, and injective, then it is of the form $f(z) = az + b$ with $a \neq 0$.*

This lemma depends on something called the Casorati-Weierstrass theorem and the open mapping theorem, which we will use but not provide here.

Proof. We write $f(z) = \sum_{n=0}^{\infty} a_n z^n$. We have two cases: the sum has finitely many terms, or it doesn't. In the second case, since $f(1/z)$ is holomorphic in $\mathbb{C} \setminus 0$, and has an essential singularity in 0, $f(B_1(0) \setminus 0)$ is dense in \mathbb{C} . Consider any point $w \in f(\mathbb{C} \setminus \overline{B_1(0)})$. Find $\varepsilon > 0$ so that $B_\varepsilon(w) \subset f(\mathbb{C} \setminus \overline{B_1(0)})$, which we can do since f is an open map by the open mapping theorem. By density, find a $z \in B_1(0) \setminus 0$ so that $f(z) \in B_\varepsilon(w)$. But since $B_1(0)$ and $\mathbb{C} \setminus \overline{B_1(0)}$ are disjoint open sets, their image is also disjoint and open (by injectivity), which this immediately contradicts. So $f(z) = \sum_{n=0}^m a_n z^n$. Suppose that there was some $a_n \neq 0$ for some $n > 1$. Then f would be a degree m polynomial with $m > 1$. Thus factor $f(z) = \prod_{i=0}^m (z_i - \alpha_i)$. If α_i is always the same, $f(z) = 1$ would have $m > 1$ solutions, a contradiction. Else, f has at least two distinct roots, another contradiction. Thus $f(z) = az + b$. Constants are not injective, so $a \neq 0$, and we are done. \square

We have seen the geometric side of things. Now, let's look at the algebraic side of things.

5 Algebra

Now define

$$\text{Aut}(\overline{\mathbb{C}}) = \left\{ \frac{az+b}{cz+d} \mid a, b, c, d \in \mathbb{C}, ad - bc \neq 0 \right\}$$

Let $\text{GL}_2(\mathbb{C})$ denote the group of 2×2 matrices who have nonzero determinant. A quick calculation shows that matrix multiplication and composition of mobius transformations are precisely the same, so we can define a group homomorphism

$$\varphi : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \frac{az+b}{cz+d}$$

from $\text{GL}_2(\mathbb{C})$ to $\text{Aut}(\overline{\mathbb{C}})$. The identity automorphism is of course $f(z) = z$. Thus $\ker(\varphi)$ consists of the matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ so that $\frac{az+b}{cz+d} = z$ for all z , and it is not hard to see that this

happens iff $a = c \neq 0, c = d = 0$. Thus the kernel is precisely

$$\ker(\varphi) = \left\{ \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \mid \lambda \in \mathbb{C} - 0 \right\}$$

We define $\text{PGL}_2(\mathbb{C})$ as the quotient $\text{GL}_2(\mathbb{C}) / \ker(\varphi)$ (relating to a group of matrices acting on \mathbb{CP} , which is also the extended complex plane!), and by the above theorem this map is onto, hence

Theorem 4. $\text{Aut}(\overline{\mathbb{C}}) \cong \text{GL}_2(\mathbb{C}) / \ker(\varphi) = \text{PGL}_2(\mathbb{C})$.

. By the above we also now have a good intuition for the a group acting on $\mathbb{CP} \cong \overline{\mathbb{C}} \cong \mathbb{S}^2$. So, what does this mean? matrices tell you how you stretch and rotate the unit square. By modding out by $\ker(\varphi)$, we may multiply by the matrix $\begin{pmatrix} \sqrt{1/\det(M)} & 0 \\ 0 & 1/\sqrt{\det(M)} \end{pmatrix}$, who's determinant is $1/\det(M)$, and represent each element of this quotient group by a matrix with determinant 1. Geometrically this means we are no longer stretching, just rotating. That is indeed the next result:

Theorem 5. Recall that $\text{SL}_2(\mathbb{C}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mid a, b, c, d \in \mathbb{C}, ad - bc = 1 \right\}$. Write $\text{PSL}_2(\mathbb{C}) = \pi(\text{SL}_2(\mathbb{C}))$, where π is the projection from $\text{GL}_2(\mathbb{C})$ to $\text{GL}_2(\mathbb{C}) / \ker(\varphi)$. Then $\text{Aut}(\overline{\mathbb{C}}) \cong \text{PGL}_2(\mathbb{C}) = \text{PSL}_2(\mathbb{C})$.

We define

Definition 3. A circle in the sphere $\mathbb{S}^2 \subset \mathbb{R}^3$ is the intersection $\mathbb{S}^2 \cap \Pi$ where Π is a plane in \mathbb{R}^3 that is intersecting and not tangent to \mathbb{S}^2 . A circle in $\overline{\mathbb{C}}$ is thus the image of this set under the prescribed homeomorphism.

A long calculation shows that if $\alpha x_1 + \beta x_2 + \gamma x_3 = \delta$ ($\alpha, \beta, \gamma \in \mathbb{R}$) is a plane corresponding to the circle C in \mathbb{S}^2 , then C is given by the equation $2\alpha \text{Re}(z) + 2\beta \text{Im}(z) + \gamma(|z|^2 - 1) = \delta(|z|^2 + 1)$. Referencing the book, this leads to there being two types of circles: regular circles in \mathbb{C} ,

Theorem 6. *If C is a circle of $\overline{\mathbb{C}}$, and if $T \in \text{PGL}_2(\mathbb{C})$, then $T(C)$ is a circle of $\overline{\mathbb{C}}$.*

Lemma 2. *Every mobius transformation is the composition of*

- (i) *Transformations of the form $e^{i\theta} z$ ($\theta \in \mathbb{R}$), representing a rotation of the sphere \mathbb{S}^2 by an angle θ about the vertical axis,*
- (ii) *The transformation $z \mapsto z^{-1}$, representing a horizontal rotation by π ,*
- (iii) *The transformations $z \mapsto rz$, $r \in \mathbb{R}_{>0}$ (stretching),*
- (iv) *The transformations $z \mapsto z + t$, ($t \in \mathbb{C}$) (translations).*

$$\begin{array}{ccccccc} & & & & T & & \\ \overline{\mathbb{C}} & \xrightarrow{z \mapsto e^{i\theta} z} & \overline{\mathbb{C}} & \xrightarrow{z \mapsto rz} & \overline{\mathbb{C}} & \xrightarrow{z \mapsto z+t} & \overline{\mathbb{C}} \\ & & & & & & \end{array}$$

Proof (Of theorem). For euclidean circles:

$$\begin{aligned} (x \cos(\theta) - y \sin(\theta) - a)^2 + (x \sin(\theta) + y \cos(\theta) - b)^2 &= R^2 \iff \\ x^2 - x2(a \cos(\theta) + b \sin(\theta)) + b^2 + y^2 + y2(a \sin(\theta) - b \cos(\theta)) + a^2 &= R^2 \end{aligned}$$

8

we do not need to add anything to complete the square, and thus our circle is

$$(x - (a \cos(\theta) + b \sin(\theta)))^2 + (y + (a \sin(\theta) - b \cos(\theta)))^2 = R^2$$

This is of course what we expected, as the center (a, b) under the aforementioned transformation gets mapped there. Similarly,

(ii) For $z \mapsto rz$, we would get

$$\begin{aligned} (rx - a)^2 + (ry - b)^2 &= R^2 \iff \\ (x - a/r)^2 + (y - b/r)^2 &= (R/r)^2 \end{aligned}$$

(iii) The final translation $z \mapsto z + (c + id)$ is clear and done in a similar fashion.

For circles of the form $\Lambda \cup \infty$ where Λ is a straight line in \mathbb{C} , the default equation is $y = mx + b$. Thus,

(i) The transformation $x + iy \mapsto x \cos(\theta) - y \sin(\theta) + i(x \sin(\theta) + y \cos(\theta))$ gives

$$\begin{aligned} x \sin(\theta) + y \cos(\theta) &= m(x \cos(\theta) - y \sin(\theta)) + b \\ x(\sin(\theta) - m \cos(\theta)) + y(\cos(\theta) + m \sin(\theta)) &= b \end{aligned}$$

Which is clearly seen to be a line. Next,

(ii) The transformation $x + iy \mapsto rx + iry$ yields

$$ry = mrx + b \iff y = mx + b/r$$

Again clearly a line. Finally,

(iii) The translation $x + iy \mapsto x + iy + (c + id)$ is clear like last time.

By the lemma every Mobius transformation is a composition of (at most three) of these. Since each of these maps circles to circles, the theorem has been proved. \square

Finally,

Theorem 7. *The map $z \mapsto \bar{z}$ is a circle-preserving homeomorphism.*

This theorem can be proved in the same exact way as before, and is therefore left as an exercise to the reader.

6 The reverse implication

Theorem 8. *If f is a circle preserving homeomorphism of $\overline{\mathbb{C}}$, then f is a Möbius transformation or the composition of complex conjugate with a Möbius transformation.*

Now this is quite the spectacular theorem. This gives rise to a very large class of homeomorphisms of $\overline{\mathbb{C}}$, which is topologically important, and has huge applications to hyperbolic geometry, which we will not discuss. First,

Lemma 3. *If $X \subset \overline{\mathbb{C}}$ is dense, and $f : \overline{\mathbb{C}} \rightarrow \overline{\mathbb{C}}$ is a continuous function so that $f(z) = z$ for all $z \in X$, then f is the identity.*

Proof. Sequential continuity is equivalent to regular continuity for first-countable spaces. It follows then that for any $z \in \overline{\mathbb{C}}$, there is some $(z_n) \rightarrow z$ with each $z_n \in X$. Now,

$$\begin{array}{ccc} f(z_n) & \longrightarrow & f(z) \\ \parallel & & \parallel \\ z_n & \longrightarrow & z \end{array}$$

□

Theorem 9. *If $m(z)$ fixes three distinct points of $\overline{\mathbb{C}}$, then $m(z)$ is the identity.*

Proof. If $m(\infty) = \infty$, then $c = 0$. This yields $m(z) = az + b$ for some a, b . Fixing two points means that the linear polynomial $(a - 1)z + b = 0$ has two solutions, which can only happen if $a = 1$ and $b = 0$. Similarly, if $c \neq 0$, then $m(\infty) \neq \infty$. So all three fixed points lie in \mathbb{C} . Notice that

$$\frac{az + b}{cz + d} = z \iff az + b = cz^2 + dz \iff cz^2 + z(d - a) - b = 0$$

is a quadratic. To have three roots, this quadratic must be equivalently 0, that is, $c = 0$, $d = a$, and $b = 0$, which is again the identity. □

Theorem 10. *There is a unique Mobius transformation sending (z_1, z_2, z_3) to (w_1, w_2, w_3) .*

Proof. (Uniqueness) Let m and n be Mobius transformations doing the above. Then $n^{-1}m$ is a Mobius transformation fixing three points and hence is the identity.

(Existence) When all the z_1, z_2, z_3 lie in \mathbb{C} , the map

$$m(z) = \frac{z - z_1}{z - z_3} \frac{z_2 - z_3}{z_2 - z_1} = \frac{(z_2 - z_3)z - z_1(z_2 - z_3)}{(z_2 - z_1)z - z_3(z_2 - z_1)}$$

sends (z_1, z_2, z_3) to $(0, 1, \infty)$ by plugging things in, and is Mobius since $ad - bc = (z_2 - z_3)(-z_3)(z_2 - z_1) - (-z_1)(z_2 - z_3)(z_2 - z_1) = (z_2 - z_3)(z_1 - z_3)(z_2 - z_1)$ which is nonzero since the z_i 's are distinct. For the map sending (∞, z_2, z_3) to $(0, 1, \infty)$, the function

$$m(z) = \frac{z_2 - z_3}{z - z_3}$$

works. The case $\infty \mapsto \infty$ results in picking $c = 0$, and is so simple I leave it as an exercise to the reader. Now, let m be the map sending (z_1, z_2, z_3) to $(0, 1, \infty)$, and let n be the map sending (w_1, w_2, w_3) to $(0, 1, \infty)$. The map $n^{-1}m$ is the desired map, and the theorem has been proved. \square

I reference the book in that three distinct points in $\overline{\mathbb{C}}$ uniquely identify a circle (or a line, if one of the points is ∞).

Proof (Of theorem). We now give the proof of the aforementioned theorem in all it's glory. We have already seen that all of those maps are circle-preserving homeomorphisms. We must prove the other direction. Let f be a circle-preserving homeomorphism, and let p be the Mobius transformation sending $(f(0), f(1), f(\infty))$ to $(0, 1, \infty)$, so $p \circ f$ satisfies $p \circ f(0) = 0$, $p \circ f(1) = 1$, and $p \circ f(\infty) = \infty$. $\overline{\mathbb{R}}$ (the extended real line) is the unique circle determined by $(0, 1, \infty)$, and thus $f(\overline{\mathbb{R}}) = \overline{\mathbb{R}}$.

I claim that $p \circ f(\mathbb{H})$ is either the upper-half or the lower-half plane. Notice that nothing in this image can lie on the extended real line by bijectivity. Case 1: there is some $z \in \mathbb{H}$ so that $p \circ f(z) \in \mathbb{H}$. Now suppose that there was some $w \in \mathbb{H}$ so that $p \circ f(w)$ had negative real part. Letting Λ be the straight line connecting z to w , Λ is connected while $(p \circ f(\Lambda) \cap \mathbb{H}, p \circ f(\Lambda) \cap -\mathbb{H})$ forms a disconnection of the image, a contradiction. Suppose, with the previous conditions still holding, there was some $a \in -\mathbb{H}$ so that $p \circ f(a)$ had positive real part. The segment connecting $f(z)$ to $f(a)$ is connected but it's preimage is not, which contradicts $p \circ f$ being a homeomorphism. This shows that $f(\mathbb{H}) = \overline{\mathbb{H}}$ in this case.

The other case is extremely similar and is left as an exercise to the reader, so the only other possibility is that $f(\overline{\mathbb{H}}) = -\mathbb{H}$.

In the first case set $m = p$, and in the second set $m = C \circ p$, where C is complex conjugation. This yields a map preserving the upper half plane, and sending $(0, 1, \infty)$ to itself. Let Z denote the set of fixed points of $m \circ f$. Notice that $\{0, 1, \infty\} \subset Z$. Since $m \circ f$ fixes ∞ , if $m \circ f$ sends Euclidean lines in \mathbb{C} (regular lines) to Euclidean lines in \mathbb{C} , and Euclidean circles in \mathbb{C} to Euclidean circles in \mathbb{C} . I claim that if X, Y are two Euclidean circles in \mathbb{C} that intersect at z_0 , with $m \circ f(X) = X$ and $m \circ f(Y) = Y$, then $m \circ f(z_0) = z_0$. This follows since $f(z_0) \in X \cap Y$, which has a unique intersection point, so z_0 is in Z in this case.

For each $s \in \mathbb{R}$, let $V(s)$ be the vertical line in \mathbb{C} through s and let $H(s)$ be the vertical line in \mathbb{C} through is .

Let H be any horizontal line in \mathbb{C} that is not \mathbb{R} . As $m \circ f(\mathbb{R}) = \mathbb{R}$ and H, \mathbb{R} are disjoint, $m \circ f(H)$ and $m \circ f(\mathbb{R}) = \mathbb{R}$ are disjoint lines in \mathbb{C} , so H is a horizontal line \mathbb{C} . As $m \circ f(\mathbb{H}) = \mathbb{H}$, H lies in \mathbb{H} iff $m \circ f(H)$ lies in \mathbb{H} .

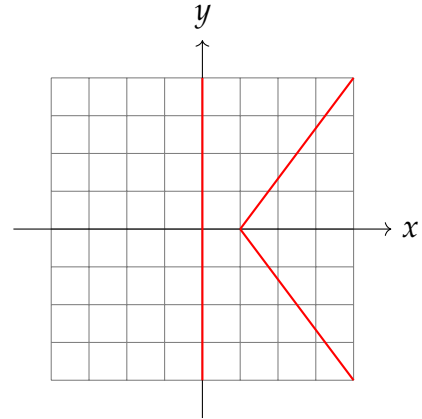
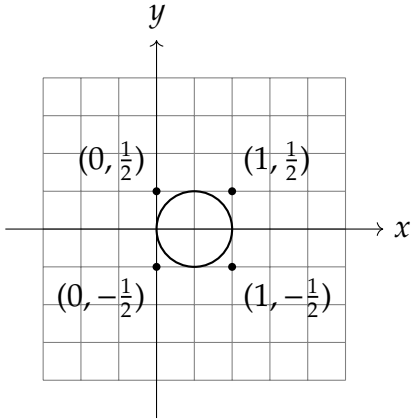
Let A denote the Euclidean circle with Euclidean center $1/2$ and radius $1/2$. As $V(0)$ is tangent (in the geometrical sense) to A at 0 , we see that $m \circ f(V(0))$ is tangent to $m \circ f(A)$ at $m \circ f(0) = 0$, since $m \circ f(A \cap V(0)) = (m \circ f(A)) \cap (m \circ f(V(0)))$ by injectivity. Similarly $m \circ f(V(1))$ is the tangent line to $m \circ f(A)$ at 1 .

As $V(0)$ and $V(1)$ are parallel lines in \mathbb{C} , $m \circ f(V(0))$ and $m \circ f(V(1))$ are also parallel lines in \mathbb{C} (by the same injectivity argument), and by rotational symmetry of the complex plane, our arguments showing that if H is horizontal then H lies in \mathbb{H} iff $m \circ f(H)$ lies in \mathbb{H} , we can conclude that if V is vertical, then V lies in the right half of the complex plane iff $m \circ f(V)$ lies in the right half of the complex plane. Necessarily then $m \circ f(V(1))$ is a vertical line going through 1 and is hence $V(1)$, and by parallel $m \circ f(V(0)) = V(0)$.

This forces $f(A) = A$, as the tangent lines through 0 and 1 of any other Euclidean circle passing through 0 and 1 are not parallel. One may see this by seeing that if $(x - a)^2 + (y - b)^2 = r^2$, and 0 and 1 lie on this circle, then $a^2 + b^2 = r^2$, and $(1 - a)^2 + b^2 = r^2$, yielding $a^2 - (1 - a)^2 = 0$, which says that $a = 1/2$. An explicit calculation shows that $r \geq 1$, and that if $r > 1$ the above claim holds. Even though $m \circ f(A) = A$, we do not know that A contains any points of Z other than 0 and 1 .

So we run the same argument with the two horizontal tangent lines to A . Considering the first tangent $H(1/2)$ to A at $1/2 + 1/2i$, we see that $m \circ f(H(1/2))$ is again a horizontal line in \mathbb{H} tangent to $m \circ f(A)$, and so $m \circ f(H(1/2)) = H(1/2)$ (A has only one tangent line that is horizontal in the upper half plane). This gives more points in Z , such as the intersection of the above horizontal and vertical lines, i.e. the points $1/2i$ and $1 + 1/2i$. Running the same argument on $H(-1/2)$ yields that the points $-1/2i$ and $1 - 1/2i$ lie in Z .

By running this exact same argument on the circle centered at $x > 0$ with radius x , we get that the entire imaginary axis is in Z , and the line segments starting from $(1/2, 0)$ with slope of 1, and starting from $(1/2, 0)$ with slope of -1.



One sees that given any two points in Z , the line segment connecting them is also in Z , since two points uniquely determine a line. For any $x > 0$, since $(0, x)$ and $(1/2 + x, x)$ are in Z , the horizontal line segment connecting them is as well. Similarly, for $x < 0$ the line segment connecting $(0, x)$ and $(1/2 + x, -x)$ lies in Z . By also getting 0 and 1/2 by continuity, this yields all the points in \mathbb{C} , thus $m \circ f$ is the identity map when restricted to \mathbb{C} , and hence also the identity map on $\overline{\mathbb{C}}$ by continuity. Then $f = m^{-1}$ is a Mobius transformation, or the composition of a Mobius transformation with complex conjugation, which concludes the proof of the theorem, and my paper. \square

References