Algebraic Number Starscapes

EDMUND HARRISS, KATHERINE E. STANGE, STEVE TRETTEL

Abstract. We study the geometry of algebraic numbers in the complex plane, and their Diophantine approximation, aided by extensive computer visualization. Motivated by the resulting images, which we have called *algebraic starscapes*, we describe the geometry of the map from the coefficient space of polynomials to the root space, focussing on the quadratic and cubic cases. The geometry describes and explains the notable features of the illustrations, and motivates a geometric-minded recasting of fundamental results in the Diophantine approximation of the complex plane. Meanwhile, the images provide a case-study in the symbiosis of illustration and research, and an entry-point to geometry and number theory for a wider audience. In particular, the paper is written to provide an accessible introduction to the study of homogeneous geometry and Diophantine approximation.

We investigate the homogeneous geometry of root and coefficient spaces under the natural $\mathrm{PSL}(2;\mathbb{C})$ action. Hyperbolic geometry and the discriminant play an important role in low degree. In particular, we rediscover the quadratic and cubic root formulas as isometries of $\mathbb{H}^2$ and its unit tangent bundle, respectively. Utilizing this geometry, we determine when the map sending certain families of polynomials to their complex roots (our starscape images) are embeddings.

We reconsider the fundamental questions of the Diophantine approximation of complex numbers by algebraic numbers of bounded degree, from the geometric perspective developed. In the quadratic case (approximation by quadratic irrationals), we consider approximation in terms of *hyperbolic* distance between roots in the complex plane and the *discriminant* as a measure of arithmetic height on a polynomial. In particular, we determine the supremum on the exponent k for which an algebraic target α has infinitely many approximations β whose hyperbolic distance from α does not exceed $\operatorname{acosh}(1+1/|\Delta_\beta|^k)$. It turns out to fall into two cases, depending on whether α lies on the image of a plane of rational slope in coefficient space (a *rational geodesic*). The result comes as an application of Schmidt's subspace theorem. Our results recover the quadratic case of results of Bugeaud and Evertse, and give some geometric explanation for the dichotomy they discovered [17]. Our statements go a little further in distinguishing approximability in terms of whether the target or approximations lie on rational geodesics.

The paper comes with accompanying software, and finishes with a wide variety of open problems.

Date: August 27, 2022.

2010 *Mathematics Subject Classification.* Primary: 11R04, 11R11, 11R16, 11J04, 11J68, 11J87, 53C30 Secondary: 11G50, 11H99, 54E99, 57M99.

This material is based upon work supported by the National Science Foundation under Grant No. DMS-1439786 and the Alfred P. Sloan Foundation award G-2019-11406 while the authors were in residence at the Institute for Computational and Experimental Research in Mathematics in Providence, RI, during the Illustrating Mathematics program. Katherine E. Stange is supported by NSF CAREER CNS-1652238.

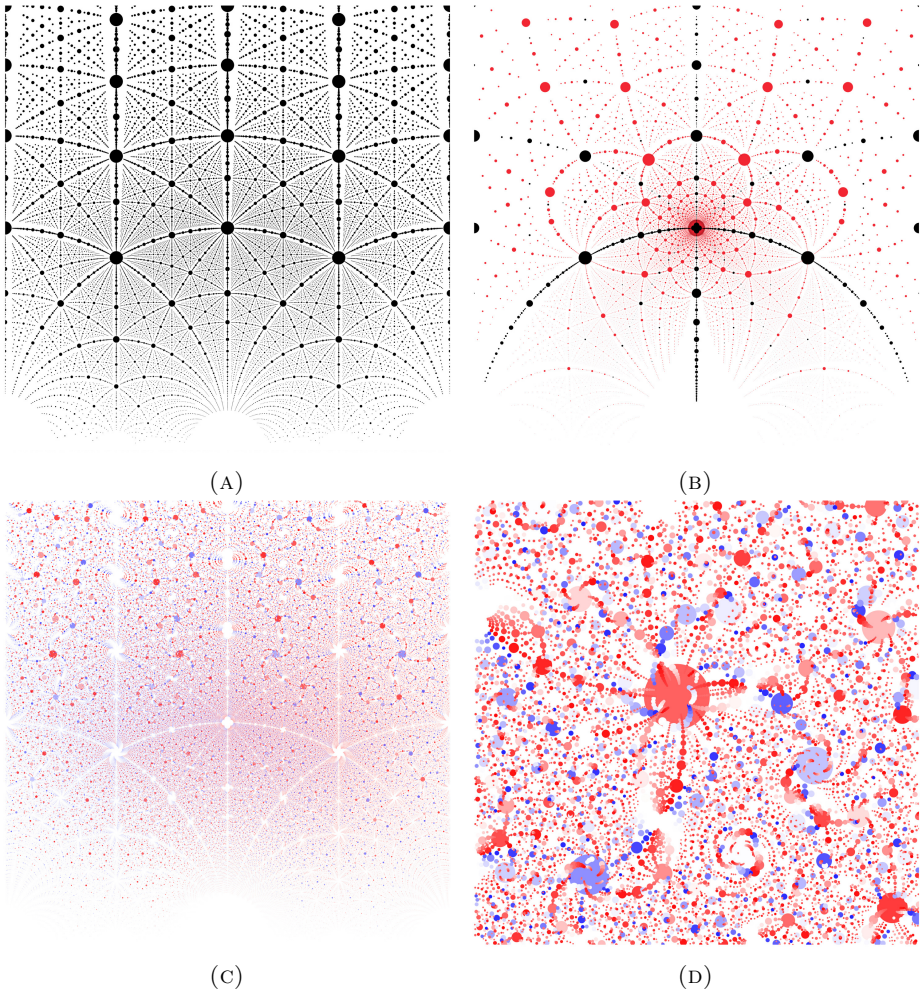
Contents

1. Introduction

We begin (and indeed this research began) with the images in Figure 1. Take a minute to look at them before continuing.

On the top left (1a) you see complex quadratic algebraic numbers plotted and sized by discriminant in the hyperbolic metric[1]. On the top right (1b) are the complex roots of polynomials

[1]For the polynomial $ax^2 + bx + c = 0$ the dot plotted will be at $\frac{-b+\sqrt{4ac-b^2}i}{2a}$ on the complex plane. Note this only considers polynomials with negative discriminant and thus complex roots. A similar image can be created for polynomials with only real roots, see Figure 9. The radius of the dot is proportional to $\frac{1}{\sqrt{4ac-b^2}}$ (one over the root discriminant) times the height above the real axis $\frac{\sqrt{4ac-b^2}}{2a}$ to adjust the radius to the hyperbolic



(A) (B)

(C) (D)

FIGURE 1. Complex algebraic numbers sized by the inverse of the discriminant of their minimal polynomial in the hyperbolic metric. All quadratics (1a). All roots (quadratic in black, cubic in red) of the polynomials $ax^3+cx^2+bx+c=0$ (1b). All cubics, coloured based on the value of the real root (1c) and the detail of the cubics around a root of $x^3+x+1=0$. The first three images, as many other figures here, are plotted from -1 to 1 in the real axis and 0 to 2 in the imaginary axis.

of the form $ax^3 + cx^2 + bx + c = 0$, and on the bottom all complex cubics coloured by their real conjugate (1c), with zoomed in detail around the complex root of $x^3 + x + 1$ (1d). In each case the dot size is inversely proportional to the root discriminant (the discriminant to the root of the degree), and the points are plotted in the hyperbolic metric (meaning, the radius is hyperbolic in the upper half-plane model). This paper grew out of our excitement at the beauty and detail present in these images and the search for mathematics that could be both seen and conjectured from their structure.

metric. This gives dots with radius proportional to $\frac{1}{2a}$. If you are plotting the points yourself it can be useful to adjust the scale of the dots (keeping the same proportions) as more are added.

1.1. **A note on our expository approach.** We believe that the images presented here provide a motivated path into several topics in geometry and number theory. Mathematical beauty can be incredibly hard to communicate to people not familiar with the details of the subject, yet these images have already appeared in an art exhibit in Iceland [32] and been used for engagement in a workshop with the Math Club for Battle Creek Area Math and Science Center in Michigan. We have therefore attempted to build a paper containing two distinct paths for two distinct audiences. For those with less background, we provide a leisurely introduction to the subject keeping the description as accessible as possible, introducing even well-known terminology as we build toward more sophisticated mathematics. We hope the paper might provide insight and interest to a motivated high school student and a mathematician in these research areas alike, and inspire REU projects. For the researcher who wishes to access the results directly, we have endeavoured to make a bypass that avoids the more leisurely parts of the paper, and provide an alternative, condensed introductory section that allows the reader to pass directly to the results.

1.2. **How to access this paper for a wider audience.** For a motivated student of mathematics, this paper should be read in chronological order, skipping Section 2. In Section 3, we follow the visual investigation with a gallery of some of the images produced, and discussion of some of the structural aspects visible in the images. We develop those observations in the next two sections studying the geometry of the roots (in Section 4) and their number theory (in Sections 5 and 6). These (geometry and number theory) are somewhat parallel, in that a reader may wish to read both to their natural stopping point (dictated by the reader's mathematical background). Each one starts from a lower level of background and ramps up.

1.3. **How researchers should read this paper.** For a working mathematician with experience in the relevant background, Section 2 should be read first, followed by Section 4, and then Section 6, dipping into Section 5 for background as needed.

1.4. **Previous illustration.** The notion of plotting algebraic numbers is not new, and there are many beautiful mathematical visualisations made from them, many in the world of blogs and online mathematical discussions. We curate those we are aware follows at the end of this introduction. Our images were studied initially from the perspective of their aesthetic, by asking, "What makes images that look interesting?" (without trying too hard to define *interesting*). However, they rapidly became both a tool to illuminate existing mathematics, including hyperbolic and projective geometry, representations of $\mathrm{PSL}(2;\mathbb{C})$, and Diophantine approximation, and also a source of new mathematics.

The richest previous investigation of imagery grows out of the study of the roots of Littlewood polynomials [43, 49, 59], which are polynomials with coefficients ± 1, and more generally polynomials with coefficients from a finite set. This produces the oldest images we have found, particularly in the work of Peter and Jonathan Borwein [7, 11, 12, 13, 52]. These collections of polynomials have the nice feature that all polynomials up to a certain degree can be considered. Other investigations of the geometry and images related to such polynomials (that we will not describe in detail) include the curiously named Thurston's master teapot [14] and the eigenvalues of Bohemian Matrices [64, 19].

In particular, a lot of interest was generated by the work of Dan Christiansen [20], shared and described by John Baez, including an incredibly intricate image created by Sam Derbyshire [5, 6, 46].

Many other people, some directly inspired by this, have also created, or discussed, the images, including Paul Nylander [48], Greg Egan [26], Andrej Bauer [8], Vincent Pantaloni [50], Daniel Wiegreffe [66], Bernat Espigulé [28], Jwalin Bhatt [9], Jonathan Lidbeck [41], and Jordan Ellenberg [27].

Pictures not limiting coefficients so strongly, closer to those we present here, are rarer. The most notable are the images of Stephen Brooks used on Wikipedia [15], that were further developed and optimised by David Moore [4]. These images also inspired a Wolfram Demonstration from Enrique Zeleny [68]. Other interesting images we have found come from David Marciel [45] and Deviant Art user Fauxtographique [29]. There have also been discussions on Stackexchange [30, 35], with the latter linking to further amazing images on the Flickr account of Stackexchange user "DumpsterDoofus" [55]. The ease of creating these images means that people quickly start making their own: for example, a twitter thread started by the first-named author [31] quickly prompted variations by Dan Anderson, Michael Pershan and Peter Farrell [1, 2, 51].

This list is extensive, but probably not exhaustive. We are interested in other versions of such patterns, especially earlier ones (before 2010, and even more before 2000) so please send us any you know of.

1.5. **Software and image generation.** We encourage the reader to explore along with us, throughout the paper, using the accompanying software, available as a Sage Mathematics Software [63] notebook, at `algebraicstarscapes.com`.

In general the images here are produced by a rather simple three step process. We first generate a list of polynomials whose coefficients lie in a region about the origin (for example, a box or ball). These polynomials are then solved to give the collection of roots, with additional data (such as polynomial discriminant) attached to the roots, data which is eventually to be used for sizing. Finally, that list of data is converted into a collection of points and plotted. Most of the images drawn involve over 50,000 dots, but some get to over 250,000.

1.6. **Acknowledgements.** The authors are grateful to the Institute for Computational and Experimental Research in Mathematics in Providence, RI, and to the semester organizers, for the opportunity to be in residence for Fall 2019 at the Illustrating Mathematics program, where this work was initiated in the grand tradition of just being in the right place at the right time. The authors would also like to thank their respective home institutions for their help in making semester residency possible. Thanks are also due to the many participants in that program for helpful discussions, including Arthur Baragar and Joseph H. Silverman. Special thanks go to Pierre Arnoux and David Dumas for especially inspiring and detailed conversations as these ideas developed.

2. A technical introduction

2.1. **Algebraic starscapes.** The images central to our story we have called *algebraic starscapes*. Formally, these images consist of dots centred at all algebraic roots of a family of polynomials, with radius a function of the coefficients of the relevant polynomial. More specifically, these families are chosen by fixing a bound on the polynomial degree and allowing the coefficients to range through all integer points in some affine subspace of the full vector space of coefficients. The sizing is typically chosen from various measures of arithmetic complexity, such as Weil height or polynomial discriminant, so that big dots correspond to low complexity. A *linear starscape* is formed when the family of polynomials is two-dimensional; these appear as beaded necklaces (see Figure 25); *planar starscapes* are formed from three-dimensional families (see Figure 2). Planar starscapes contain infinitely many linear starscapes. The reader is invited to examine the examples in Figures 2a through 2f. These pictures can all be considered generalisations of Figure 15. The 'repulsion' of large dots from one another illustrates an analog to Dirichlet's approximation theorem (Theorem 3.1), stating that rational numbers cannot well approximate other rational numbers (see also Figure 14).

As an aside, although we restrict our attention to complex roots, there's no reason one cannot seek analogous visualizations for real roots. For example, Figure 9 shows the real pairs that are roots of quadratics. Or, for cubics with one real root, one could parametrize polynomials by a

complex root in the upper half plane together with a real root on its ideal boundary: this gives us a starscape picture living in a solid torus as in Figure 10. See Section 7 for further avenues.

The purpose of this paper is to study algebraic numbers in the complex plane, including their Diophantine approximation properties, as a function of the homogeneous geometry of the coefficients-to-roots map. That is, we aim to describe exactly the map from coefficient space (the space of coefficient vectors of polynomials of fixed degree), and its affine subspaces, to the root space (collections of points in the complex plane). Starscapes represent the images of affine subspaces (strictly speaking, the starscape is formed by plotting all the roots occurring in the root space image). The philosophy is that, with a sufficient understanding of the geometry, we can state and prove concrete Diophantine results in low degree.

2.2. **Geometry.** The arithmetic complexity of an algebraic number, although measured in a variety of ways (see Figures 12 and 13), is typically correlated to the size of the integer coefficients of its minimal polynomial. The big dots in the image, then, are the image of the short integer vectors in the affine subspace. If the geometric map from coefficient space to root space is sufficiently convoluted, these dots may end up close together in the complex plane. If we can control the behaviour of the geometric map, we can control this effect and prove Diophantine results.

There are several key geometric points:

(1) The geometric description of the coefficients-to-roots map is studied in detail in degrees two and three. The spaces of coefficients and roots decompose as a union of homogeneous spaces for $\mathrm{PSL}(2;\mathbb{R})$, and in these low degrees we may describe the cases of interest (real polynomials with complex roots) completely in terms of the geometry of the hyperbolic plane and its unit tangent bundle. See Sections 4.2 and 4.3.

(2) In degree two, we demonstrate that the coefficients-to-roots map (more colloquially known as the quadratic formula) realizes an *isometry between two models of the hyperbolic plane*. See Theorem 4.9, Corrolary 4.10 and the associated Figure 22 for a precise formulation. While surely classically known, none of the authors had previously encountered this surprising aspect of the quadratic formula, and so we have provided details accessible to students of hyperbolic geometry.

(3) In degree three, the $\mathrm{PSL}(2,\mathbb{R})$ action naturally identifies the space of real cubics possessing a complex root with the unit tangent bundle to the hyperbolic plane. This identification allows a recasting of many familiar algebraic results in geometric terms. As a particular example we see the fact that every real cubic has a real root provides a preferred trivialization $\mathrm{UT}\mathbb{H}^2$ (Proposition 4.17), and the cubic formula can be interpreted as a means of explicitly identifying cubics in coefficient-space with their coordinates with respect to this trivialization (Theorem 4.27).

(4) Concerning the roots map in degree three, the projection onto complex roots naturally identifies with the bundle map $\mathrm{UT}\mathbb{H}^2 \to \mathbb{H}^2$. We study the interaction of the lattice of integer points with this map: for example, can rational affine subspaces be contained in the fibres? We discuss this, and its implications for starscapes, in Section 4.3.4.

2.3. **Diophantine Approximation.** Diophantine approximation can be described as the quantitative study of the trade-off that is required to approximate a real number from a set of approximations (for example, rational numbers), namely between the precision of the approximation and the complexity of the approximant. Most of the classical story lives on the real line, so that, given $\alpha \in \mathbb{R}$, we ask for $p/q \in \mathbb{Q}$ (if approximating with rationals), so that

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{q^k}$$

for various positive k. Dirichlet's Theorem (Theorem 3.1) asserts that for $\alpha \notin \mathbb{Q}$, and $k = 2$, there are infinitely many such approximations, while for algebraic α, and $k > 2$, Roth's Theorem

(Theorem 5.2) asserts that there are only finitely many. Thus, the exponent $k = 2$ is a critical exponent for approximation of algebraic numbers by rationals. We might choose approximations from other sets, such as algebraic numbers of bounded degree. Koksma defines $k_d(\alpha)$ to be the supremum of all k such that there are infinitely many algebraic β of degree $\leq d$ satisfying

$$|\alpha - \beta| < \frac{1}{H(f_\beta)^k}.$$

Here, H refers to the naïve height, and f_β to the minimal polynomial of β.

These questions have naturally been extended to the rest of the complex plane, typically by reference to the euclidean distance in the complex plane. Thus one may define $k_d(\alpha)$ for $\alpha \in \mathbb{C}$ also. Sprindžuk showed that the real and complex cases are essentially different: $k_d(\alpha) = d+1$ for almost all $\alpha \in \mathbb{R}$, but $k_d(\alpha) = (d+1)/2$ for almost all $\alpha \in \mathbb{C}$ (see Theorem 5.3). Bugeaud and Evertse where able to determine $k_d(\alpha)$ for most algebraic $\alpha \in \mathbb{C}\backslash\mathbb{R}$ (Theorem 5.4). In particular, in the quadratic case they discovered that $k_2(\alpha) = 2$ or $3/2$, depending upon whether the quantities 1, $\alpha\overline{\alpha}$ and $\alpha + \overline{\alpha}$ are linearly dependent or not, respectively.

However, motivated by the geometry of the coefficients-to-roots map which describes the visual features of the algebraic starscapes, we propose that it make sense to ask about Diophantine approximation in the hyperbolic metric and with a sensitivity to the action of $\mathrm{PSL}(2;\mathbb{Z})$ implied by the homogeneous geometry. Thus we consider instead the critical exponent k when asking for approximations in the following sense:

$$d_{hyp}(\alpha, \beta) \leq \operatorname{arcosh}\left(1 + \frac{1}{|\Delta_\beta|^k}\right).$$

Here, Δ_β is the discriminant of the minimal polynomial of β, and d_{hyp} represents the hyperbolic distance. This inequality is preserved under $\mathrm{PSL}(2;\mathbb{Z})$ transformations, and we demonstrate, using Schmidt's Subspace Theorem, that the critical exponent is either 2 or $3/2$, the difference controlled as in Bugeaud and Evertse's result. The dependence of 1, $\alpha\overline{\alpha}$ and $\alpha + \overline{\alpha}$ can be reinterpreted as the condition that α lie on a *rational geodesic* (see Section 4.2.5), i.e. the image of a plane of rational slope in the coefficient space. These are exactly the linear starscapes of Figure 1a, which are more densely packed with quadratic irrationals.

This essentially recovers (up to some nuances discussed in Section 6.1.1) the result of Bugeaud and Evertse for degree 2. But it begs the question if higher degree cases (some of which are not completely settled) can be similarly described in terms of the geometry of the coefficients-to-roots map.

2.4. **Reading Guide.** In Section 4, we describe in detail the map from coefficient space to root space. A research audience may wish to skip Section 4.1, briefly looking at Section 4.1.3 for the roots map in the linear case, and begin with Section 4.2.4, which concerns quadratic polynomials. Section 4.2.5 concerns the consequences of the quadratic geometric story for integer points, which is relevant to the Diophantine approximation we will do later. Section 4.3 tells a similar story for cubics using the geometry of the unit tangent bundle to the hyperbolic plane, and interprets the cubic formula from this perspective 4.27).

In Section 5, we give the necessary background to Diophantine approximation. The expert will not find anything new in Section 5.1–5.1.1, but we place the Diophantine results in context in Section 5.1.2, and argue for the importance of $\mathrm{PSL}(2;\mathbb{Z})$ in Section 5.2.1. Sections 5.2.2 and 5.2.3 can again be skipped by the expert. Section 5.2.4–5.2.5 discuss the appropriate choice of measure of arithmetic complexity.

The number theorist interested in the quadratic results will therefore wish to read Sections 4.2.4–4.2.5, Sections 5.1.2–5.2.1 and then focus on Section 6 (possibly returning to Sections 4.3 for the cubic geometric story, although our Diophantine results are restricted to quadratics). In Section 6, we revisit the description of the critical exponent for Diophantine approximation of complex numbers by algebraic numbers of degree 2. We use Schmidt's subspace theorem, and we

find that the exponent depends upon whether the complex numbers lies on a rational hyperbolic geodesic (defined in Section 4.2.5). The main theorems are Theorems 6.3 and 6.4 (analogous to Dirichlet's Theorem, asserting infinitely many good quadratic irrational approximations) and Theorem 6.6 (analogous to Roths' Theorem, asserting that there are only finitely many better quadratic irrational approximations).

3. Gallery

3.1. **In a galaxy far, far away.** The starscapes shown in Figure 1 of the quadratics (1a) and the family of cubics (1b) have a striking self-similar curvilinear structure: the figures seem to be populated with beaded necklaces of discs subdividing the plane into smaller regions criss-crossed with similar, finer, necklaces.

This basic pattern occurs quite generally. To give all polynomials of a given degree each coefficient must be allowed to change freely. This defines topological space, the "coefficient space," of polynomials where each coefficient gives a dimension[2]. The total dimension of all polynomials of a given degree is one greater than the degree. It is natural to think of subspaces of this space and the pattern we describe above seems to appear any time one works with a three dimensional linear subspace in coefficient space; more examples appear in Figure 2.

As the degree increases, the curves and their relationships become more complicated, but the basic motif continues. We hope these images justify the term "algebraic starscapes." When the family of polynomials is two-dimensional, we obtain a "linear starscape" (that is, a single beaded necklace threading through the plane; see Figure 25). When a family of polynomials is three-dimensional, we obtain "planar starscapes" such as those in Figure 2. One may continue this to higher dimension, but in four-dimensional families such as all cubics (Figure 1c), the collapse to the complex plane produces much more complicated, less immediately patternful, although nonetheless enticing, pictures.

The simplest version of the basic motif that seems to pervade these images appears when looking at a lattice in perspective, as you can see in Figure 3. In this case the points only form straight lines. The one dimensional version of this effect is sometimes called the orchard illusion and can be experienced when driving past an orchard planted on a grid (Figure 4).

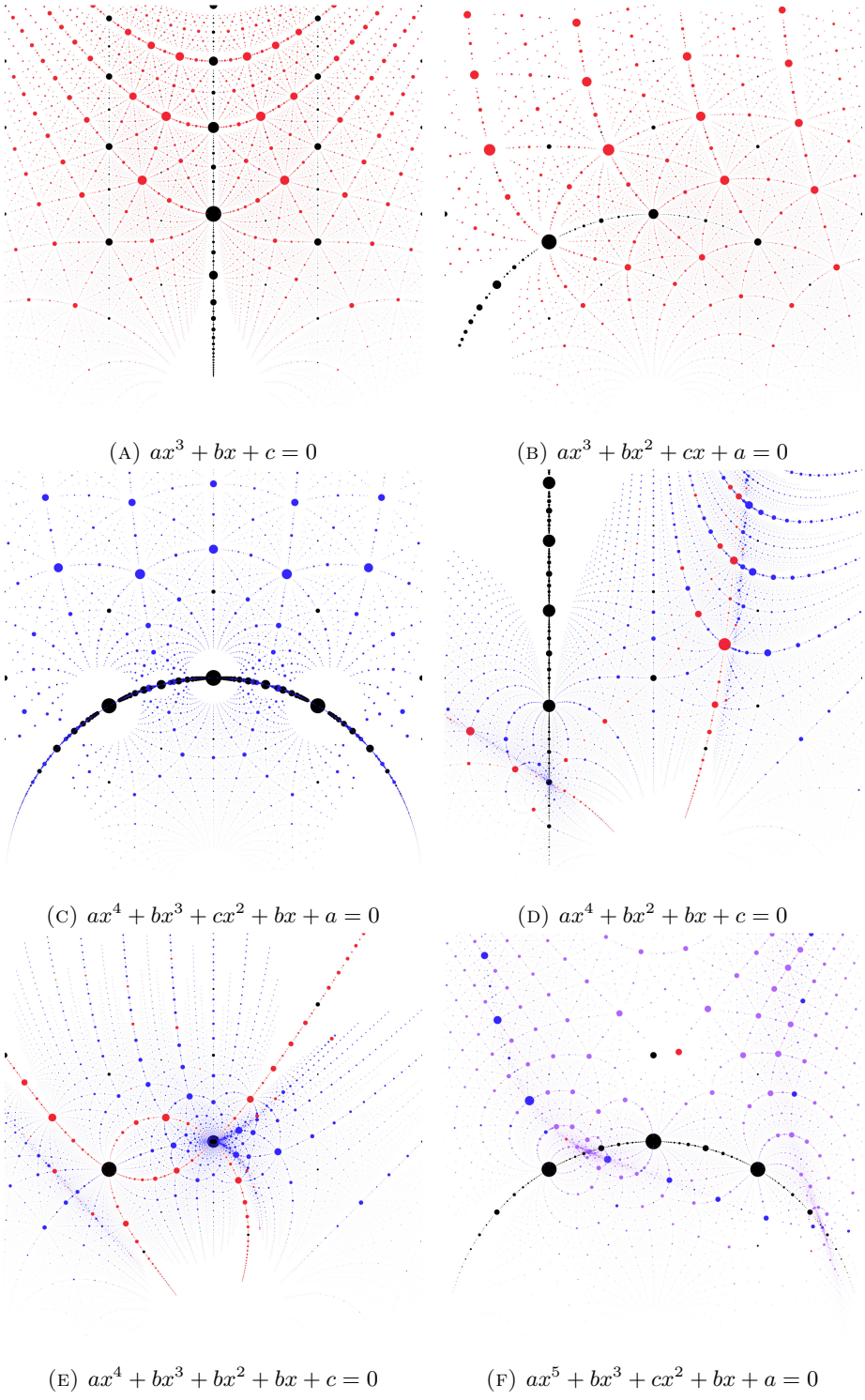
In the quadratic case (Figure 1a) the lines might actually seem familiar to geometers as they are the straight lines (geodesics), not in euclidean geometry, but in the upper half plane model of the hyperbolic plane. In fact Figure 1a is a regular tiling of the upper half plane by a single tile, under the action of a group of symmetries preserving the hyperbolic distance. More precisely, the action is that of the *modular group* $\mathrm{PSL}(2;\mathbb{Z})$ (Figure 5).

While the link to hyperbolic geometry and $\mathrm{PSL}(2;\mathbb{Z})$ is especially strong in the quadratic case, it is more generally useful in understanding starscapes. From an aesthetic perspective, using the hyperbolic metric improves the look of the images close to the real line, as seen in Figure 6. There are more mathematical justifications for this approach described in later sections.

The observations we have described so far constitute a naïve visual approach to the images produced: we are simply asking what it is we are seeing. The underlying geometric explanation, especially for the quadratic and cubic cases, is developed in far more detail in Section 4, and the deeper connections to the study of Diophantine approximation in Sections 5 and 6.

3.2. **Mostly harmless.** The images shown in Section 3.1 are compelling enough to encourage wider investigation. For example, one might move the families considered away from 0 to give affine subspaces of the coefficient space, as shown in Figure 7.

[2]This space starts as $\mathbb{R}^n$, but with polynomials multiplying all coefficients by a constant does not change the roots of the polynomials. It is natural, therefore to consider such polynomials as equivalent. This gives projective geometry one dimension lower, as discussed in Section 4.1.



(A) $ax^3 + bx + c = 0$

(B) $ax^3 + bx^2 + cx + a = 0$

(C) $ax^4 + bx^3 + cx^2 + bx + a = 0$

(D) $ax^4 + bx^2 + bx + c = 0$

(E) $ax^4 + bx^3 + bx^2 + bx + c = 0$

(F) $ax^5 + bx^3 + cx^2 + bx + a = 0$

FIGURE 2. Complex algebraic numbers (roots of each polynomial family with $a, b, c \in \mathbb{Z}$) sized by the root discriminant in the hyperbolic metric. Quadratic numbers are black, cubics red, quartics blue and quintics purple.

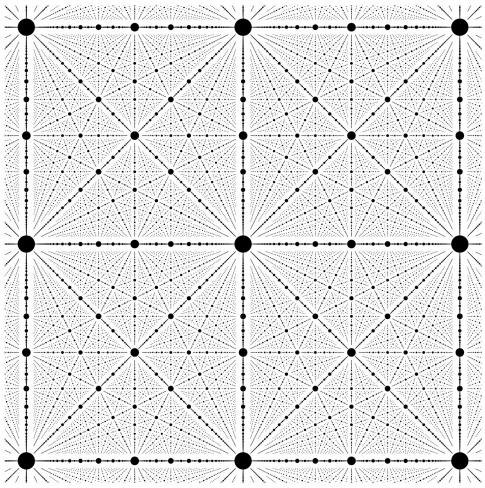


FIGURE 3. The points $(b/a, c/a)$, for $a, b, c \in \mathbb{Z}$ with radius proportional to $1/a$. To see how this links to the view of a lattice for an observer, see Figure 17.

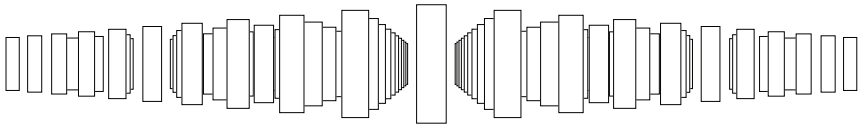


FIGURE 4. A sideways view of a grid of rectangles in perspective forming the orchard illusion.

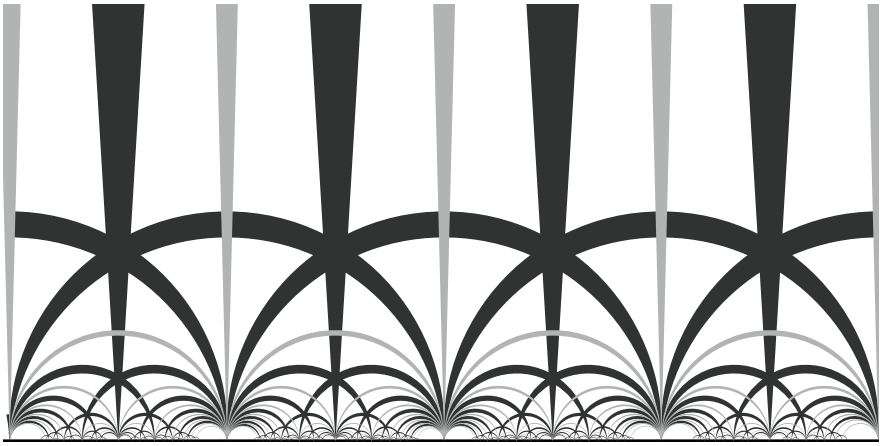


FIGURE 5. A tiling by infinite polygons on the hyperbolic plane (a partition of the hyperbolic plane by the action of $\mathrm{PSL}(2;\mathbb{Z})$). The curves seen here are some of those seen in Figure 1a.

In these affine subspaces, a couple have particular interest: the algebraic integers (where the leading coefficient is 1) and the algebraic integer units (where the leading and constant coefficients are both 1). These are shown for the cubics and quartics in Figure 8.

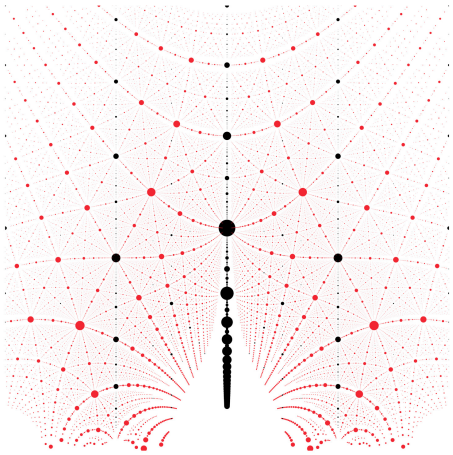


FIGURE 6. The depressed cubics, sized by the root discriminant, plotted in the euclidean metric (meaning, radii are interpreted as euclidean). Note how the plot is very dense at the bottom, but light at the top.

Another approach is to consider real roots, or tuples of roots. For example, quadratics with real roots can be plotted in $\mathbb{R} \times \mathbb{R}$ (Figure 9), though the lack of an ordering on the roots will cause each to appear twice.

A powerful property of the images of the quadratics with complex roots is that they show *all* the information about the roots, as the complex roots come as a complex conjugate pair (so that all the information is shown with just one of them). To extend this to the cubics requires an additional dimension. In this case the cubic polynomials with complex roots always have an additional real root.

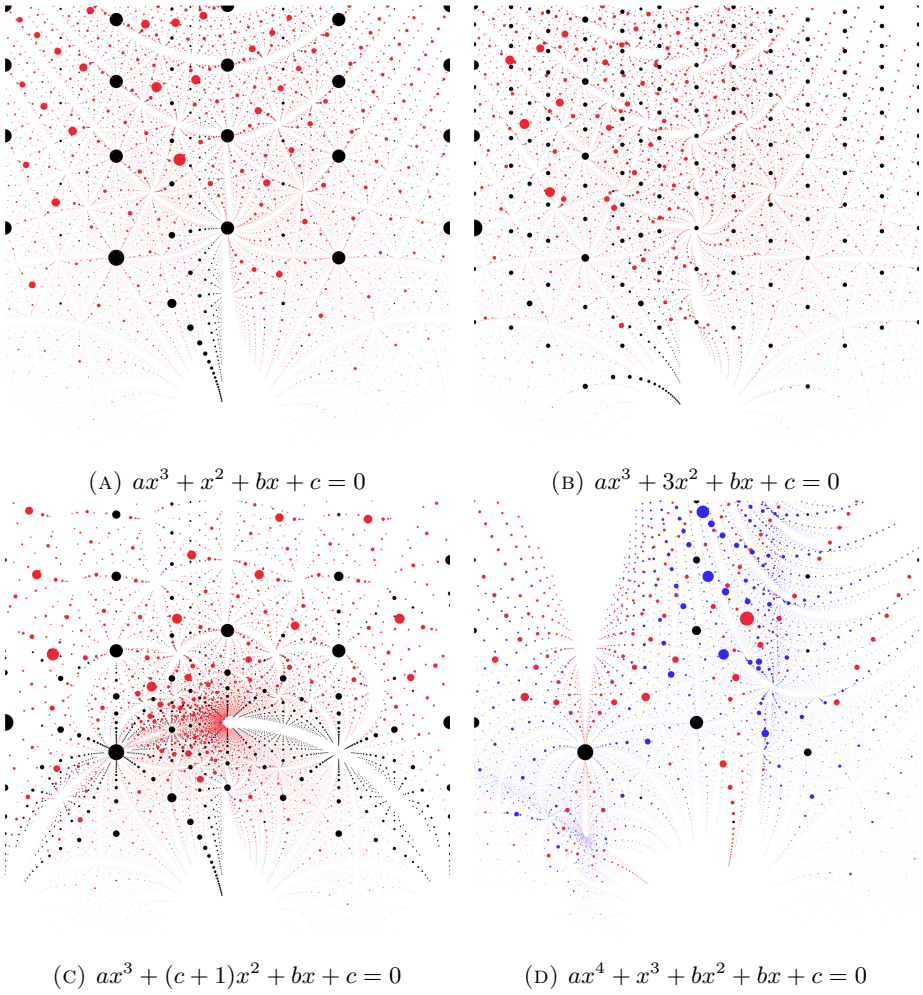
To take into account this additional information, we could take a space in $\mathbb{C} \times \mathbb{R}$ given by the upper half-plane for the complex root and the real line for the real root. We can then add some rather beautiful geometry to this space, described in detail in Section 4.3. This corresponds to considering the upper half plane of $\mathbb{C}$ as the hyperbolic plane. In this model the real line in $\mathbb{C}$ lies on the boundary of the hyperbolic plane, which contains one additional point, the point at infinity that pulls the real line back into a circle by connecting the two ends. This can be considered as the difference between considering the slope and angle of a line on the plane. As the slope gets more positive or negative the line gets closer to vertical. As a slope this is not obtainable, but it is can be considered as an angle. The complex root can therefore look towards any point on this circular boundary.

Considering the pair of a point and direction in a space gives the geometry of the "unit tangent bundle". For hyperbolic geometry this can be considered to be a solid torus, where the circular slices are a disk model of the hyperbolic plane. This torus is a finite region of three dimensional space so, although we took a bit of a journey the result gives a powerful way to see the roots of cubic polynomials, as shown in Figure 10.

The cubic families we have considered nicely embed into this picture as 2d surfaces. The polynomials $ax^3 + bx^2 + ax + c = 0$ having a complex root even create a Möbius strip (Figure 10b). Both of these images are even more powerful when you can manipulate them yourself in 3d, and we encourage you to check out David Dumas' beautiful software SL(View) that we used to make these images ourselves [25]; see `algebraicstarscapes.com` for the datasets.

A different approach to using the unit tangent bundle is to draw arrows, rather than dots on the plane. Such images are shown in Figures 29 and 32.

There are many other spaces to explore that have the potential to reveal many aspects of the structure of algebraic numbers and illustrate various geometric ideas. Using the colour of the



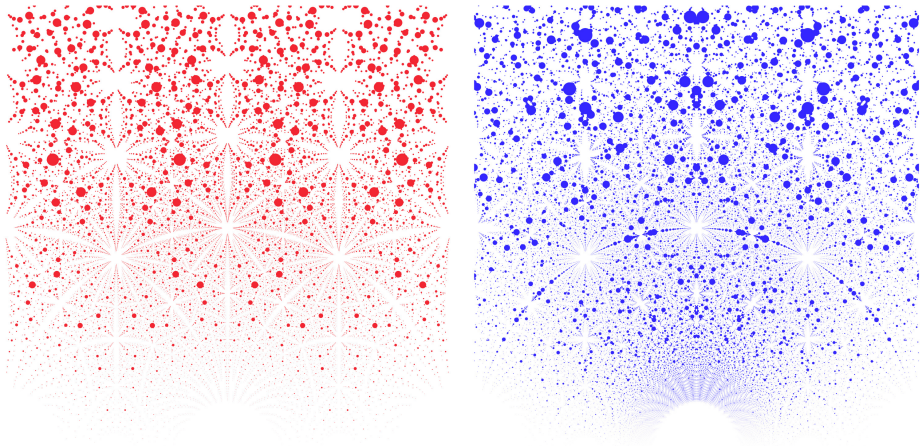
(A) $ax^3 + x^2 + bx + c = 0$

(B) $ax^3 + 3x^2 + bx + c = 0$

(C) $ax^3 + (c+1)x^2 + bx + c = 0$

(D) $ax^4 + x^3 + bx^2 + bx + c = 0$

FIGURE 7. Roots on affine planes in coefficient space.



(A) Cubic algebraic integers

(B) Quartic *unit* algebraic integers

FIGURE 8. Algebraic integers and unit integers.

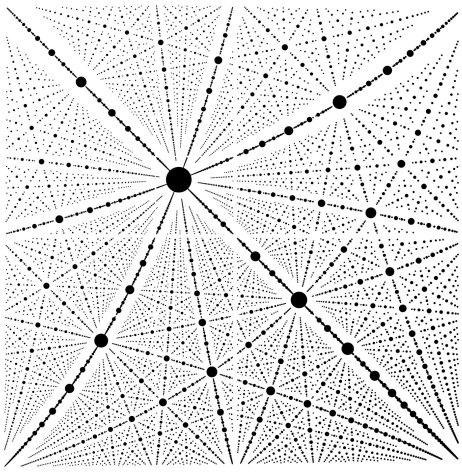
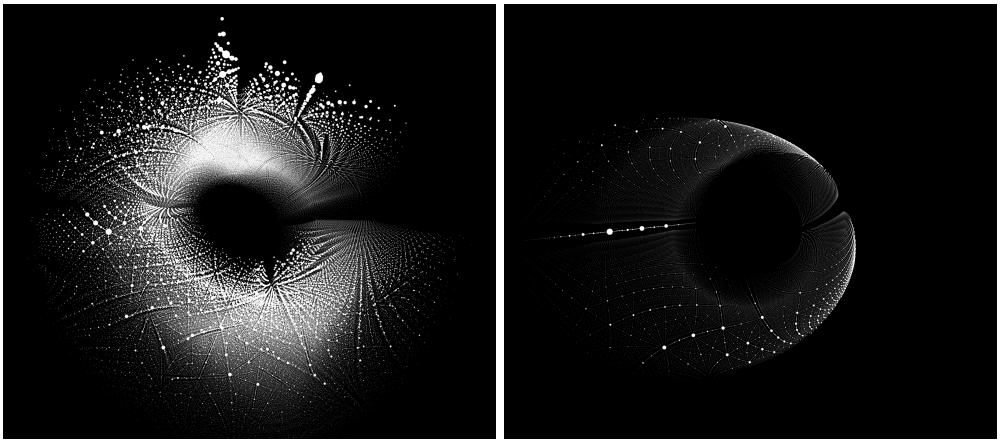


FIGURE 9. The real quadratics plotted against their algebraic conjugate, with one root (x-axis) between -1 and 0 and the other between 1 and 2. The large dot is the Golden ratio $(\frac{1-\sqrt{5}}{2}, \frac{1+\sqrt{5}}{2})$.



(A) (B)

FIGURE 10. All cubic polynomials shown in the unit tangent bundle to $\mathbb{H}^2$ (10a) and the roots of $ax^3 + bx^2 + ax + c = 0$ (10b), forming a Möbis strip within it. Both are shown from the same angle so if you look carefully you can see the second image in the first. Images are screenshots from SL(View) by David Dumas [25].

dots has the potential to give pictures with up to 6 dimensions of information (3 spatial and 3 colour (red, green and blue, for example)). As an example, more of the structure of Figure 2d is revealed when real roots are used to colour points, as shown in Figure 11.

The challenge is to produce images that are attractive, informative or, ideally, both. From our explorations so far we have found that chasing the first is a surprisingly reliable (though not guaranteed) path to the second.

3.3. **A celestial dance.** In all our images so far, there is a feeling that larger dots repel each other, like binary stars locked in a mutual orbit, unable to approach. We also see remarkable

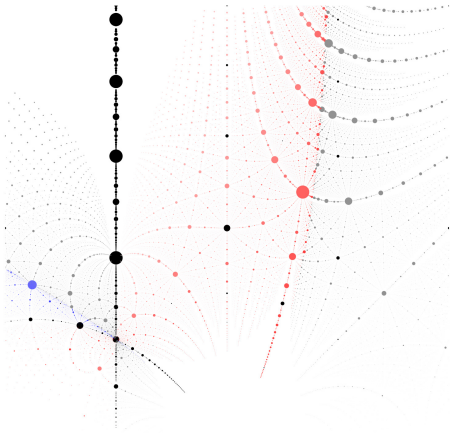


FIGURE 11. The starscape shown in Figure 2d coloured by a real root (simply the first given by Sage), red for negative and blue for positive, each fading to white as the absolute value increases. Quadratic Points are black and points with no real conjugate are shown in gray.

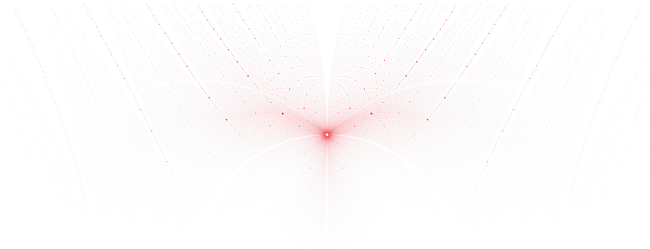
spiraling trajectories of shrinking dots surrounding large ones, like the arms of small galaxies, as in Figure 1d. We have been using the root discriminant as the way to size the dots: this provides a rough measure of the "complexity" of an algebraic number. The notion of needing increasingly complicated algebraic numbers to successively better and better approximate a target is a classic idea in number theory, and is explored in the field of *Diophantine approximation*.

The paradigm is that as the polynomial gets more complicated, the dots representing its roots are drawn smaller. There are many notions of complexity one might use to size the points. Ideas of "complicated" are generally tied to the coefficients (as in the discriminant); Figure 12 shows some classic metrics on the coefficients. In Figure 13a we show the Weil height (or Mahler measure), a more sophisticated way to show complexity. These are all discussed in more detail in Section 5. Compare with the discriminant in Figure 13b.

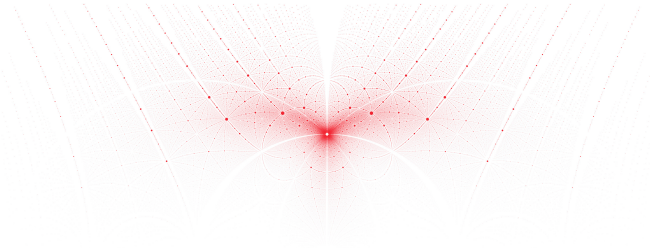
From a geometric perspective, we've seen that this type of repulsion effect can arise from a lattice in projection, with the dots appearing smaller in some sense proportional to their distance. If we consider the lattice of points (p, q) for $p, q \in \mathbb{Z}$, we can map to a line by taking $\frac{p}{q}$ and get the rational numbers, as in Figure 16 and its result, Figure 15. In this setup, the points are sized inversely to their denominators. But as it turns out, the traditional measure of arithmetic complexity of a rational number is in fact, simply its denominator! The geometric aligns with the arithmetic. This is the most elementary example of the relationship between geometry and arithmetic that plays out throughout the paper.

A natural arithmetic question is to ask how well a real number can be approximated by rationals. The first answer to this, of course, is that the rational numbers are dense, so you can get approximations as close as you wish. On the other hand, one might consider $\frac{22}{7}$ to be a surprisingly good approximation for π, given the simplicity of the numerator and denominator. After all, $\left|\pi - \frac{22}{7}\right| < 0.0013$. To quantify this, it's convenient to imagine a cost (the complexity) of a rational number, with higher denominators considered more expensive. We can then ask which approximations are particularly good value for money.

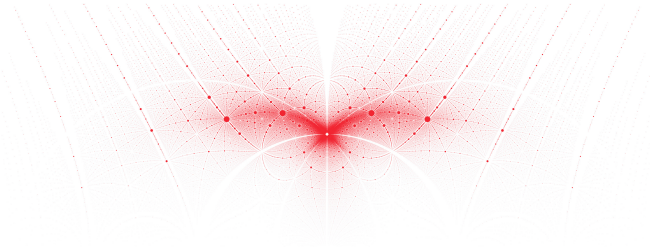
To illustrate such approximations, we could surround each rational number with a disk, where more expensive points have smaller disks; then we could call an approximation to $\alpha \in \mathbb{R}$ *good* if the corresponding disk covers α. This notion is, of course, very dependent on the disk



(A) The 1-Norm on the coefficients (sum of absolute value).



(B) The standard euclidean norm on the coefficients.



(C) The ∞-Norm giving the largest absolute coefficient.

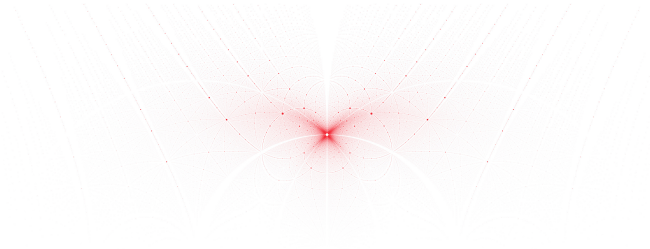
FIGURE 12. Complex roots of polynomials of the form $ax^3+bx^2+cx+a=0$, using different dot sizing from the geometry of the coefficients, given by the inverse of the given quantity. All points with a, b, c between -40 and 40 are plotted, in the region between -2.5 and 2.5 on the real axis and 0 and 2 on the imaginary axis, for a total of 231710 roots.

sizing. It turns out a sizing of $\frac{1}{q^2}$ is a sort of cusp in behaviour, as demonstrated by the following theorem[3] illustrated in Figure 14.

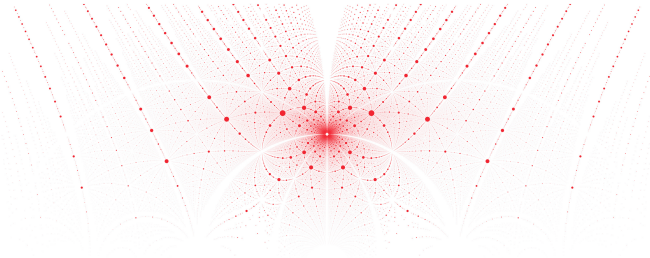
Theorem 3.1 (Dirichlet's Approximation Theorem)**.** *For any $\alpha \in \mathbb{R}$, α is irrational if and only if there exist infinitely many distinct $p/q \in \mathbb{Q}$ such that*

$$|\alpha - p/q| < 1/q^2.$$

[3]The theorem as stated here is most commonly known as Dirichlet's Approximation Theorem, but this theorem was already known to Legendre [39] as a result of the study of continued fractions. The proof we give in Section 5.1 is Dirichlet's and actually gives a stronger *asymptotic approximation* result.



(A) The Weil Height, or Mahler measure.



(B) The root discriminant as used more generally in the paper.

FIGURE 13. Companion to Figure 12, showing two sizings based on the polynomial and number theory of the roots.

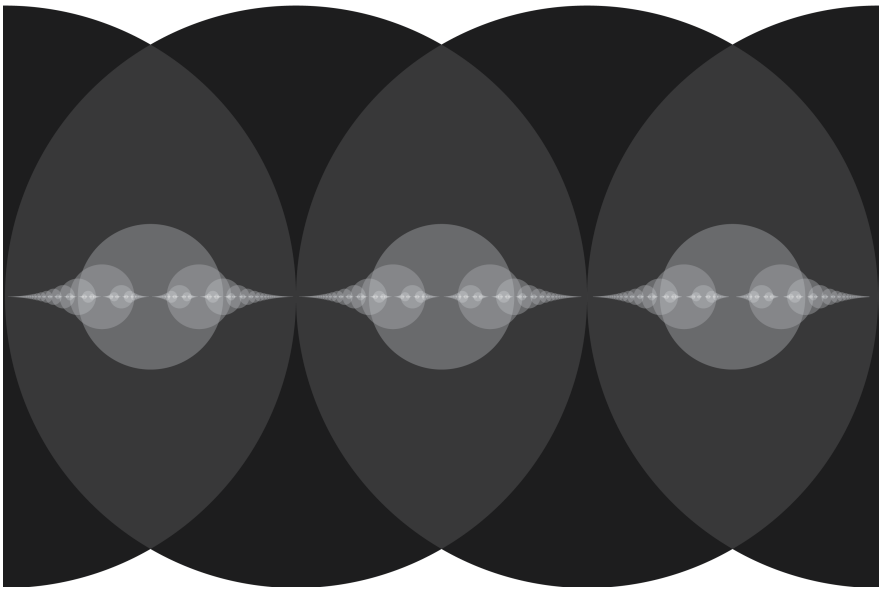


FIGURE 14. Disks centred at rational numbers $\frac{p}{q}$ between 0 and 3 with radius $\frac{1}{q^2}$, illustrating Dirichlet's theorem.

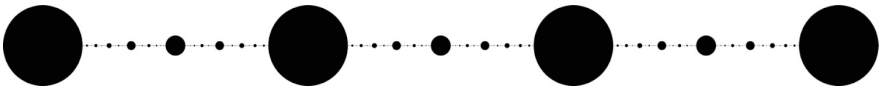


FIGURE 15. Disks centred at rational numbers $\frac{p}{q}$ between 0 and 3 with radius $\frac{3}{20q^2}$.

This gives good approximations to irrational numbers (the p/q that satisfy Theorem 3.1) and also shows that the rational numbers cannot be easily approximated by other rationals, demonstrating the idea that the points are repelling each other. By scaling the points down, the repulsion rather than the notion of approximation becomes clearer, and we have an image close to a one dimensional version of the starscapes (Figure 15).

These ideas and pictures lie at the heart of Diophantine approximation. Linking geometry and number theory, they provide many of the key ideas that can be extended to the algebraic numbers. It is therefore with geometry that we will start the next section, moving from a study of images to the mathematical ideas expressed in them. In later sections, we will return to Diophantine approximation, motivated by the geometry to recast and reprove some variations on and extensions of standard results.

4. Geometry

The gallery images show some structures in the lattice of polynomials over the integers, viewed in two dimensions by drawing their roots in the complex plane. Because polynomials are basic and important objects across mathematics, it should perhaps be no surprise (although it was one to the authors!) that their geometric story involves many familiar characters: from projective geometry and discrete groups, to hyperbolic space, the representation theory of $\mathrm{SL}_2(\mathbb{C})$ and symmetric powers of the sphere. A recurring theme:

> *Patterns in the distribution of algebraic numbers are shadows of the geometry of their lattice of minimal polynomials.*

Throughout the geometry section, we will formulate precise versions of the above statement in degrees two and three, where there is a beautiful connection of the geometry of polynomials to the geometry of the hyperbolic plane. We attempt to increase the required prerequisites only gradually, to allow readers at many levels to chart their own paths through the material. In particular, researchers may want to read the high level overview *Section 4.0.1: A birds eye view* immediately below, and then move directly to any theorem of particular interest. Students may wish instead to skip this fast-paced summary and begin reading at Section 4.1, which provides an introduction to projective space.

4.0.1. *A Bird's Eye View.* The main actor in this story is the roots map $\mathcal{R}$ sending the coefficients of a complex polynomial to its multi-set of roots. This provides a two-way bridge between the space of polynomial solutions and their coefficients, relating the structures in these images directly to structures already present in their sets of minimal polynomials. Making use of this bridge requires understanding what kind of information (topological, representation-theoretic, geometric) survives the trip. We summarize three main themes here:

- The fundamental theorem of algebra implies that the roots map is a *homeomorphism* between the spaces of roots and coefficients over $\mathbb{C}$. Thus topological properties of collections of algebraic numbers are equivalent to topological properties of their corresponding sets of minimal polynomials.
- For $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, the roots map $\mathcal{R}$ is equivariant with respect to natural $\mathrm{PSL}(2;\mathbb{F})$ actions on the spaces of roots and coefficients. This equips each of these with a notion of geometry preserved by $\mathcal{R}$. Thus, geometric properties of the space of minimal polynomials determine geometric properties of the algebraic numbers.

- In small degree, this action has finitely many orbits, decomposing the space of polynomials into a union of homogeneous geometries for $\mathrm{PSL}(2;\mathbb{F})$.

By 'geometric' throughout, we mean in the sense of homogeneous geometry, following Felix Klein. In a vast generalisation of euclidean geometry, Klein proposed in his 1872 Erlagen Programm that a geometry is *defined as a space X, together with the transitive action of a group G*. This group action is interpreted as the allowable, or 'rigid' motions of this geometry, its transitivity implies the geometry is *homogeneous* or behaves the same at every point[4].

In the sections that follow, we give a detailed analysis of the roots map in low degree, and introduce the necessary geometric objects as they arise. In particular, we use the geometry of homogeneous spaces for $\mathrm{PSL}(2;\mathbb{R})$ to give an interpretation of the quadratic and cubic formulas in terms of familiar geometric spaces:

- The space of real quadratics (those with real coefficients) with complex roots identifies with the hyperbolic plane. Restricted to this subset, the roots map (the quadratic formula) is an isometry between the projective model (in coefficient space) and upper half plane model (in root space).
- The space of real cubics with complex roots is topologically a solid torus, and identifies with the unit tangent bundle to the hyperbolic plane. The roots map is an isometry between the models of this geometry constructed from coefficients and roots, respectively.
- Viewing the space of these cubics as the unit tangent bundle to $\mathbb{H}^2$ gives a geometric factorization of the cubic formula. Computing the complex and real roots amounts to a projection onto the base and fiber (with respect to a fixed trivialization) respectively.

As the algebro-geometric study of polynomials of low degree spans centuries, these are almost certainly not new, but we do not know of a reference. We prove them in this section (Theorems 4.3, 4.10, 4.14, 4.26 and 4.27) for the benefit of the reader. These geometric interpretations provide both insight into the gallery images and new perspectives on results from number theory. We summarize some of these insights here.

- The starscapes are naturally interpreted through hyperbolic geometry. Inversion in the unit circle and translation along a horocycle through ∞ preserve integer polynomials, explaining the evident $\mathrm{SL}(2;\mathbb{Z})$ symmetry in figures such as Figure 1a and Figure 1c.
- The roots map is an isometry, so we can measure distances between quadratic numbers explicitly using the discriminant quadratic form on their minimal polynomials, which we will use in Section 6.
- The identification of cubics having complex conjugate roots with the unit tangent bundle to $\mathbb{H}^2$ suggests new ways of visualizing cubic numbers: as subsets of the solid torus (Figures 10a, 28b, 30) and as vector fields in $\mathbb{C}$ (Figure 29).
- Under this identification, two parameter families of cubics embed in $\mathbb{C}$ via the roots map when everywhere transverse to the fibers of the unit tangent bundle. This provides a condition on precisely when the projection onto the complex root is not a homeomorphism: compare and contrast Figures 7c and 32b.

4.0.2. *Notation.* We briefly collect here some useful notation and conventions used throughout. A *complex polynomial of one variable* is a function $f\colon \mathbb{C} \to \mathbb{C}$ of the form $f(x) = a_n x^n + \cdots + a_1 x + a_0$ for $a_0, \ldots a_n \in \mathbb{C}$. Throughout this section we focus instead on *homogeneous polynomials*, which are polynomials in x, y where each term has constant total degree, i.e. of the form $a_n x^n + a_{n-1} x^{n-1} y \cdots + a_1 x y^{n-1} + a_0 y^n$ for $a_0, \ldots, a_n \in \mathbb{C}$. These generalize the single variable case (setting $y = 1$ returns its single variable counterpart) with several advantages.

[4]From this perspective the euclidean plane is $\mathbb{R}^2$ equipped with the group of rotations, translations and (glide) reflections. Projective geometry, hyperbolic geometry, and de Sitter space are other common examples, which we will encounter throughout our journey.

Chiefly, binary forms of degree n naturally include all single variate polynomials of lower degree[5], providing spaces to study all algebraic numbers of bounded degree.

For a fixed degree n, we let Coefs_n be the space of all coefficients of all degree n binary forms, and Roots_n be the space of all multi-sets of their roots (we drop the subscript when no confusion results). For binary forms, it is convenient to allow roots to lie in the extended complex plane $\mathbb{CP}^1 = \mathbb{C} \cup \{\infty\}$. In this way, every polynomial of degree n has exactly n roots (with multiplicity)[6]. The roots map $\mathcal{R}_n \colon \text{Coefs}_n \to \text{Roots}_n$ returns all roots of a polynomial as a function of its coefficients. As the roots of a polynomial are unchanged by rescaling all coefficients by a constant, we abuse notation and also write $\mathcal{R}_n$ for the roots map on scaling classes $\mathcal{R}_n \colon \mathbb{P}\text{Coefs}_n \to \text{Roots}_n$. Similarly, we write $\Delta_n \colon \text{Coefs} \to \mathbb{C}$ for the discriminant, dropping n when unambiguous.

Often, we will identify a certain subset of polynomials with some geometric or topological space X. Depending on our perspective, we may find it useful to think of these as either being built out of the polynomial's coefficients or its roots. To keep these two conceptually distinct, we will often decorate X accordingly, writing $X_{\text{Coefs}} \subset \text{Coefs}$ and $X_{\text{Roots}} \subset \text{Roots}$. Some of the important spaces that arise in this way (and will be defined in the following sections) are the projective spaces $\mathbb{RP}^1, \mathbb{CP}^1$, their generalisations $\mathbb{RP}^n, \mathbb{CP}^n$, the symmetric powers of the sphere $\text{SP}^n(\mathbb{CP}^1)$, and the hyperbolic plane $\mathbb{H}^2$.

4.1. **Projective geometry and rational numbers.** An analysis of the rational numbers $\mathbb{Q}$ provides a gentle introduction to the recurring themes of this section, and an excellent first introduction to projective geometry. Secretly of course, this is just the story of linear polynomials over $\mathbb{Z}$ and their roots. The simplicity of the roots map $\mathcal{R}(ax - b) = b/a$ allows us to suppress much of this formalism and focus on examples of our main idea: *complex patterns that reveal themselves when viewed as shadows of some higher dimensional structure.*

After introducing projective geometry, we provide two examples of this. We give the geometric explanation for the otherwise strange mediant addition law $\frac{p}{q} \oplus \frac{r}{s} = \frac{p+r}{q+s}$ used in the construction of Farey sequences, and enumeration of the rationals via the Stern-Brocot tree.

4.1.1. *Projective Space.* Recall that every rational number r is the quotient of two integers $r = p/q$, so it is natural to expect our analysis of $\mathbb{Q}$ to be deeply intertwined with the lattice $\mathbb{Z} \times \mathbb{Z}$. However, to actually build $\mathbb{Q}$ from $\mathbb{Z} \times \mathbb{Z}$, there are two minor complications to be resolved: (1) a pair $(p, q) \in \mathbb{Z} \times \mathbb{Z}$ does not uniquely determine a rational number, as $\frac{p}{q} = \frac{np}{nq}$, and (2) not all pairs $(p, q) \in \mathbb{Z} \times \mathbb{Z}$ determine rational numbers, as division by zero is undefined. Both of these issues evaporate upon the realization that as *ratios* of integers, we should not model rationals by *points in* $\mathbb{Z} \times \mathbb{Z}$ but rather as *slopes*. This is more memorably stated as follows:

> *The rational numbers* $\mathbb{Q}$ *are what the integer lattice* $\mathbb{Z} \times \mathbb{Z}$ *looks like if you stand at the origin* $(0, 0)$ *and look around (Figure 16).*

Formally, considering objects up to scaling is called *projectivization*, and scaling classes of pairs of integers form the *rational projective line* or $\mathbb{QP}^1$:

$$\mathbb{QP}^1 = \{(p, q) \neq \vec{0} \mid p, q \in \mathbb{Z}\} \Big/ (p, q) \sim (np, nq)\,.$$

The points of $\mathbb{QP}^1$ are denoted $[p : q]$. Note that by definition these points are unchanged by a global scaling of their coordinates, resolving issue (1) above: for example $[p : q] = [2p : 2q] = [\frac{p}{q} : 1]$. Thus, one may recover the usual perspective of the rational numbers as points on a line

[5]For example the linear polynomial $2x + 3$ can be thought of as the homogeneous linear polynomial $2x + 3y$, or the homogeneous quadratic $2xy + 3y^2$, or the homogeneous cubic $2xy^2 + 3y^3$, etc.

[6]Again taking the univariate polynomial $2x + 3$ as an example, thought of as a linear homogeneous equation this has a single root. But thought of as a quadratic, such linear equations have an additional root "at infinity".

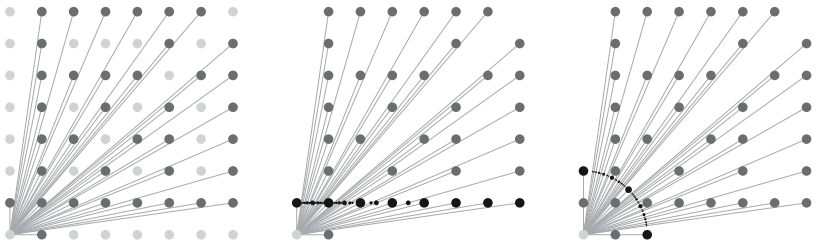


FIGURE 16. The rational numbers as the projection of integer lattice (left) onto an affine patch (middle) and visual sphere (right). The darker point on the left represent the first point on each rational slope, as viewed from the origin (lower left dot).

from this, by imagining a screen (called an *affine patch*) placed a unit distance in front of one's eyes. The rational number $[p : q]$ projects onto this screen to $[\frac{p}{q} : 1]$ (Figure 16, middle).

This definition automatically resolves point (2) as well, and gives a rigorous interpretation for $1/0$ as the projective point $[1 : 0] \in \mathbb{QP}^1$ associated to the horizontal axis in the plane. This line does not intersect the affine patch $\{[x : 1] \mid x \in \mathbb{Q}\}$, so it is not a rational number but a *point at infinity*, often denoted $[1 : 0] = \infty$. The rational projective line fixes the asymmetry of $\mathbb{Q}$'s relation to $\mathbb{Z} \times \mathbb{Z}$ by adding a single point: $\mathbb{QP}^1 = \mathbb{Q} \cup \{\infty\}$. There are many contexts in which scaling classes are the right objects to consider, and this construction generalizes far beyond the rational numbers.

Definition 4.1. Let $\mathbb{F}$ be a field, and $n \in \mathbb{N}$. Then n-dimensional projective space over $\mathbb{F}$, denoted $\mathbb{FP}^n$, is given by scaling classes of the nonzero elements of $\mathbb{F}^{n+1}$ up to elements of $\mathbb{F}^\times$:

$$\mathbb{FP}^n = \{\mathbf{x} = (x_1, \ldots, x_{n+1}) \neq \vec{0} \mid x_i \in \mathbb{F}\} \Big/ (\mathbf{x} = a\mathbf{x}, \quad a \in \mathbb{F}^\times).$$

An equivalence class in $\mathbb{FP}^n$ is denoted $[\mathbf{x}] = [x_1 : x_2 : \cdots : x_{n+1}]$. Given $X \subset \mathbb{F}^{n+1} \backslash \vec{0}$, we write $\mathbb{P}X = \{[x] \mid x \in X\} \subset \mathbb{FP}^n$ for the scaling classes of all elements in X. An *affine patch* of $\mathbb{FP}^n$ is a subset homeomorphic to $\mathbb{F}^n$ equivalent to $\{[x_1 : x_2 : \cdots : x_n : 1]\}$ under some change of coordinates.

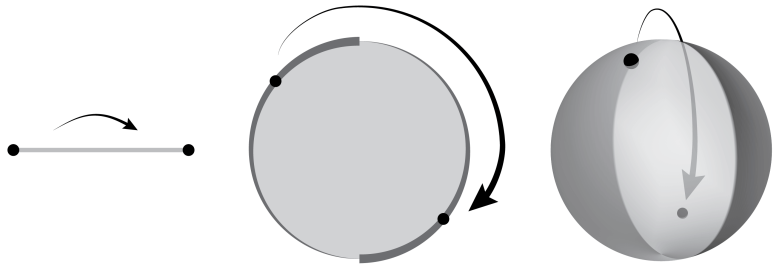


FIGURE 17. Real projective space can be constructed from a solid ball by identifying its boundary via the antipodal map. Here we see this in dimensions 1, 2 and 3.

Remark 4.2. For any field $\mathbb{F}$, we may see that $\mathbb{FP}^1 \cong \mathbb{F} \cup \{[1 : 0]\} = \mathbb{F} \cup \{\infty\}$ by the same argument as for $\mathbb{Q}$. In fact, $\mathbb{FP}^1$ is the *one point compactification of* $\mathbb{F}$. For example, $\mathbb{RP}^1$ is a circle and $\mathbb{CP}^1$ is a sphere.

For us, a main advantage of this perspective[7] is to aid in switching back and forth between thinking about objects in $\mathbb{FP}^n$ and objects in the higher dimensional space $\mathbb{F}^{n+1}$. We give two simple, but surprisingly beautiful applications of this below.

4.1.2. $\mathbb{QP}^1$ *and the Stern-Brocot Tree.* A means of representing the lowest-terms representative of each rational number by an infinite binary tree was discovered independently by Stern and Brocot in the mid 1800s. This tree is depicted in Figure 18. The construction of the tree is inductive, where the elements of the n^{th} row are produced from those in prior rows by taking *mediants*. The *mediant of two fractions* $\frac{p}{q}$ and $\frac{r}{s}$ in lowest terms is the fraction $\frac{p+q}{r+s}$, which appears to be a rather algebraically unnatural construction[8]. Taking mediants is a common step in such enumeration sequences (including the Farey sequence, a source of much beautiful mathematics on $\mathbb{QP}^1$), but is best understood not as an *algebraic operation* on $\mathbb{Q}$ but rather as a *geometric operation* on $\mathbb{QP}^1$. Indeed, before projectivizing, the points (p, q) and (r, s) are vectors in $\mathbb{R}^2$, and here the mediant operation is simply *vector addition*! This is completely natural geometrically on $\mathbb{QP}^1$: given two points, lift to their representatives in $\mathbb{Z}^2$ which are closest to the origin. These two vectors determine two sides of a parallelogram, whose main diagonal connects you directly to the mediant.

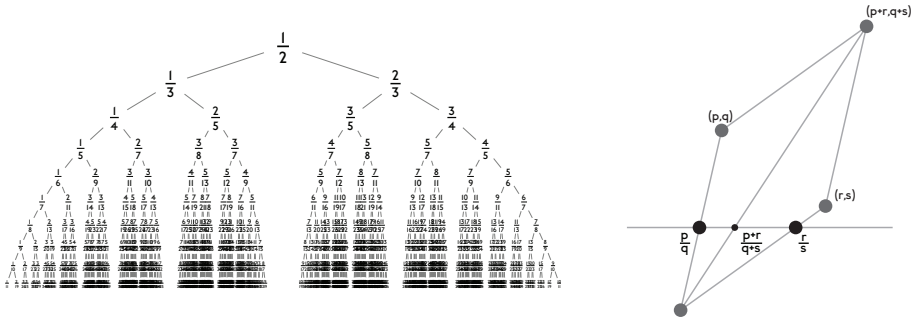


FIGURE 18. The Stern-Brocot Tree enumerating the rationals, and a geometric description of the mediant of two fractions, essential to its construction.

4.1.3. *The Roots Map.* We briefly record the roots to linear equations $ax + b$ in our standard notation, for later use when discussing cubics. Over $\mathbb{C}$, the space of coefficients identifies naturally with $\mathbb{C}^2 = \{(a, b) \mid a, b \in \mathbb{C}\}$, and so up to scaling we have $\mathbb{P}\text{Coefs} = \mathbb{CP}^1$. The space of their roots is also $\mathbb{CP}^1$, thought of as the extended complex plane; as $ax + b$ (thought of as the homogeneous equation $ax + by$ on $\mathbb{CP}^1$), has the unique root $[-b : a] = [-b/a : 1]$ when $a \neq 0$ and otherwise $\infty = [1 : 0]$. Thus the roots map here is a linear isomorphism of $\mathbb{CP}^1$ with itself:

$$\mathcal{R}_1 \colon \mathbb{CP}^1_{\text{Coefs}} \to \mathbb{CP}^1_{\text{Roots}} \qquad [a : b] \mapsto [-b : a]$$

This gives a natural identification between the spaces of roots and (projectivized) coefficients, allowing us to simplify the stories above. In the following sections, $\mathcal{R}$ will continue to be a homeomorphism $\mathbb{P}\text{Coefs} \to \text{Roots}$, and topologically the mapping from a polynomial's coefficients to its solutions is still equivalent to projectivization. However, $\mathcal{R}$ is no longer such a

[7] In addition to managing to rigorously make sense of 1/0 of course!

[8] However, it is certainly a visually natural thing to try given our representation of rationals as quotients - indeed probably the most common mistake when first learning arithmetic is to add rationals by taking their mediant!

simple isomorphism, and much of the work involved in accurately transferring information from coefficients to roots involves a careful analysis of $\mathcal{R}$ and the symmetries it preserves.

4.2. **Hyperbolic geometry and quadratic numbers.** Some of the striking images in the gallery involve *quadratic numbers*, or solutions in $\mathbb{C}$ to degree-two polynomials with integer coefficients $ax^2 + bx + c$. The suggestively "hyperbolic" nature of these (Figure 1a) is no accident; and the goal of this section is to make this connection explicit. In particular, we prove the following.

Theorem 4.3. *Let $\mathbb{H}^2_{\text{Coefs}}$ be the projectivized set of coefficients of real quadratics with complex roots and $\mathbb{H}^2_{\text{Roots}}$ be the set of their root-sets, equipped with the following metrics:*

- $\mathbb{H}^2_{\text{Coefs}} = \{[a : b : c] \mid b^2 < 4ac\}$ *is given the projectively invariant metric it inherits as a convex subset of $\mathbb{RP}^2$,*
- $\mathbb{H}^2_{\text{Roots}} = \{\{x \pm iy\} \mid x, y \in \mathbb{R}, y > 0\}$ *is given the Möbius-transformation invariant metric from identification with the upper half plane $\subset \mathbb{CP}^1$.*

Both of these spaces are isometric to the hyperbolic plane. Furthermore, the quadratic formula, given as the map $[a : b : c] \mapsto \left\{\frac{-b \pm i\sqrt{4ac-b^2}}{2a}\right\}$ is an isometry between these two metrics.

This theorem is our main goal in Section 4.2; we will take some time to explain the metrics and invariances just referenced. Throughout we attempt to emphasize how how natural symmetry considerations might lead one to have conjectured this theorem in the first place.

4.2.1. *Complex Quadratics.* Irreducible quadratics over $\mathbb{R}$ have pairs of complex conjugate roots, which makes the complex numbers integral to our discussion. As such, we begin with a discussion of homogeneous quadratics over $\mathbb{C}$. This has several advantages[9], and after getting comfortable here we will restrict back to real coefficients to prove Theorem 4.3. In this section we precisely define the space of coefficients, space of roots, and the map $\mathcal{R}\colon \text{Coefs} \to \text{Roots}$ for complex homogeneous quadratics. This lays the foundation for a bridge between properties of quadratic numbers and properties of their minimal polynomials, by the following observation.

Observation 4.4. *The roots map $\mathcal{R}\colon \text{Coefs} \to \text{Roots}$ is a homeomorphism from the space of complex homogeneous quadratics to the multi-sets of their roots in $\mathbb{CP}^1$.*

The polynomial $f = ax^2 + bxy + cy^2$ is determined by its coefficients $a, b, c \in \mathbb{C}$, so $\text{Coefs} \cong \mathbb{C}^3$. We denote the scaling class of a quadratic $f = ax^2 + bxy + cy^2$ by $[f]$, with coefficients $[a : b : c] \in \mathbb{P}\text{Coefs} = \mathbb{P}(\mathbb{C}^3) = \mathbb{CP}^2$ in the complex projective plane. The roots map $\mathcal{R}$ is just the quadratic formula; taking the polynomial with coefficients $[a : b : c]$ with $a \neq 0$ to its solutions $[\frac{-b}{2a} \pm \frac{\sqrt{b^2-4ac}}{2a} : 1]$, or $[-b \pm \sqrt{b^2 - 4ac} : 2a]$ after clearing denominators[10]. The space of roots comes with no natural ordering. Thus, Roots does not identify with the space of *ordered pairs* $\mathbb{CP}^1 \times \mathbb{CP}^1$ of points in the extended complex plane, but rather the set of *unordered pairs in* $\mathbb{CP}^1$. This space is called the *second symmetric power of* $\mathbb{CP}^1$, and we write $\text{Roots} = \text{SP}^2(\mathbb{CP}^1)$. An explicit construction of this space results from the quotient of $\mathbb{CP}^1 \times \mathbb{CP}^1$ by the action of $\mathbb{Z}/2\mathbb{Z}$ swapping coordinates.

Remark 4.5. It may be helpful to pause here to gain some intuition from the lowest-dimensional example of a nontrivial symmetric power. The second symmetric power of a circle $\text{SP}^2(\mathbb{S}^1)$ is the quotient of a torus $\mathbb{S}^1 \times \mathbb{S}^1$ by the involution $(x, y) \mapsto (y, x)$ This fixes the diagonal (x, x), and the quotient is a Möbius strip with these points as the boundary curve (Figure 19).

[9] As $\mathbb{C}$ is algebraically closed, the space of roots is easy to describe, defining the roots map does not require passing to a field extension, and the $\text{PSL}(2; \mathbb{C})$ symmetry can be exhibited at its most natural level of generality.

[10] It is quick to check that as $r \to \infty$ the coefficients of the polynomial $(x - ry)(ax + by)$ converge projectively to $[0 : a : b]$ and the roots to $\{[1 : 0], [-b : a]\} = \{\infty, -b/a\}$. Thus the quadratic formula extends continuously to linear equations interpreted as 'quadratics with roots at infinity' as noted in Section 4.1.

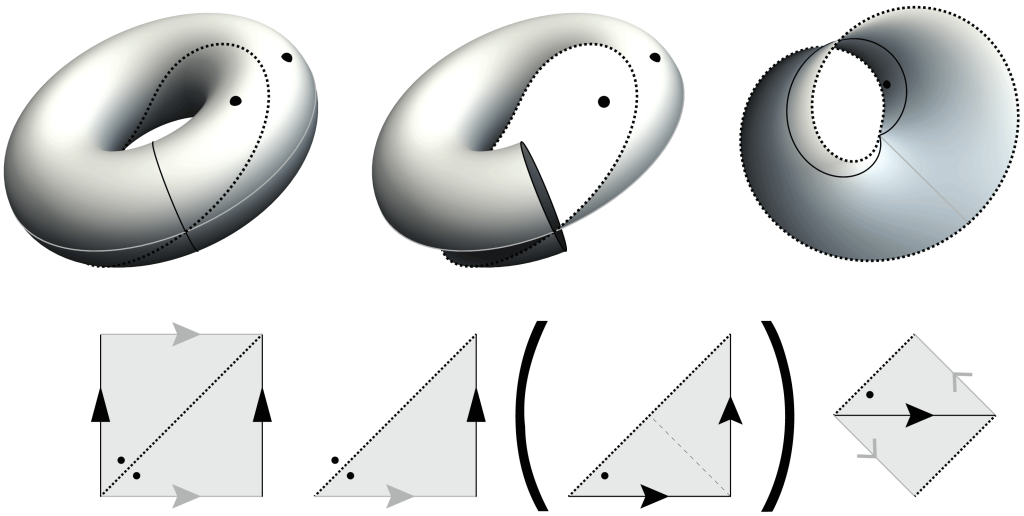


FIGURE 19. The space $\mathrm{SP}^2(\mathbb{RP}^1)$ of unordered pairs of points on the circle $\mathbb{RP}^1 \cong \mathbb{S}^1$ is homeomorphic to a Möbius band.

The space of quadratic polynomials parameterized by their roots is $\mathrm{SP}^2(\mathbb{CP}^1)$, a space built from the sphere in much the same way as the Möbius strip $\mathrm{SP}^2(\mathbb{S}^1)$ was built from the circle in Remark 4.5. See the text following Observation 4.4, for the discussion of this. Not only do every pair of points in $\mathbb{CP}^1$ determine a scaling class of quadratics, but as a consequence of the fundamental theorem of algebra, every scaling class has two roots (with multiplicity) determining it. Together with the continuity of $\mathcal{R}$, this implies that the roots map is a homeomorphism[11] from $\text{Coefs} = \mathbb{CP}^2$ to $\text{Roots} = \mathrm{SP}^2(\mathbb{CP}^1)$. Thus topological properties of (projectivized) collections of quadratic polynomials completely determine the topology of their collection of roots. To strengthen this connection, we next focus on natural symmetries of the space of polynomials which are preserved by the roots map. We will see these symmetries are actually *isometries* of natural choices of metrics on both the space of coefficients and roots, which will be instrumental in our proof of the main theorem, 4.3.

4.2.2. $\mathrm{PSL}(2;\mathbb{C})$ *Symmetry.* From both the roots and coefficients perspectives, quadratics are intimately tied to complex projective spaces. But these spaces are only half the story. To work geometrically, we must also describe their groups of allowable motions. As projective space $\mathbb{CP}^n$ is just scaling classes of vectors in $\mathbb{C}^{n+1}$, the most natural group of symmetries is the linear group $\mathrm{GL}(n+1;\mathbb{C})$ acting on these scaling classes via $A.[v] = [Av]$ for $A \in \mathrm{GL}(n;\mathbb{C})$, $[v] \in \mathbb{CP}^n$. For simplicity, we may consider these matrices up to constant multiples as well, and take the *projective special linear group* $\mathrm{PSL}(n+1;\mathbb{C}) = \{[A] \mid \det A = 1\}$ as the symmetries of $\mathbb{CP}^n$. For points in $\mathbb{CP}^1$ thought of as $\mathbb{C} \cup \{\infty\}$, these are known as *Möbius transformations*: the transformation corresponding to $\left(\begin{smallmatrix} p & q \\ r & s \end{smallmatrix}\right)$ acts on $z \in \mathbb{C}$ as

$$\begin{pmatrix} p & q \\ r & s \end{pmatrix}.z = \begin{bmatrix} p & q \\ r & s \end{bmatrix}.[z:1] = \left[\begin{pmatrix} p & q \\ r & s \end{pmatrix}\begin{pmatrix} z \\ 1 \end{pmatrix}\right] = \left[\begin{pmatrix} pz+q \\ rz+s \end{pmatrix}\right] = \left[\frac{pz+q}{rz+s} : 1\right] = \frac{pz+q}{rz+s}.$$

Thus, $\mathbb{P}\text{Coefs} \cong \mathbb{CP}^2$ naturally has the symmetries of $\mathrm{PSL}(3;\mathbb{C})$. As $\text{Roots} = (\mathbb{CP}^1 \times \mathbb{CP}^1)/\mathbb{Z}_2$ is built from the complex projective line, the natural symmetries of the space of roots are

[11] This is an important argument in its own right, for it shows topologically the symmetric power $\mathrm{SP}^2(\mathbb{S}^2)$ is the complex projective plane in disguise. In particular, it is a closed manifold. Compare this with Remark 4.5 where $\mathrm{SP}^2(\mathbb{S}^1)$ is a manifold with boundary. This generalizes to arbitrary degree n, and the roots map provides a homeomorphism $\mathcal{R}_n \colon \mathbb{CP}^n \xrightarrow{\sim} \mathrm{SP}^n(\mathbb{CP}^1)$.

$\mathrm{PSL}(2;\mathbb{C})$, inherited from $\mathbb{CP}^1$ via $[A].\{[z],[w]\} = \{[Az],[Aw]\}$. Initially these two notions of geometry are completely independent, constructed from our models of Roots and Coefs respectively. However, given that $\mathcal{R}\colon \text{Coefs} \to \text{Roots}$ is a homeomorphism, we may use it to compare the symmetries on one side to the other. The main observation of this section is that these two actions are actually *compatible*[12] with each other.

Observation 4.6. *Moving the space of roots by a (projective) linear transformation is represented on the space of coefficients by a (projective) linear transformation as well. That is, conjugation of this action by $\mathcal{R}^{-1}$ embeds $\mathrm{PSL}(2;\mathbb{C})$ into $\mathrm{PSL}(3;\mathbb{C})$.*

One may then directly convert any knowledge about this representation into geometric statements binding the spaces of roots and coefficients yet closer together. This $\mathrm{PSL}(2;\mathbb{C})$ action on the space of coefficients has a natural description: conjugation by the roots map simply declares that $[A] \in \mathrm{PSL}(2;\mathbb{C})$ send the polynomial with roots $\{[z],[w]\}$ to the polynomial with roots $\{[Az],[Aw]\}$. Writing this out explicitly, let $A = \left(\begin{smallmatrix} p & q \\ r & s \end{smallmatrix}\right) \in \mathrm{PSL}(2;\mathbb{C})$ and f be a homogeneous quadratic with coefficients $[f] = [a:b:c]$. Then $[A].[f] = [f \circ A^{-1}]$, which gives the following.

$$f\left(\begin{bmatrix} s & -q \\ -r & p \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}\right) = f\left(\begin{bmatrix} sx-qy \\ py-rx \end{bmatrix}\right) = a(sx-qy)^2 + b(sx-qy)(py-rx) + c(py-rx)^2$$

Expanding and collecting like terms in $x^i y^j$ shows that the new coefficient vector is *a linear transformation of* $[a:b:c]$ *involving* p,q,r,s, which confirms Observation 4.6. Specifically, the coefficients of $[A].[f]$ satisfy

$$\begin{bmatrix} p & q \\ r & s \end{bmatrix}.[a:b:c] = \left[\begin{pmatrix} s^2 & -rs & r^2 \\ -2qs & qr+ps & -2pr \\ q^2 & -pq & p^2 \end{pmatrix}\begin{pmatrix} a \\ b \\ c \end{pmatrix}\right]. \tag{1}$$

This 3×3 matrix is the *representation* of $\left(\begin{smallmatrix} p & q \\ r & s \end{smallmatrix}\right)$ acting on $\text{Coefs} = \mathbb{CP}^2$. Let $\rho\colon \mathrm{PSL}(2;\mathbb{C}) \to \mathrm{PSL}(3;\mathbb{C})$ denote this representation. By construction, for each $A \in \mathrm{PSL}(2;\mathbb{C})$ we have $f(A.[x:y]) = (\rho(A).f)([x:y])$. This implies the roots map is *ρ-equivariant*, satisfying the following important identity:

$$\mathcal{R}\left(\rho(A).f\right) = A.\mathcal{R}(f). \tag{2}$$

In principle, this formula for the action lets us completely compute anything we desire about the relationship between the spaces of roots and coefficients (including the quadratic formula itself, Corollary 4.10). This $\mathrm{PSL}(2;\mathbb{C})$ action divides the space of quadratics into two components: those with a double root, and (the generic case) those with distinct roots. Each of these components forms a singe $\mathrm{PSL}(2;\mathbb{C})$ orbit, which we may see as follows. For quadratics with a double root, note for any $a = [a_1:a_2] \in \mathbb{CP}^1$ the symmetry[13] $A = \left(\begin{smallmatrix} a_2 & -a_1 \\ a_1 & a_2 \end{smallmatrix}\right) \in \mathrm{PSL}(2;\mathbb{C})$ takes $f = x^2$ to the quadratic $(a_2x - a_1y)^2$ with double root a. More abstractly, one may deduce this for the component containing quadratics with distinct roots from the fact that $\mathrm{PSL}(2,\mathbb{C})$ acts freely and transitively on distinct triples in $\mathbb{CP}^1$ (see [53] for this and other useful facts in projective geometry). That is, given any quadratic f with roots $a,b \in \mathbb{CP}^1$, choosing any arbitrary $c \notin \{[0:1],[1:1],a,b\}$ there is a unique $A \in \mathrm{PSL}(2;\mathbb{C})$ taking (a,b,c) to $([0:1],[1:1])$, and thus taking f to the fixed quadratic $x(x-y)$. Because the action of $\mathrm{PSL}(2;\mathbb{C})$ is transitive on each of these components, they can be interpreted as homogeneous geometries in the sense of Klein. This allows us to use geometric properties to understand the behavior of the roots map.

[12] This picture may already be familiar from representation theory: the action of $\mathrm{PSL}(V)$ on a 2-dimensional complex vector space V induces actions on symmetric powers of V, which are the irreducible representations of $\mathrm{PSL}(V)$. Projectivizing this picture under an identification $V = \mathbb{C}^2$ yields the result we exposit here.

[13] indeed there are many choices for A: if $a \in \mathbb{C}$ then $A = \left(\begin{smallmatrix} 1 & a \\ 0 & 1 \end{smallmatrix}\right)$ also works.

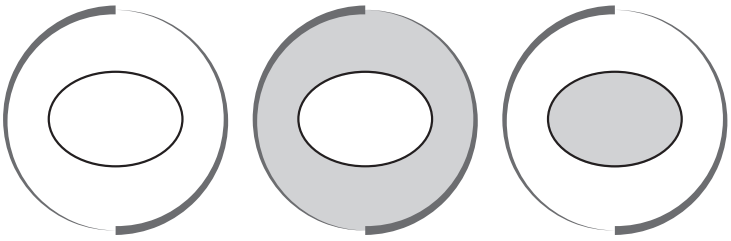
4.2.3. *Real Quadratics.* We now turn to quadratics with real coefficients. The spaces of real quadratics naturally inherit their topology and geometry as subsets of the corresponding spaces over $\mathbb{C}$. This reduces the space of coefficients from $\mathbb{C}^3$ to $\mathbb{R}^3$, so we work with the projective plane $\mathbb{P}\text{Coefs} = \mathbb{RP}^2$, and its symmetries $\text{PSL}(3;\mathbb{R})$. The space of roots is more complicated to describe due to the fact that real quadratics may have complex roots. However, as the roots map is a homeomorphism over $\mathbb{C}$, we can immediately determine its topology: $\text{Roots} = \mathcal{R}(\mathbb{P}\text{Coefs}) = \mathcal{R}(\mathbb{RP}^2) \cong \mathbb{RP}^2$. We will denote these two models of projective space as $\mathbb{RP}^2_{\text{Coefs}}$, $\mathbb{RP}^2_{\text{Roots}}$ in what follows.

$$\mathbb{RP}^2_{\text{Coefs}} = \left\{[a:b:c] \mid (a,b,c) \in \mathbb{R}^3 \backslash \mathbf{0}\right\} \subset \mathbb{CP}^3$$
$$\mathbb{RP}^2_{\text{Roots}} = \left\{\{x,y\} \mid x,y \in \mathbb{RP}^1\right\} \bigcup \left\{\{x \pm iy\} \mid x,y \in \mathbb{R}, y > 0\right\} \subset \text{SP}^2(\mathbb{CP}^1)$$

The natural symmetry group on the roots restricts to $\text{PSL}(2;\mathbb{R})$, and equation (1) confirms that the representation faithfully translates this to a subgroup of $\text{PSL}(3;\mathbb{R})$ on the coefficients. Thus $\mathcal{R}\colon \mathbb{RP}^2_{\text{Coefs}} \to \mathbb{RP}^2_{\text{Roots}}$ remains equivariant with respect to the $\text{PSL}(2;\mathbb{R})$ symmetries on each side. Recall that over $\mathbb{C}$, this action divided the space of quadratics into two orbits. Any two polynomials with distinct roots are related to one another by a symmetry transformation, as are any two with a double root; but $\text{PSL}(2;\mathbb{C})$ cannot convert one type into the other. Over $\mathbb{R}$ the story gets more interesting, as the generic case (polynomials with distinct roots) splits.

Observation 4.7. *The* $\text{PSL}(2;\mathbb{R})$ *action divides the space of real quadratics into three orbits*[14]*: (1) quadratics with a double real root, (2) quadratics with distinct real roots, and (3) quadratics with complex conjugate roots.*

The orbit (1) is homeomorphic to a circle[15] as any point $[p:q] \in \mathbb{RP}^1$ determines a quadratic $f = (px - qy)^2$ with double root at $[p:q]$. The orbit (2) of quadratics with distinct real roots identifies with the space of unordered pairs of points of $\mathbb{RP}^1$, which we saw in Remark 4.5 to be the Möbius band. We are mainly interested in orbit (3), which consists of unordered pairs of *complex conjugates* in $\mathbb{C}\backslash\mathbb{R}$. Each such pair contains a unique point $x + iy$ with $y > 0$, so we identify this with the upper half plane in $\mathbb{C}$. As the $\text{PSL}(2;\mathbb{R})$ action is transitive on each orbit, all three of these pieces inherit the structure of a homogeneous geometry, listed below.



(A) Double Root (B) 2 Real Roots (C) 2 Complex Conjugate Roots

FIGURE 20. The four $\text{PSL}(2;\mathbb{R})$? orbits on the $\mathbb{RP}^2$ of real quadratics, as subsets of $\mathbb{RP}^2$. These are the three homogeneous spaces of Observation 4.8 As in Figure 17, the topology of each component is recovered by identifying points on the boundary circle via the antipodal map. Compare to the analogous decomposition of $\mathbb{RP}^3$ for cubics in Figure 26.

[14] On $\mathbb{RP}^2_{\text{Coefs}}$, the orbit of quadratics with double root is exactly the discriminant locus (the set of polynomials whose discriminant is equal to zero); its complement is the union of the other two orbits.

[15] In fact, from here, elementary topology completes the story as all separating circles divide $\mathbb{RP}^2$ into a Möbius strip and a disk, giving the topological type of (2) and (3) respectively.

Observation 4.8. *The* $\mathrm{PSL}(2;\mathbb{R})$ *action on the projective plane of real quadratics divides it into three disjoint homogeneous geometries.*

(1) *The geometry of quadratics with a double root is the familiar geometry of the real projective line.*
(2) *The geometry of quadratics with distinct real roots is* de Sitter space*, a two dimensional geometry relevant to relativistic physics*[16].
(3) *The geometry of quadratics with complex conjugate roots is a hyperbolic plane (the disk in Figure 20c equipped with the projective action of* $\mathrm{PSL}(2;\mathbb{R})$.

4.2.4. *Hyperbolic Geometry and the Roots Map.* It is the real quadratics with complex roots which are responsible for some of the beautiful images in the gallery such as Figure 1a, so we study the hyperbolic geometry which underlies them[17]. An abstract understanding of hyperbolic space is insufficient for our goals, which rely on an explicit understanding of the geometry of the roots map $\mathcal{R}$. Thus we need to consider both the model $\mathbb{H}^2_{\text{Coefs}}$ of hyperbolic geometry given by the coefficients of these polynomials, and model $\mathbb{H}^2_{\text{Roots}}$ formed by their roots. For an excellent exposition of these models and more, see [18].

We begin with $\mathbb{H}^2_{\text{Roots}}$. Identifying this space with the upper half plane $\{x+iy \mid y>0\} \subset \mathbb{C}$, the Riemannian metric[18] for hyperbolic geometry is $ds^2 = (dx^2+dy^2)/y^2$, which follows directly from the $\mathrm{PSL}(2;\mathbb{R})$ action, translating the standard metric dx^2+dy^2 at i around by the group action.

The length of paths in the hyperbolic plane is computed via integrating this infinitesimal arc length: given a curve $\gamma(t) = (x(t), y(t))$ defined for $t \in [a,b]$, its length is $\text{Length}(\gamma) = \int_a^b \frac{x'(t)^2+y'(t)^2}{y(t)^2}dt$. The geodesics of hyperbolic geometry are given by arcs of semicircles whose centers lie on the boundary $y=0$, together with vertical lines (circles of infinite radius). These geodesics determine all length minimizing segments in the hyperbolic plane: if p, q are two points in hyperbolic space and γ is a geodesic passing through p and q, then the segment of γ connecting them achieves the minimum distance among all curves joining p to q. This allows us to compute explicitly the distance function on hyperbolic space: if x_1+iy_1 and x_2+iy_2 in $\mathbb{H}^2_{\text{Roots}}$, the distance between them is given by

$$d_{\text{Roots}}(x_1+iy_i, x_2+iy_2) = \operatorname{acosh}\left(1+\frac{(x_2-x_1)^2+(y_2-y_1)^2}{2y_1y_2}\right) \tag{3}$$

Hyperbolic geometry has a well-defined notion of an *ideal boundary*, consisting of points at infinity. For the upper half plane model, we may describe these points as the idealized endpoints of geodesics: if γ is any hyperbolic geodesic, the limits $\lim_{t\to\pm\infty}\gamma(t)$ lie on the ideal boundary. Concretely, this consists of all points on the real line $y=0$ (the endpoints of all semicircle geodesics, and one endpoint of each vertical geodesic) together with a single additional point traditionally labeled ∞ (denoting the idealized endpoint of all vertical geodesics not lying on the real line). See [3], Chapter 1 for more details on the upper half plane model and its ideal boundary.

Next, we turn to $\mathbb{H}^2_{\text{Coefs}}$. Here the basic geometry is likely familiar from the study of conic sections, often studied in high school mathematics. Generic real quadratics have either two real roots or a pair of complex conjugate roots, as determined by the discriminant $\Delta(ax^2+bx+c) =$

[16]De Sitter geometry is a particular geometry not of space, but rather of *spacetime*. In this particular case, De Sitter 2-space describes a world with one space and one time dimension of positive curvature. Geometrically, this is just the hyperboloid of 1-sheet $x^2+y^2-z^2=1$ in $\mathbb{R}^3$ equipped with the action of $\mathrm{SO}(2,1)$.

[17]Hyperbolic geometry, commonly denoted $\mathbb{H}^2$, is the unique two dimensional geometry with constant negative curvature, and was the first non euclidean geometry discovered. Negative curvature implies that $\mathbb{H}^2$ violates Euclid's fifth postulate with an infinitude of parallel lines to a given line through any point not on it. For an introductory treatment of the hyperbolic plane, see [3].

[18]A Riemannian metric is a choice of inner product for each tangent space, which allows one to measure the length of vectors, and hence the arc length of curves.

b^2-4ac being positive or negative, respectively. In fact, this and more can be explicitly recovered from studying the $\mathrm{PSL}(2;\mathbb{R})$ action given by the representation ρ on $\mathbb{RP}^2_{\text{Coefs}}$. We focus here on the polynomials with complex roots; similar reasoning applies in the other case. As $\mathrm{PSL}(2;\mathbb{R})$ acts transitively on $\mathbb{H}^2_{\text{Roots}}$, we may recover the entire space as the orbit of any point. Using the defining property $\mathcal{R}(\mathbb{H}^2_{\text{Coefs}}) = \mathbb{H}^2_{\text{Roots}}$ and equation (2), we see that

$$\mathbb{H}^2_{\text{Roots}} = \mathrm{PSL}(2;\mathbb{R}).\{\pm i\} \implies \mathbb{H}^2_{\text{Coefs}} = \rho\left(\mathrm{PSL}(2;\mathbb{R})\right).[1:0:1], \tag{4}$$

as $[1:0:1]$ are the coefficients of x^2+1, with roots $\{\pm i\}$. Computing this orbit[19], the hyperbolic space $\mathbb{H}^2_{\text{Coefs}}$ is the projectivization of the negative cone[20] of the discriminant Δ; that is $\mathbb{H}^2_{\text{Coefs}} = \{[a:b:c] \mid 4ac > b^2\}$. Topologically $\mathbb{H}^2_{\text{Coefs}}$ is a disk, but depending on the affine patch of $\mathbb{RP}^2$ that we choose it may take different forms. Indeed, the patch $[1:b:c]$, corresponding to taking monic representatives, represents $\mathbb{H}^2_{\text{Coefs}}$ as the interior of a paraboloid, with one point on its ideal boundary at infinity: $\mathbb{H}^2_{\text{Coefs}} = \{[1:b:c] \mid 4c > b^2\}$.

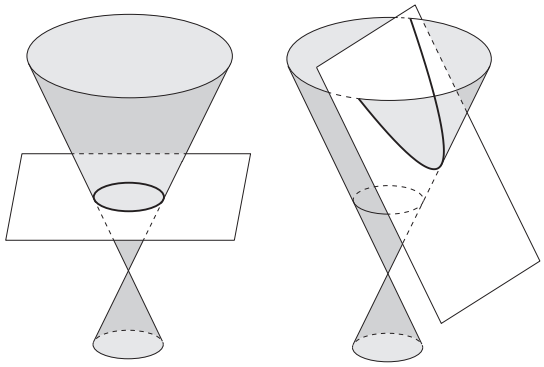


FIGURE 21. The cone determined by the discriminant in $\mathbb{R}^3$, together with two affine patches giving the usual Klein model (left) and a parabola model (right) of $\mathbb{H}^2$.

As $\mathbb{H}^2_{\text{Coefs}}$ is a properly convex set in $\mathbb{RP}^2$, it can be endowed with a natural metric invariant under projective transformations[21]. See [18] for a more detailed analysis of this projective, or *Klein Model* of the hyperbolic plane.

As the discriminant carries all the geometry associated with the $\mathrm{PSL}(2;\mathbb{R})$ symmetries, there is a nice description of this metric in terms of Δ. This is most apparent for the Riemannian metric after pulling back to the hyperboloid in $\mathbb{R}^3$: if $v \in \mathbb{R}^3$ is any tangent vector to the

[19]The entire action of $\mathrm{PSL}(2;\mathbb{R})$ on $\mathbb{R}^3$ via the representation ρ preserves the quadratic form Δ, and the inner product $\langle (a_1,b_1,c_1),(a_2,b_2,c_2)\rangle = b_1b_2 - 2a_1c_2 - 2a_2c_1$ for which $\Delta(v) = \langle v,v\rangle$. Thus the symmetries of $\mathbb{H}^2_{\text{Coefs}}$ are contained in the *special orthogonal group* of this form $\mathrm{SO}(\Delta)$, the orthogonal group of a quadratic form q is the group of all matrices with determinant 1 whose action leaves q invariant: $\mathrm{SO}(q) = \{A \in \mathrm{SL}(n;\mathbb{R}) \mid q(v,w) = q(Av,Aw)\}$. The indefinite orthogonal group $\mathrm{SO}(\Delta)$ has two components, determined by whether or not a symmetry preserves or swaps the two sheets of the hyperboloid. The representation ρ is actually an isomorphism onto the connected component of the identity: $\mathrm{PSL}(2;\mathbb{R}) \cong \mathrm{SO}_0(\Delta)$.

[20]A similar description can be given for the two other geometries of quadratic polynomials. The $\mathbb{RP}^1$ of polynomials with double roots is the zero set of Δ, or the projectivization of the light cone $b^2 = 4ac$. The space of polynomials with distinct real roots corresponds to points on which Δ is positive, which projectively forms a Möbius band. This space also has a natural notion of geometry, coming not from the hyperbolic plane but from relativity (it is called $1+1$ dimensional de Sitter space, but is beyond the scope of this paper.)

[21]This is called the Hilbert metric. Such a metric may be defined for any convex subset $\Omega \subset \mathbb{RP}^2$ not containing any entire projective line, and realizes a model of hyperbolic geometry precisely when Ω is bounded by a nonsingular conic section.

hyperboloid, than its infinitesimal arc length is simply

$$ds^2 = \Delta(v).$$

The geodesics in this metric are straight line segments in any affine patch containing $\mathbb{H}^2_{\text{Coefs}}$, and the distance between two points $f_1 = [a_1 : b_1 : c_1]$ and $f_2 = [a_2 : b_2 : c_2]$ is given by

$$d_{\text{Coefs}}(f_1, f_2) = \operatorname{acosh}\left(\frac{-\langle f_1, f_2\rangle}{\sqrt{\Delta(f_1)\Delta(f_2)}}\right) = \operatorname{acosh}\left(\frac{2a_1c_2 + 2a_2c_1 - b_1b_2}{\sqrt{(4a_1c_1 - b_1^2)(4a_2c_2 - b_2^2)}}\right) \tag{5}$$

Bringing this all together, we have seen the natural $\mathrm{PSL}(2;\mathbb{R})$ actions on both the spaces of roots and coefficients endow the space of real quadratics with complex conjugate roots with the homogeneous geometry of the hyperbolic plane. We are now in a position to prove the main theorem of this section, following the outline proposed at the beginning.

Theorem 4.9. *Let $\mathbb{H}^2_{\text{Coefs}} = \{[a : b : c] \in \mathbb{RP}^2 \mid 4ac > b^2\}$ be equipped with the projectively invariant metric (equation 5) it inherits as a convex subset of $\mathbb{RP}^2$, and $\mathbb{H}^2_{\text{Roots}} = \{\{x \pm iy\} \mid x, y \in \mathbb{R}, y > 0\}$ be equipped with the hyperbolic metric (equation 3) arising from its identification with the upper half plane. Then the restricted roots map $\mathcal{R}\colon \mathbb{H}^2_{\text{Coefs}} \to \mathbb{H}^2_{\text{Roots}}$ is an isometry.*

Proof. Observation 4.4 implies the roots map is a homeomorphism on the total space of projectivized quadratics and their roots, so restricting to polynomials with a complex conjugate pair of roots, $\mathcal{R}$ remains a homeomorphism from $\mathbb{H}^2_{\text{Coefs}}$ to $\mathbb{H}^2_{\text{Roots}}$. Equipped with their respective metrics d_{Coefs} and d_{Roots} referenced in the theorem statement, we show that $\mathcal{R}$ is an isometry by proving for each pair of quadratics f, g with complex roots,

$$d_{\text{Coefs}}(f, g) = d_{\text{Roots}}(\mathcal{R}(f), \mathcal{R}(g)).$$

Fix such an f and g. Because hyperbolic geometry is homogeneous, there is some isometry A of $\mathbb{H}^2_{\text{Coefs}}$ which takes f to any point of our choosing. To leverage this symmetry in our computations, we choose A to be the isometry taking f to $[1 : 0 : 1]$, which represents the polynomial $x^2 + 1$. But furthermore the hyperbolic plane is isotropic (looks the same in every direction); consequently we may find another isometry B which fixes $Af = [1 : 0 : 1]$ and rotates about it, taking $A.g$ to a point of the form $[1 : 0 : a^2]$ lying on the geodesic $[1 : 0 : t]$ through $[1 : 0 : 1]$. As $C = BA$ is an isometry it leaves distances invariant, and so

$$d_{\text{Coefs}}(f, g) = d_{\text{Coefs}}(C.f, C.g) = d_{\text{Coefs}}\left([1 : 0 : 1], [1 : 0 : a^2]\right)$$

This final quantity is straightforwward to compute directly from the definition of the distance function:

$$d_{\text{Coefs}}\left([1 : 0 : 1], [1 : 0 : a^2]\right) = \operatorname{acosh}\left(\frac{2 + 2a^2 - 0}{\sqrt{(-4)(-4a^2)}}\right) = \operatorname{acosh}\left(\frac{1}{2}\left(a + \frac{1}{a}\right)\right) \tag{6}$$

Now, we turn to the computation of $d_{\text{Roots}}(\mathcal{R}(f), \mathcal{R}(g))$. The symmetry C we leveraged in Equation 6 is a linear transformation preserving the hyperboloid $\Delta(f) = -1$ in $\mathbb{R}^3$, and hence C lies in the image of the representation ρ, so we may write $C = \rho(M)$ for some $M \in \mathrm{PSL}(2;\mathbb{R})$. Now, we can use the equivariance of the root map $\mathcal{R}$ to simplify things:

$$d_{\text{Roots}}\left(\mathcal{R}(f), \mathcal{R}(g)\right) = d_{\text{Roots}}\left(M.\mathcal{R}(f), M.\mathcal{R}(g)\right) = d_{\text{Roots}}\left(\mathcal{R}\left(\rho(M).f\right), \mathcal{R}\left(\rho(M).g\right)\right)$$

where the first equality follows as M is an isometry of d_{Roots}, and the second is equivariance (Equation 2). But this is just the distance between the roots of $C.f = x^2 + 1$ and $C.g = x^2 + a^2$; which is the length of the hyperbolic geodesic connecting i to ia in the upper half plane. Using the expression for d_{Roots} in Equation 3, we see

$$d_{\text{Roots}}(i, ia) = \operatorname{acosh}\left(1 + \frac{(0 - 0)^2 + (1 - a)^2}{2(1)(a)}\right) = \operatorname{acosh}\left(\frac{1}{2}\left(a + \frac{1}{a}\right)\right) \tag{7}$$

Putting these two computations together, we see that for any f, g the distance $d_{\text{Coefs}}(f, g)$ in the domain is equal to the distance $d_{\text{Roots}}(\mathcal{R}(f), \mathcal{R}(g))$ in the range, so $\mathcal{R}$ is an isometry as claimed.

□

We may use this to derive the quadratic formula from hyperbolic geometry. Fix any quadratic f (say, $x^2 + 1$) with complex root r (here i) in the upper half plane, and let $A \in \mathrm{PSL}(2;\mathbb{R})$. By equivariance, the polynomial with root $A.r$ is $\rho(A).f$ for $\rho\colon \mathrm{PSL}(2;\mathbb{R}) \to \mathrm{PSL}(3;\mathbb{R})$ the representation from Equation 1. Writing this out, we see the polynomial with roots $x \pm iy = \left(\begin{smallmatrix} y & x \\ 0 & 1 \end{smallmatrix}\right).\{\pm i\}$ has coefficients $\rho\left(\left(\begin{smallmatrix} y & x \\ 0 & 1 \end{smallmatrix}\right)\right).[1 : 0 : 1] = [1 : -2x : x^2 + y^2]$. This is an explict formula for the inverse of the roots map, $\mathcal{R}^{-1}\colon \mathbb{H}^2_{\text{Roots}} \to \mathbb{H}^2_{\text{Coefs}}$, sending the roots $\{x \pm iy\}$ to the polynomial with (projectivized) coefficients $[1 : -2x : x^2 + y^2]$. Inverting the relation $[1 : -2x : x^2 + y^2] = [a : b : c]$ gives $x = -b/2a$ and $y = \sqrt{c/a - x^2}$, or

$$x + iy = \frac{-b}{2a} + i\frac{\sqrt{4ac - b^2}}{2a}.$$

Theorem 4.9 allows us to compute any geometric quantity of interest using either the roots or coefficients. This simplifies certain calculations. In particular, if $\alpha, \beta \in \mathbb{C}$ are complex roots of the real quadratics f_α, f_β respectively, we may avoid using equation (3) to compute $d_{\text{Roots}}(\alpha, \beta)$, and instead compute $d_{\text{Coefs}}(f_\alpha, f_\beta)$ using equation (5). This is used for the results in Section 6.

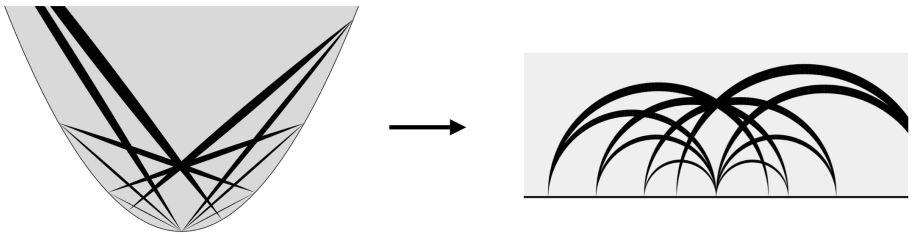


FIGURE 22. Isometry from the projective model $\{[a : b : c] \mid 4ac > b^2\}$ and the upper half plane model of the hyperbolic plane, given by the quadratic formula.

Corollary 4.10. *After the projective change of coordinates $\mathbb{RP}^2 \to \mathbb{RP}^2$ given by $[a : b : c] = [\frac{w+u}{2} : v : \frac{w-u}{2}]$, the quadratic formula is precisely the usual isometry from the Klein disk model of $\mathbb{H}^2$ to the upper half plane model,*

$$[u : v : 1] \mapsto \frac{-v \pm i\sqrt{1 - u^2 - v^2}}{1 + u}.$$

Qualitatively, this provides a complete understanding of the roots map[22]: affine lines in the space of coefficients are geodesics in the projective model of $\mathbb{H}^2$, and so the roots of such a 1-parameter family of polynomials form a geodesic in the upper half plane model: generalized circles orthogonal to $\mathbb{R} \subset \mathbb{C}$.

There is also a very nice geometric interpretation of the roots map for quadratics with two real roots, a patch of which is visualized in Figure 9. We do not give many details here, as these quadratics do not occur in the starscape images. Nonetheless, we cannot resist telling the beginning of the story.

[22]As particular examples, the 1-parameter families of quadratics with coefficients $[1 : 0 : t]$, and $[1 : t : 1]$ project under the roots map to the vertical geodesic and unit circle through $i \in \mathbb{C}$ respectively.

Observation 4.11. *The roots map for quadratics with a double root is the continuous extension of* $\mathcal{R}\colon \mathbb{H}^2_{\text{Coefs}} \to \mathbb{H}^2_{\text{Roots}}$ *to the ideal boundary of the hyperbolic plane. For polynomials of positive discriminant, the roots map decomposes geometrically as follows:*

(1) *Send the polynomial with coefficient vector* $v = [a : b : c]$*, thought of as a point in the Möbius band* $\mathbb{RP}^2 \backslash \mathbb{H}^2_{\text{Coefs}}$*, to the set* $\{w_1, w_2\} \in \partial_\infty \mathbb{H}^2$*, where the line through* v *and* w_i *is tangent to the ideal boundary. See figure 23.*

(2) *Follow by applying to each* w_i *the homeomorphism* $\partial_\infty \mathbb{H}^2_{\text{Coefs}} \to \partial_\infty \mathbb{H}^2_{\text{Roots}}$ *sending the projectivized lightcone to the extended real line in* $\mathbb{CP}^1$*. The resulting two points are the roots of* $f(x) = ax^2 + bx + c$*.*

We may reduce the proof of this observation to checking its truth at a single point using two facts: the transitivity of this $\mathrm{PSL}(2;\mathbb{R})$ action (via ρ) on the space of quadratics with positive discriminant, and the fact that as linear linear transformations preserving the lightcone, $\rho(A)$ preserves the colledtion of tangent lines to the ideal boundary for each $A \in \mathrm{PSL}(2;\mathbb{R})$. Choosing a point at which to verify the assertion: note the polynomial $f(x) = (x+1)(x-1)$ has coefficients $v = [1 : 0 : -1]$ and roots $\{1, -1\} \in \mathbb{R}$. These roots are identified with the polynomials $(x \pm 1)^2$ on the lightcone, so $w_i = [1 : \pm 2 : 1]$, and the lines through v and w_i are easily verified to be tangent to the discriminant locus, as claimed.

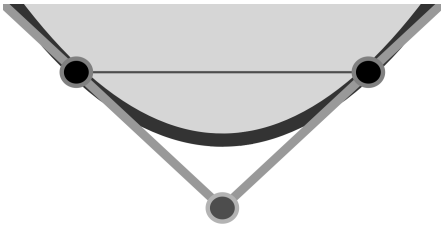


FIGURE 23. The roots map for quadratics with two real roots associates a point exterior to the hyperbolic plane in $\mathbb{RP}^2_{\text{Coefs}}$ to its two points of tangency with $\partial_\infty \mathbb{H}^2_{\text{Coefs}}$.

4.2.5. *Applications to quadratic algebraic numbers.* We now can apply this to the original case of interest: quadratic algebraic numbers and integer quadratic polynomials. The integer polynomials form a $\mathbb{Z}^3$ lattice in the space $\mathbb{R}^3$ of coefficients, and its image in $\mathbb{RP}^2$ can be interpreted as what it would look like to see the integer lattice from the origin (much like we saw for $\mathbb{QP}^1$ and $\mathbb{Z}^2$ in Figure 16). The quadratics of interest lie inside of the cone cut out by the discriminant, with planes through the origin projecting to lines in the disk $\mathbb{H}^2_{\text{Coefs}}$. And as the roots map realizes an isometry onto the upper half plane model (for specificity, by selecting the root in the complex conjugate pair with positive imaginary part), we know the image of these geodesics are also geodesics - here represented by circles orthogonal to the boundary. This explains the small scale patterns visible everywhere in the picture of the integer quadratic numbers - they are just the perspective view of a cubic lattice, distorted by the isometry taking the Klein model to the upper half plane model.

The natural $\mathrm{PSL}(2;\mathbb{C})$ action on roots and coefficients does not preserve the $\mathbb{Z}^3$ lattice. The subgroup which does is isomorphic to $\mathrm{PSL}(2;\mathbb{Z})$ (it is the intersection of $\rho(\mathrm{PSL}(2;\mathbb{C}))$ with $\mathrm{PSL}(3;\mathbb{Z})$, and it will play an important role in Sections 5 and 6).

The two-dimensional sublattices of $\mathbb{Z}^3$ will play a special role in what is to come. The planes such a sublattice can span are exactly planes with rational normal vector. These project to lines in $\mathbb{H}^2_{\text{Coefs}}$ which we will call *rational geodesics*. These are the dominant features in Figure 1a. There are a few important facts to collect about rational geodesics; these are just immediate consequences of the geometry.

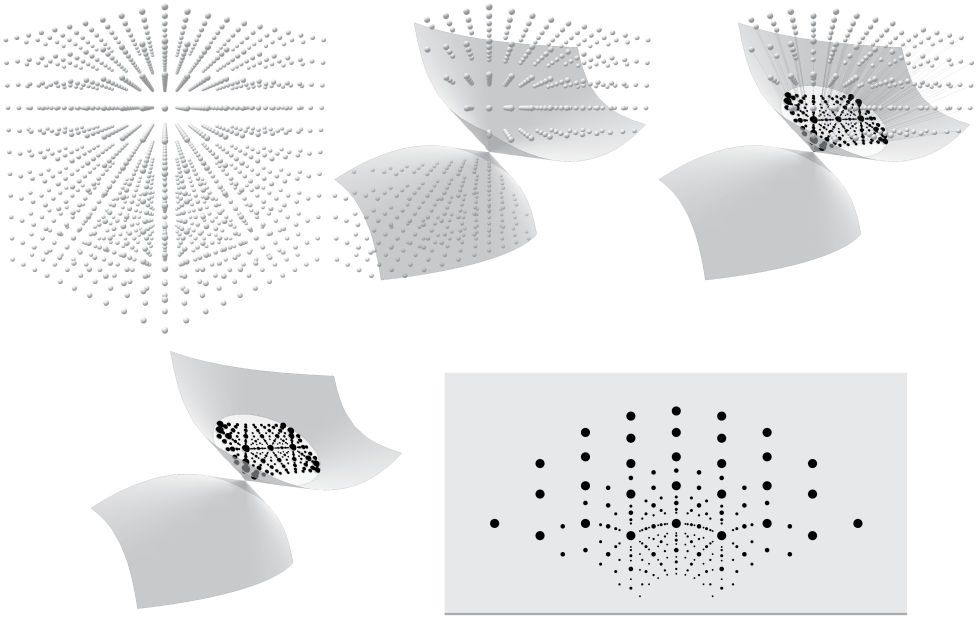


FIGURE 24. The lattice of integer quadratic polynomials, and its image under the roots map, decomposed as a sequence of geometric steps, from top left to bottom right. The lattice points lying within the light cone of the discriminant represent quadratics with a complex conjugate roots. Their projectivization is a collection of points in a projective model of the hyperbolic plane, and the roots map is an isometry onto the upper half plane model. See Theorem 4.9.

Observation 4.12. (1) *For $z \in \mathbb{C}$, its corresponding point $\{z, \overline{z}\}$ in root space pulls back to $[1 : -z - \overline{z} : z\overline{z}]$, and so z lies on a rational geodesic if and only if 1, $z + \overline{z}$ and $z\overline{z}$ are $\mathbb{Q}$-linearly dependent.*

(2) *Any two quadratic irrationalities $z, w \in \mathbb{C}$ share a unique rational geodesic, since any two points in $\mathbb{RP}^2$ determine a unique line. Any quadratic irrationality lies on infinitely many rational geodesics, and any two rational geodesics intersect at a quadratic irrationality.*

(3) *The images of rational geodesics under the roots map are exactly the hyperbolic geodesics of the upper half plane given by the upper half circles centred on a rational number, whose radius squared is rational. In other words, the limit points form a conjugate pair of points in a real quadratic field, or a pair of rational points.*

(4) *The group* $\mathrm{PSL}(2;\mathbb{Z})$ *acts as change of variables on the quadratic form associated to the geodesic*[23].

The geometric description of the roots map for quadratics with two real roots (Observation 4.11) has a nice interpretation for geodesics. Realizing such a geodesic γ as the intersection of a plane $\mathcal{S}$ with the cone of positive discriminant, denote the normal to $\mathcal{S}$ by n (well defined up to scaling; lying outside the lightcone) and the endpoints of γ by x, y on the ideal boundary of $\mathbb{H}^2_{\mathrm{Coefs}}$ (the projectivized lightcone of the discriminant). In the projective model one may

[23]In this way the orbits of rational geodesics under $\mathrm{PSL}(2;\mathbb{Z})$ are identified with the narrow ideal classes of real quadratic fields K. See [21, Section B.7].

recover these endpoints x, y (and hence the geodesic γ itself) directly from the normal n via a purely geometric construction. There are precisely two planes containing the line defining n which are tangent to the lightcone, and projectivizing these lines of tangency gives the ideal endpoints $\{x, y\}$ of the geodesic associated to n. But by Observation 4.11, the map sending n to its two points of tangency with $\partial_\infty \mathbb{H}^2$, followed by the isometry from the projective to upper half plane model of $\mathbb{H}^2$ is none other than the roots map on quadratics of positive discriminant. Thus, if we think of a geodesic as determined by its normal vector $n = (n_2, n_1, n_0)$ (often computationally an attractive thing to do), we may directly recover[24] the geodesic from n as its endpoints are precisely the roots of $n_2x^2 + n_1x + n_0 = 0$. This[25] is visible in Figure 23.

4.3. **Beyond discworld: the geometry of cubics.** Similarly to the quadratic case, we begin with the space of all complex homogeneous polynomials together with the $\mathrm{PSL}(2;\mathbb{C})$ action arising from precomposition with Möbius transformations. Cubics are the highest degree[26] where this action is enough to equip the various open subsets of generic cubics (components of the complement of the discriminant locus) with the structure of homogeneous geometries. Because $\mathrm{PSL}(2;\mathbb{R})$ is so tightly linked to $\mathbb{H}^2$, the hyperbolic plane remains a prominent actor in this story. Indeed, restricting to real cubics with negative discriminant, the complex root determines a point in $\mathbb{H}^2$ and the real root a direction - identifying this geometry as the unit tangent bundle[27] to the hyperbolic plane. Both the spaces of coefficients and roots form models of this geometry, related by the roots map, resulting in an analogous theorem to Theorem 4.3.

Theorem 4.13. *Let* $\{[a : b : c : d] \mid \Delta_3(a, b, c, d) < 0\}$ *be the set of real cubics with exactly one real root, where* Δ_3 *is the discriminant for cubics, and* $\{\{r, z, \overline{z}\} \mid r \in \mathbb{RP}^1, z \in \mathbb{C}\backslash\mathbb{R}\}$ *be the set of their root-sets. Each of these spaces admits a natural* $\mathrm{PSL}(2;\mathbb{R})$ *action (the former by precomposing the polynomial with a linear transformation, the latter by applying a Möbius transformation to each root). Finally, equipped with these actions, each of these spaces is isomorphic to the unit tangent bundle to the hyperbolic plane.*

We will prove this theorem in two pieces, Proposition 4.17 and Corollary 4.18. Beyond this, we will see that there is a natural way to equip each of these spaces with a Riemannian metric and with respect to these metrics, the roots map $\mathcal{R}$ is actually an *isometry* (exactly analogous to the quadratic case). After developing the necessary pieces, this is stated precisely and proven in Theorem 4.26. Utilizing this geometry both on the spaces of roots and coefficients provides a geometric description of the cubic formula, which we again pre-emptively state here. We state and prove the full version in Theorem 4.27.

Theorem 4.14. *On the set* $X \subset \mathbb{P}\mathrm{Coefs}$ *of polynomials which have exactly one real root, we define the* complex-root-map $\mathcal{R}_\mathbb{C}$ *as follows. For* $f \in X$ *a polynomial, let* $\mathcal{R}_\mathbb{C}(f) = z$ *be the unique complex root of* f *with positive imaginary part. Then, equipping domain with the*

[24] Topologically, this describes a map which takes the projectivization of the exterior of the lightcone (a Möbius band) to the set of unordered pairs of distinct points on the circle (or $\mathrm{SP}^2(\mathbb{S}^1)$, which is also a Möbius band, as depicted in Figure 19).

[25] Geometrically one may tell a beautiful story here quite analogous to Theorem 4.9, where the roots map is an isometry between a pair of Lorentzian metrics defined on each of these Möbius bands, though investigation of this would take us too far afield from the goals of this paper.

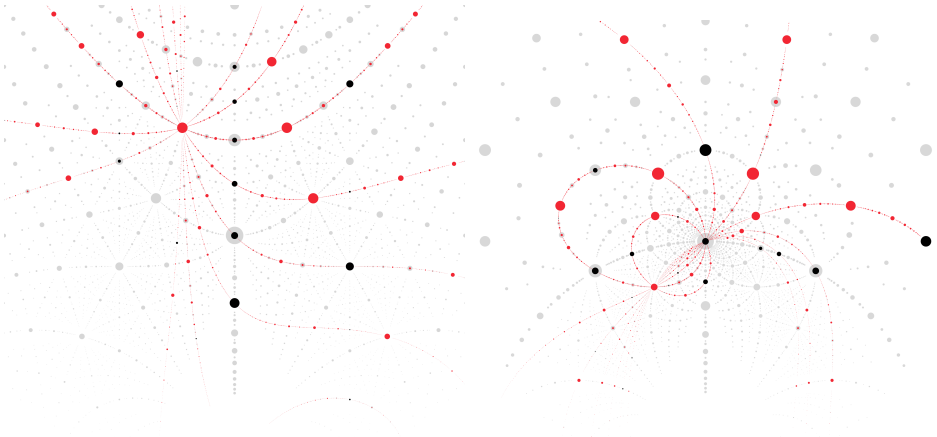
[26] This action has finitely many orbits as $\mathrm{PSL}(2;\mathbb{C})$ acts simply transitively on ordered triples of distinct elements of $\mathbb{CP}^1$. Thus, some of these orbits are open. In higher degree, there are a continuum of orbits, which are parameterized by the moduli of n possibly indistinct points in $\mathbb{CP}^1$ up to projective transformations (see also Remark 4.16).

[27] Given a manifold X, the set of all tangent vectors to X at a point p is called the *tangent space* to X at p. You can think of this as a 'linear approximation' to X near that point. If we restrict our attention to only unit vectors, we define the *unit tangent space* at p (for 2-dimensional geometries X, the unit tangent space at every point is just a circle). Collecting all the tangent spaces for every point of X gives the *tangent bundle* TX to X, and collecting only all unit vectors gives the *unit tangent bundle* $\mathrm{UT}X$. For an introduction smooth manifolds and their tangent bundles, see [38].

geometry of the hyperbolic plane's unit tangent bundle, the map $\mathcal{R}_{\mathbb{C}}$ *factors geometrically as the projection* $\mathrm{UT}\mathbb{H}^2 \to \mathbb{H}^2$ *onto the projective model* $\mathbb{H}^2_{\mathrm{Coefs}}$ *of the hyperbolic plane, followed by the isometry* $\mathbb{H}^2_{\mathrm{Coefs}} \to \mathbb{H}^2_{\mathrm{Roots}} \subset \mathbb{C}$ *of Theorem 4.9.*

$$\begin{array}{ccccc} & & \overset{\mathcal{R}_{\mathbb{C}}}{\frown} & & \\ \mathrm{Coefs} & \overset{\cong}{\to} & \mathrm{UT}\mathbb{H}^2 & & \\ & & \downarrow{\scriptstyle \pi} & & \\ & & \mathbb{H}^2_{\mathrm{Coefs}} & \xrightarrow{\mathcal{R}_2} \mathbb{H}^2_{\mathrm{Roots}} & \hookrightarrow \mathbb{C} \end{array}$$

Said briefly, the map sending a real cubic to its complex root in the upper half plane used to draw cubic starscapes is topologically conjugate to the projection $\mathrm{UT}\mathbb{H}^2 \to \mathbb{H}^2$ defining the hyperbolic plane's unit tangent bundle. We may use this geometry to understand some of the interesting interesting 1 and 2 dimensional families of cubic polynomials, producing linear and planar starscapes respectively. Some such linear starscapes are highlighted in the planar starscapes below (Figure 25). In particular, this gives a simple condition for when a starscape (thought of as a projective subspace of the set of cubics) embeds in the complex plane under the projection onto the complex root (Figure 25a) and when it is singular, collapsing some curve to a point (Figure 25b).



(A) (B)

FIGURE 25. Some linear starscapes, shown in red and black, within planar starscapes, the other points are shown in grey. The black points are quadratic roots, where the real root is rational. In (25a) some of the projective lines (corresponding to 2d sublattices of the lattice of coefficients) through the root of $x^3 - 2x^2 + 1$ are plotted in the depressed cubic starscape (also shown in Figure 2a). In (25b) some of the lines through the root of $x^3 - x^2 - 1$ are plotted in the starscape for the family $ax^3 + cx^2 + bx + c$ (also shown in Figure 1b). In the latter, these lines all also pass through i, which is the projection of a fibre $(dx + e)(x^2 + 1) = dx^3 + ex^2 + dx + e$. This can be seen in a different way on the right in Figure 32b.

Corollary 4.15. *Let* $\mathcal{S}$ *be a one or two dimensional affine space of real cubics in the space of projectivized coefficients, and* $\mathcal{R}_{\mathbb{C}}|_{\mathcal{S}} \colon \mathcal{S} \to \mathbb{C}$ *the projection onto the complex root. Under*

the geometric identification of this space with the unit tangent bundle of the hyperbolic plane (Theorem 4.13), the space of coefficients is foilated by simple closed curves (fibers of the unit tangent bundle). Then $\mathcal{R}_{\mathbb{C}}|_{\mathcal{S}}$ is an embedding if and only if $\mathcal{S}$ is everywhere transverse to this fibration.

For this failure of embedding to actually be *visible* in a cubic starscape, it must occur at some cubic or quadratic number, and accounting for this gives the more refined statement of Corollary 4.30. Quadratics are too low-degree for any interesting analogs of this behavior to occur[28], but it persists in higher degree (Figure 2), making cubics an important testing ground. We fill in the details of this picture below.

4.3.1. *General Complex and Real Coefficients.* A homogeneous cubic in two variables has the form $ax^3 + bx^2y + cxy^2 + dy^3$, so the sets of coefficients of all such cubics naturally identifies with $\mathbb{C}^4$. As previously, we are mostly concerned with cubics only up to a global scaling, and so take the space of projectivized coefficients $\mathbb{P}\text{Coefs} = \mathbb{CP}^3$, together with the natural action of $\mathrm{PSL}(4;\mathbb{C})$ as our starting point.

From the perspective of their solutions, cubics may be identified with unordered triples of (possibly coincident) points in the extended complex plane, so $\text{Roots} = \mathrm{SP}^3(\mathbb{CP}^1)$ is the third symmetric power of the sphere. The symmetric power admits a natural action of $\mathrm{PSL}(2;\mathbb{C})$ coming directly from its usual action on $\mathbb{CP}^1$ by Möbius transformations. More precisely, the action of $A = \left(\begin{smallmatrix} p & q \\ r & s \end{smallmatrix}\right)$ on the triple of points $\{z_1, z_2, z_3\}$ is

$$A.\{z_1, z_2, z_3\} = \left\{ \frac{pz_1 + q}{rz_1 + s}, \frac{pz_2 + q}{rz_2 + s}, \frac{pz_3 + q}{rz_3 + s} \right\}. \tag{8}$$

Again, the roots map $\mathcal{R}\colon \mathbb{P}\text{Coefs} \to \text{Roots}$ realizes a homeomorphism from $\mathbb{P}\text{Coefs} = \mathbb{CP}^3$ to $\text{Roots} = \mathrm{SP}^3(\mathbb{CP}^1)$. Conjugating the $\mathrm{PSL}(2;\mathbb{C})$ action on Roots by $\mathcal{R}$ gives an action on $\mathbb{P}\text{Coefs}$, which is compatible with its natural $\mathrm{PSL}(4;\mathbb{C})$ action: it is the projectivization of the representation $\mathrm{SL}(2;\mathbb{C}) \to \mathrm{SL}(4;\mathbb{C})$. As in Section 4.2.2, the explicit formula for this representation is defined by $\rho(A).f(x) = f(A^{-1}.x)$:

$$\begin{pmatrix} p & q \\ r & s \end{pmatrix} \xrightarrow{\rho} \begin{pmatrix} s^3 & -rs^2 & r^2s & -r^3 \\ -3qs^2 & 2qrs + ps^2 & -qr^2 - 2prs & 3pr^2 \\ 3q^2s & -q^2r - 2pqs & 2pqr + p^2s & -3p^2r \\ q^3 & pq^2 & -p^2q & p^3 \end{pmatrix} \tag{9}$$

Recall the $\mathrm{PSL}(2;\mathbb{C})$ action on the plane acts triply transitively (see [53] for an introduction to real and complex projective geometry). This implies that the space of cubics divides up into three distinct orbits under this action: cubics with three distinct roots, cubics with a pair of coincident roots, and cubics with a triple root. As each component is an orbit of the action, $\mathrm{PSL}(2;\mathbb{C})$ acts transitively on each.

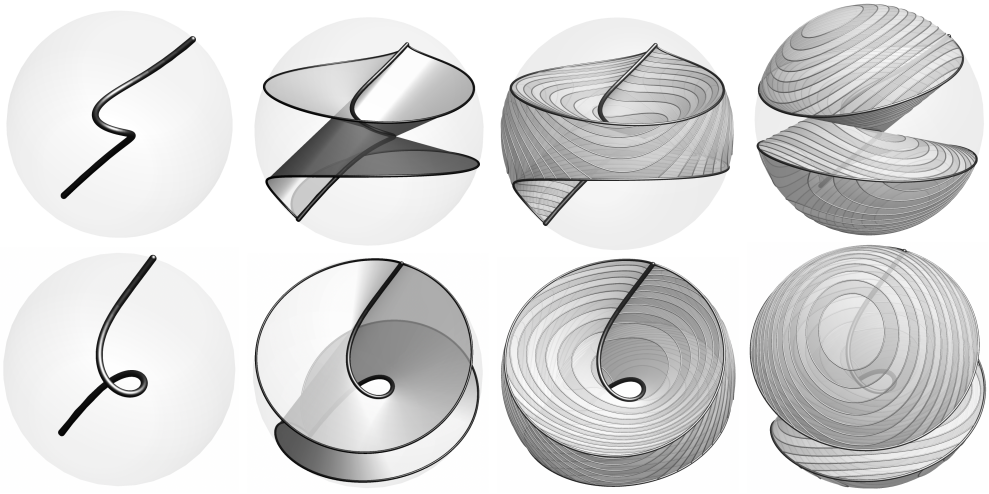
Remark 4.16. Here the (real) dimension of the space of homogeneous cubics is 6, which coincides with the dimension of $\mathrm{PSL}(2;\mathbb{C})$. Thus, cubics are the last dimension on which this $\mathrm{PSL}(2;\mathbb{C})$ action remains transitive on the open subset of generic cubics, giving it the structure of a $\mathrm{PSL}(2;\mathbb{C})$ homogeneous space.

Restricting to real cubics replaces the space of coefficients $\mathbb{CP}^3$ with $\mathbb{RP}^3$, and its image under the roots map is an embedding of real projective 3-space in $\mathrm{SP}^3(\mathbb{CP}^1)$. The restricted $\mathrm{PSL}(2;\mathbb{R})$ action divides $\mathbb{RP}^3$ into four components. The discriminant locus $\Delta^0 := \mathbb{P}\Delta^{-1}(0)$ consists of the union of the orbits with triple and double roots, and is homeomorphic to a torus[29]. Over $\mathbb{R}$ the

[28]For quadratics, the roots map is a homeomorphism on the entire space: there are no 2-dimensional subfamilies, and all 1-dimensional curves are geodesics.

[29]Note this torus is not smoothly embedded in the space of coefficients, and is singular along the circle parameterizing cubics with a triple root.

generic case splits into polynomials with three distinct real roots Δ^+ (positive discriminant) and those with a complex conjugate pair of complex roots Δ^- (negative discriminant). We see below each of $\Delta^\pm$ are individually homeomorphic to a solid torus, forming the standard Heegaard decomposition of $\mathbb{RP}^3$.



(A) Triple Root (B) Double Root (C) 3 Real Roots (D) 1 Real, 2 Complex Conjugate Roots

FIGURE 26. The four $\mathrm{PSL}(2;\mathbb{R})$ orbits on the $\mathbb{RP}^3$ of real cubics described in Section 4.3.1, as subsets of $\mathbb{RP}^3$. Compare with the quadratic case in Figure 20. As in Figure 17, the topology of each component is recovered by identifying points on the boundary sphere via the antipodal map.

The action of $\mathrm{PSL}(2;\mathbb{R})$ is transitive on all four components, so we may consider each as a homogeneous geometry. Like in the complex case, via dimension count we see that $\dim \Delta^+ = \dim \Delta^- = \dim \mathrm{PSL}(2;\mathbb{R})$, so this action has at most a discrete stabilizer. For Δ^+, this stabilizer is the symmetric group on three elements[30], as $\mathrm{PSL}(2;\mathbb{R})$ acts simply transitively on the space of ordered distinct triples in $\mathbb{RP}^1$. For Δ^- however, this action is free. In the following section we look in detail at this geometry both from the coefficients and roots perspectives.

4.3.2. *Cubics and the Unit Tangent bundle to* $\mathbb{H}^2$. Each cubic in Δ^- has a real root and complex conjugate pair of complex roots. The image under $\mathcal{R}$ in $\mathrm{SP}^3(\mathbb{CP}^1)$ is easily described, as

$$\mathcal{R}(\Delta^-) = \left\{\{r, z, \bar{z}\} \mid r \in \mathbb{RP}^1,\ z = x + iy \text{ for } x, y \in \mathbb{R}, y > 0\right\}.$$

This is exactly an embedding of the product of $\mathbb{RP}^1$ and the upper half plane in $\mathrm{SP}^3(\mathbb{CP}^1)$, so this describes the homeomorphism of Δ^- with a solid torus, as observed in Figure 26d. As z is a point in the upper half plane being acted on by $\mathrm{PSL}(2;\mathbb{R})$ it is tempting to think of it as a point of $\mathbb{H}^2$. From this perspective, the real root $r \in \mathbb{RP}^1$ lies on the ideal boundary of the upper half plane model of $\mathbb{H}^2$, so we may think of it as specifying a *direction*. We make this precise in the following proposition, which is the first half of Theorem 4.14 highlighted in the introduction to this section.

Proposition 4.17. *Let* $\mathcal{R}(\Delta^-) = \{\{r, z, \bar{z}\} \mid z = x + iy, x \in \mathbb{R}, y \in \mathbb{R}_+, r \in \mathbb{R}\}$ *denote the space consisting of the roots of all cubics in* Δ^-*, equipped with the* $\mathrm{PSL}(2;\mathbb{R})$ *action by*

[30]Thus the space of cubics with three distinct roots is an infinite volume three dimensional orbifold with a geometric structure modeled on $\widetilde{SL}_2(\mathbb{R})$

Möbius transformations acting on each root described above, and let $\mathrm{UT}\mathbb{H}^2 = \{(p, v) \mid p \in \mathbb{H}^2, v \in T_p\mathbb{H}^2, \|v\| = 1\}$ *denote the unit tangent bundle to the hyperbolic plane. Define the map* $\Phi\colon \mathcal{R}(\Delta^-) \to \mathrm{UT}\mathbb{H}^2$ *by sending each triple* $\{r, z, \overline{z}\}$ *to the point* z *and the unit tangent vector* $v \in T_z\mathbb{H}^2$ *such that* v *is the initial tangent vector to the geodesic ray* γ *starting at* z *and limiting to the ideal point* $r = \gamma(+\infty)$*. Then* Φ *is an isomorphism of geometries.*

Proof. First we note that Φ is in fact a homeomorphism: given any point z in the upper half plane and any real number r, there is a unique circle passing through z which intersects the real axis at r. Taking the tangent vector to this circle at z defines the desired unit tangent v and noting that geodesics are uniquely determined by these tangent vectors completes the proof that Φ is a bijection. That Φ is in fact continuous with continuous inverse follows directly from constructions in Euclidean geometry[31] (since, in the upper half plane model here, all hyperbolic geodesics are simply Euclidean circles meeting the real axis along a diameter).

To see this is an isomorphism of geometries, we must further show the natural actions of $\mathrm{PSL}(2;\mathbb{R})$ are preserved by Φ. That is, we need that $A.\Phi(\{z, \overline{z}, r\}) = \Phi(A.\{z, \overline{z}, r\})$ for all $A \in \mathrm{PSL}(2;\mathbb{R})$ and all points of $\mathcal{R}(\Delta^-)$. The action on $\mathcal{R}(\Delta^-)$ is given by equation (8), where A acts on both z, r by the same Möbius transformation of $\mathbb{CP}^1$. The action on $\mathrm{UT}\mathbb{H}^2$ is given by the differential of its action on $\mathbb{H}^2$ by isometries. The compatibility of these actions follows immediately from the fact that geodesics (and thus their endpoints at infinity) are determined by their initial conditions. That is, $\Phi(\{A.z, A.\overline{z}, A.r\})$ is the tangent vector based at $A.z$ pointing in the direction of $A.r$, which is the image of the tangent vector $\Phi(\{z, \overline{z}, r\})$ under A, by existence and uniqueness of solutions to the geodesic equation.

□

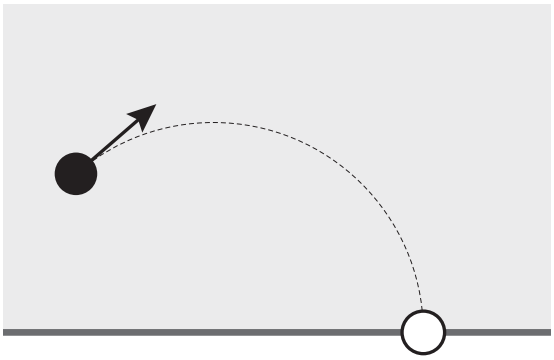


FIGURE 27. Cubic on upper halfplane (black point) and the real conjugate (white point), showing the geodesic between them and the direction from the complex root looking along the geodesic to the real root.

Above we saw that the space $\mathcal{R}(\Delta^-)$ of roots of the cubics in Δ^- can be identified with the unit tangent bundle to $\mathbb{H}^2$ by constructing a homeomorphism between these spaces which was equivariant with respect to the $\mathrm{PSL}(2;\mathbb{R})$ action on each side. But, as we have already seen, the roots map $\mathcal{R}$ is itself a homeomorphism from Δ^- to $\mathcal{R}(\Delta^-)$, which is equivariant with respect to the $\mathrm{PSL}(2;\mathbb{R})$ actions on each. Thus, this map defines an isomorphism of geometries $\Delta^- \cong \mathcal{R}(\Delta^-)$, and thus, transitively, between Δ^- and $\mathrm{UT}\mathbb{H}^2$.

[31] Computing expressions for Φ and its inverse we see they are compositions of elementary (continuous) functions: $\Phi(x \pm iy, r) = (v_1, v_2)_{x+iy} = \left(\frac{w}{\sqrt{1+w^2}}, \frac{1}{\sqrt{1+w^2}}\right)_{x+iy}$ for $w = \frac{2x(x-r)}{x^2+y^2-r^2}$ and $\Phi^{-1}(v_1, v_2)_{x+iy} = \{x \pm iy, r\} = \left\{x \pm iy, x + \left(\frac{v_2}{v_1} + \sqrt{1 + \left(\frac{v_2}{v_1}\right)^2}\right) y\right\}$

Corollary 4.18. *The geometry of* Δ^- *equipped with the* $\mathrm{PSL}(2;\mathbb{R})$ *action defined by Equation 9 is isomorphic to the geometry of the unit tangent bundle to the hyperbolic plane.*

From now on, we will denote this collection of polynomials as $\Delta^- = \mathrm{UT}\mathbb{H}^2_{\mathrm{Coefs}}$ and their associated roots as $\mathcal{R}(\Delta^-) = \mathrm{UT}\mathbb{H}^2_{\mathrm{Roots}}$, to emphasize this geometric structure. To be able to measure distances in these geometries (which right now are only spaces together with a group of symmetries) we need to specify a distance function on each of $\mathrm{UT}\mathbb{H}^2_{\mathrm{Coefs}}$ and $\mathrm{UT}\mathbb{H}^2_{\mathrm{Coefs}}$. As in the case of the hyperbolic plane (Section 4.2.3) we specify this metric indirectly: first we define a norm on each tangent space (induced by an inner product: the Riemannian metric), then we define the length of curves by integrating the norm of their tangent vectors with respect to this, and finally we define the distance between two points to be the length of the shortest curve between them. See [38] for an introductory account of Riemannian manifolds and their metrics.

Definition 4.19 (The Metric on $\mathrm{UT}\mathbb{H}^2_{\mathrm{Roots}}$)**.** As a point in $\mathrm{UT}\mathbb{H}^2_{\mathrm{Roots}}$ may be represented unambiguously as a pair $\{a, z\}$ for $a \in \mathbb{RP}^1$ and z in the upper half plane, a tangent vector to $\mathrm{UT}\mathbb{H}^2_{\mathrm{Roots}}$ at $\{a, z\}$ is a pair $\{u, v\}$ where $x \in \mathbb{R}$ and $v \in \mathbb{C}$. Fixing the basepoint $\{0, i\}$, we define the norm squared of the tangent vector $\{u, v\}$ at $\{0, \pm i\}$ as

$$\|\{u, v\}\|^2 = u^2 + |v|^2.$$

That is, if we think of the tangent vector $(u, v_{\mathsf{Re}}, v_{\mathsf{Im}})$ as a vector in $\mathbb{R}^3$, we are employing the standard Euclidean inner product at the basepoint. We use the free and transitive action of $\mathrm{PSL}(2;\mathbb{R})$ to translate this to every tangent space, equipping $\mathrm{UT}\mathbb{H}^2_{\mathrm{Roots}}$ with a $\mathrm{PSL}(2;\mathbb{R})$ invariant Riemannian metric.

Definition 4.20 (The Metric on $\mathrm{UT}\mathbb{H}^2_{\mathrm{Coefs}}$)**.** As a point in $\mathrm{UT}\mathbb{H}^2_{\mathrm{Coefs}}$ may be represented unambiguously as a monic cubic f (with coefficients the projective point $[1 : a : b : c]$), a tangent vector to $\mathrm{UT}\mathbb{H}^2_{\mathrm{Coefs}}$ at f is an infinitesimal deformation of this cubic which leaves it monic. That is, tangent vectors to $\mathrm{UT}\mathbb{H}^2_{\mathrm{Coefs}}$ are represented by quadratic equations ux^2+vx+w (we write this as (u, v, w) when thinking of it as a tangent vector to the 3-dimensional affine patch $[1 : a : b : c]$). Fixing the basepoint $f = x(x^2 + 1)$, we define the norm squared of the tangent vector (u, v, w) as

$$\|(u, v, w)\|^2 = \left(\frac{u - w}{2}\right)^2 + \left(\frac{v}{2}\right)^2 + w^2.$$

We use the free and transitive action of $\mathrm{PSL}(2;\mathbb{R})$ to translate this to every tangent space, equipping $\mathrm{UT}\mathbb{H}^2_{\mathrm{Roots}}$ with a $\mathrm{PSL}(2;\mathbb{R})$ invariant Riemannian metric.

We quickly comment on the form of the metric on $\mathrm{UT}\mathbb{H}^2_{\mathrm{Coefs}}$, which is the natural choice[32], with respect to the correct choice of affine patch. We may embed the space of quadratics with complex roots into the space of cubics by taking a quadratic f to the polynomial $xf(x)$ which has a unique real root at $x = 0$. Following the previous section, as the space of quadratics can be identified with the hyperbolic plane, we may expect this collection of cubics to look something like an embedding of a projective model of the hyperbolic plane. And it does – with respect to the original affine patch it appears as the *paraboloid model* of the hyperbolic plane, but changing patch gives a round Klein disk, as in Figure 22: and it is with respect to an affine patch of this form that we take the metric to look Euclidean at the basepoint[33].

[32] In fact, while any choice of inner product on the tangent space to our basepoint can be promoted to a Riemannian metric where $\mathrm{PSL}(2;\mathbb{R})$ acts by isometries, this metric is the most symmetric possible choice: its isometry group is 4-dimensional, whereas a generic inner product only leads to a 3-dimensional isometry group.

[33] More precisely, we take the Euclidean form of the metric on the affine patch where the following three curves of polynomials are orthogonal at the polynomial $x(x^2+1)$: (1)those with fixed imaginary root i, varying real root t, (2) those with fixed real root 0, varying real part of complex root, and (3) those with real root 0, varying imaginary part of complex root

These choices of Riemannian metrics will make clear the importance of hyperbolic geometry to the study of cubics. Indeed, we will see in Proposition 4.26 that these metrics are actually *isometric to each other*, and that the roots map provides an isometry between them, in direct analogy to what the quadratic formula provided for $\mathbb{H}^2$.

To make use of this geometry in our analysis of cubics, we need to introduce some facts about the hyperbolic plane's tangent bundle. First, the circle of unit tangent vectors at each point provides a foliation by circles: every point of $\mathrm{UT}\mathbb{H}^2$ is the unit tangent vector to *some unique point of* $\mathbb{H}^2$ and thus lies on a unique one of these circles. But we may also define a collection of sections of $\mathbb{H}^2 \to \mathrm{UT}\mathbb{H}^2$ of the unit tangent bundle such that every point lies on exactly one such hyperbolic plane. For a given ideal point $r \in \partial_\infty \mathbb{H}^2$, at each $z \in \mathbb{H}^2$ we select the unit vector $v_r \in T_z\mathbb{H}^2$ which points[34] to r. This mapping $z \mapsto v_r \in T_z\mathbb{H}^2$ provides one such section, and varying $r \in \partial_\infty \mathbb{H}^2$ foliates the unit tangent bundle with translates of this.

It is straightforward to give an explicit example of one of each of these, passing through the basepoint for the space. Considering $\mathrm{UT}\mathbb{H}^2_{\mathrm{Coefs}}$ with $f = x(x^2+1)$ as the basepoint (the corresponding statements are equally true for $\mathrm{UT}\mathbb{H}^2_{\mathrm{Roots}}$ and $\{0, \pm i\}$ as the basepoint), the circle fiber through f is just all cubics in $\mathrm{UT}\mathbb{H}^2_{\mathrm{Coefs}}$ with x^2+1 as their irreducible quadratic factor, and the hyperbolic plane through f is all such cubics having x as their linear factor. Using the metric from Definition 4.20, these two spaces are seen to intersect each other orthogonally at f. But as the metric on the entire space is built from the metric at this basepoint, we conclude that the same behavior is observed at every point. We note this precisely in the following observation.

Observation 4.21. *The foliation of* $\mathrm{UT}\mathbb{H}^2_{\mathrm{Roots}}$ *by circles (the fibers of the bundle) correspond to collections of cubics with a fixed complex root* $\mathbb{RP}^1_z = \{\{a, z, \overline{z}\} \mid a \in \mathbb{RP}^1\}$*, as the real root varies. Conversely, fixing the real root* $a \in \mathbb{RP}^1$ *and letting the complex conjugate roots vary over* $\mathbb{C}\backslash\mathbb{R}$ *gives the foliation by disks* $\mathbb{H}^2_a = \{\{a, z, \overline{z}\} \mid z \notin \mathbb{R}\}$*. with respect to the Riemannian metric on* $\mathrm{UT}\mathbb{H}^2_{\mathrm{Roots}}$*, these two foliations are orthogonal at every point. The first consists of geodesic circles all of length* 2π*, and the second of isometrically embedded hyperbolic planes.*

Together, these foliations provide a *trivialization* of the unit tangent bundle: an explicit choice of homeomorphism[35] $\mathrm{UT}\mathbb{H}^2 \to \mathbb{H}^2 \times \mathbb{S}^1$, sending each $v \in \mathrm{UT}\mathbb{H}^2$ to its basepoint $z \in \mathbb{H}^2$, and the angle θ that v makes with respect to the direction field associated with some fixed ideal point. Working in the upper half plane model, it is easiest to choose this as the direction field associated to ∞; as in the euclidean coordinates $z = x + iy$ this is simply the direction made with the vertical at every point. Explicitly, the point (z, r) in the space of roots, identified with the unit tangent vector $v \in T_z\mathbb{H}^2$, is sent to (z, θ) for

$$\theta = 2 \arctan\left(\frac{r - x}{y}\right). \tag{10}$$

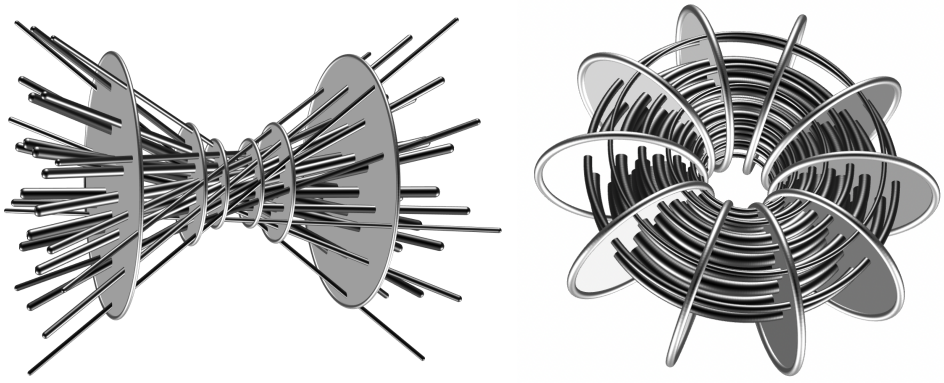
Remark 4.22. Using a Möbius transformation to send the upper half plane to the unit disk, we may represent the space of roots as the interior of a solid torus of revolution in $\mathbb{R}^3$, as in Figure 28b. This homeomorphism provides the beautiful pictures visible in Figure 10 produced by David Dumas' wonderful program.

This geometric perspective provides a nice way of thinking about the space of roots, avoiding the complicated space $\mathrm{SP}^3(\mathbb{CP}^1)$ in which it was originally defined. The upper half plane model of $\mathbb{H}^2$ is a subset of $\mathbb{CP}^1$, so the map Φ in Proposition 4.17 above explicitly identifies $\mathcal{R}(\Delta^-)$ with a subset of $\mathrm{UT}\mathbb{CP}^1$. But, as $\mathbb{CP}^1$ is topologically the 2-sphere (Remark 4.2), $\mathrm{UT}\mathbb{CP}^1$ is just

[34] That is, choose v so that the geodesic γ_v with initial tangent v at z has $\lim_{t\to\infty+} \gamma(t) = r$.

[35] Note however this homeomorphism is *not* an isometry of the metrics we have defined with the product metric on $\mathbb{H}^2 \times \mathbb{R}$.

the unit tangent bundle of the sphere[36], which is topologically $\mathbb{RP}^3$. As the space of coefficients is naturally a subset of $\mathbb{RP}^3$, this provides a uniform means of drawing both spaces, see Figures 28a and 30.



(A) (B)

FIGURE 28. Foliation of $\mathrm{UT}\mathbb{H}^2_{\mathrm{Roots}}$ by circles and hyperbolic planes. In (A) we view $\mathrm{UT}\mathbb{H}^2_{\mathrm{Roots}}$ in an affine patch of $\mathrm{UT}\mathbb{CP}^1 = \mathbb{RP}^3$, where the hyperbolic foliation is visible as parallel disks. In (B), it is realized as the interior of a torus of revolution in $\mathbb{R}^3$ via Remark 4.22.

Remark 4.23. This also provides the means to represent cubics literally as unit tangent vectors in $\mathbb{H}^2$: As θ is measured with respect to geodesics limiting to ∞ in the upper half plane, we may depict the cubic with roots $\{z, \overline{z}, r\}$ by the unit vector at z pointed along the geodesic[37] to r. As this depicts a 3-dimensional space using 2-dimensions, we cannot understand the entire space of cubics this way - however it provides a useful means of looking at 2-dimensional families, such as Figure 29.

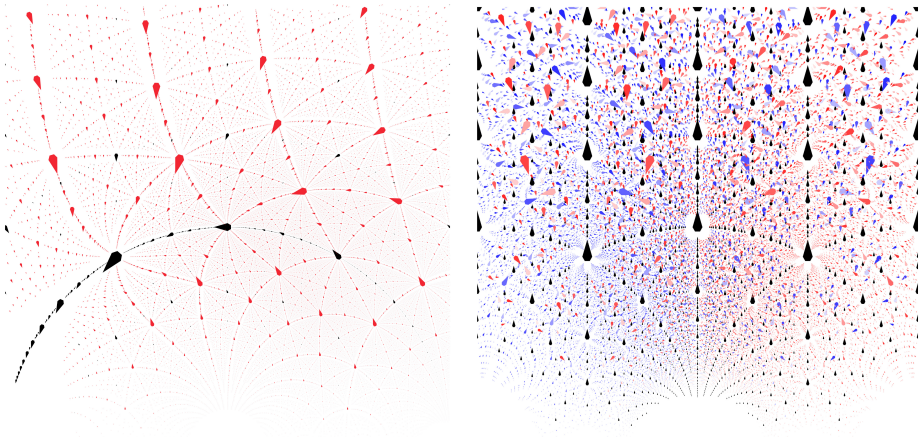
We next turn to the description of this geometry on the space of coefficients, which we likewise denote $\mathrm{UT}\mathbb{H}^2_{\mathrm{Coefs}}$. Here, the surface cut out by the discriminant represents the ideal boundary of the geometry, whose points correspond to the polynomials with negative discriminant, depicted in Figure 26d. The important geometric information is a description of the $\mathbb{S}^1$ fibers, which provide the structure of the unit tangent bundle, and a choice of section $\mathbb{H}^2 \to \mathrm{UT}\mathbb{H}^2_{\mathrm{Coefs}}$ giving a trivialization. Via Observation 4.21, the choice made in the roots model has a convenient description in terms of their corresponding polynomials, which we summarize below.

Observation 4.24. *The trivialization in Observation 4.21 is expressed in the space of coefficients as:*

- *Each fiber of the $\mathbb{S}^1$ foliation passes through a unique cubic with real root at infinity, identified with the quadratic $ax^2 + bx + c \in \mathbb{H}^2$. The fiber passing through this point is parameterized by $[a : b - ra : c - rb : rc]$ in the space of coefficients, for $r \in \mathbb{RP}^1 = \mathbb{R} \cup \{\infty\}$.*
- *Each fiber of the $\mathbb{H}^2$ foliation has constant real root $r \in \mathbb{RP}^1$. This fiber is parameterized by $[1 : -2u - r : 2ru - u^2 - v^2 : -ru^2 - rv^2]$ for $u, v \in \mathbb{R}, v > 0$ representing the varying complex root $u + iv \in \mathbb{H}^2$.*

[36]One shows the isometries of $\mathbb{S}^2$ act freely and transitively on the unit tangent bundle, which then provides a diffeomorphism from $\mathrm{Isom}(\mathbb{S}^2) = \mathrm{SO}(3)$ and $\mathrm{UT}\mathbb{S}^2$. Finally, we recall that $\mathrm{SO}(3)$ is topologically real projective 3-space (for example, by noting that it is double covered by $\mathrm{SU}(2) \cong \mathbb{S}^3$).

[37]Using the coordinates $z = x + iy$ on the upper half plane, this vector is in the direction $(2y(r - x), (r - x)^2 - y^2)$.



(A) (B)

FIGURE 29. Cubic numbers as a vector field, as shown in Figure 27. The roots of the polynomial family $ax^3 + bx^2 + cx + a$ (29a), and all cubics (29b).

Exactly as in Observation 4.21, with respect to the choice of metric in Definition 4.20, these two spaces of fibers are orthogonal at every point of intersection. Every circle fiber has metric length 2π and every plane in the $\mathbb{H}^2$ foliation is an isometrically embedded hyperbolic plane (thus justifying the name).

As in the quadratic case, the natural geometry is more difficult to see at first from the coefficient perspective. However, drawing $\mathrm{UT}\mathbb{H}^2_{\mathrm{Coefs}}$ in the affine patch which puts quadratics at infinity, we see the hyperbolic foliation consists of copies of the now familiar *parabola model* of $\mathbb{H}^2$. Choosing other affine patches may render (some of) these as copies of the more familiar Klein disk model. These two foliations of the space of cubics, by hyperbolic planes and by circles, provide convenient means of keeping track of information about cubics. We see some examples of this below as we seek a geometric understanding of the roots map.

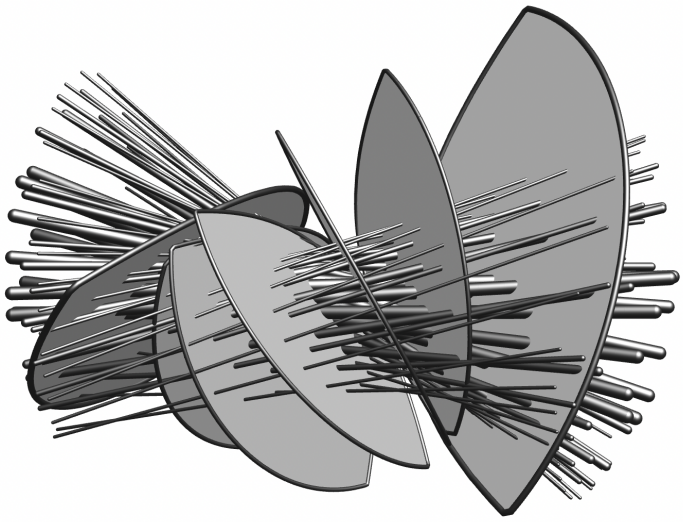


FIGURE 30. Foliation of $\mathrm{UT}\mathbb{H}^2_{\mathrm{Coefs}}$ by hyperbolic planes and circle fibers

4.3.3. *Geometry of the Roots Map.* The roots map $\mathcal{R}\colon \mathbb{P}\text{Coefs} \to \text{Roots}$ restricts on the space of real cubics of negative discriminant to a homeomorphism $\text{UT}\mathbb{H}^2_{\text{Coefs}} \to \text{UT}\mathbb{H}^2_{\text{Roots}}$ equivariant with respect to the $\text{PSL}(2;\mathbb{R})$ actions given by equations (8) and (9). We noted earlier (Proposition 4.17 and Corollary 4.18) that this implies $\mathcal{R}$ is an isomorphism of geometries in the sense of Klein. Following this, we saw that both the space of roots and coefficients can be equipped with a natural choice of Riemannian metric in definitions 4.19 and 4.20. We now strengthen our original proposition, and show that with respect to these two metrics in fact $\mathcal{R}$ is an isometry. Because computing the actual geodesic metric distance here is quite challenging (see [24] for a computation of the geodesic curves), we adopt a different approach than our proof of Theorem 4.9 for quadratics, and work locally, leveraging the equivariance of $\mathcal{R}$ with respect to the group actions to reduce the problem to showing $\mathcal{R}$ induces an isometry of tangent spaces at a single point. This technique relies on a useful lemma of Riemannian geometry, stated below.

Lemma 4.25. *Let (X, g_X) and $Y, g_Y)$ be Riemannian manifolds, each equipped with a transitive action of some Lie group G by isometries. Suppose further that $f\colon X \to Y$ is a diffeomorphism which is equivariant with respect to these G actions. Then, f preserves the inner product at any point $p \in X$, it is an isometry.*

The proof of this lemma is a straightforward computation using the fact that G acts on both sides via isometries. See again [38] for an introduction to the tools utilized in such arguments. We sketch the proof below.

Sketch. Let p be any point in X. We wish to see that the map f preserves the inner product at p. But using the homogeneity of the G action we can choose some isometry taking the fixed point x in the theorem statement to p. Then using equivariance of the G action with respect to f, we see this same isometry element takes $f(x)$ to $f(p)$. We can use this isometry to transfer any local computation at p to a local computation at x, and similarly for $f(p)$ and $f(x)$. Putting this all together, we see that our map f preserves the inner product at p if and only if it does at x. But this latter condition was precisely our assumption: thus f is a Riemannian isometry.

We now use this result to prove the proposition of interest.

Proposition 4.26. *Let $\text{UT}\mathbb{H}^2_{\text{Coefs}}$ be the set of real cubics with exactly one real root, $\text{UT}\mathbb{H}^2 = \{[a:b:c:d] \mid \Delta_3(a,b,c,d) < 0\}$, and $\text{UT}\mathbb{H}^2_{\text{Roots}} = \{\{r, z, \overline{z}\} \mid r \in \mathbb{RP}^1, z \in \mathbb{C}\backslash\mathbb{R}\}$ be the set of their root-sets. Equipping each with the $\text{PSL}(2;\mathbb{R})$ metric identifying them with the unit tangent bundle to the hyperbolic plane, the roots map $\mathcal{R}\colon \text{UT}\mathbb{H}^2_{\text{Coefs}} \to \text{UT}\mathbb{H}^2_{\text{Roots}}$ is an isometry.*

Proof. By the above proposition and the equivariance of the $\text{PSL}(2;\mathbb{R})$ actions by isometries, it is enough to show that $\mathcal{R}$ defines a Riemannian isometry at any single point. We choose to compute at the basepoint $f = x(x^2+1)$ in $\text{UT}\mathbb{H}^2_{\text{Coefs}}$ and its image $\mathcal{R}(f) = \{0, \pm i\}$ in $\text{UT}\mathbb{H}^2_{\text{Roots}}$. Because writing an actual formula for the roots map in degree 3, we opt instead to work with its inverse (if $\mathcal{R}^{-1}$ is a Riemannian isometry then so is $\mathcal{R}$). This inverse takes the point $\{r, z, \overline{z}\}$ to the polynomial that has r and $z, \overline{z}$ as roots: written in the affine patch of monic cubics, this has the explicit formula

$$\mathcal{R}^{-1}(\{r, z, \overline{z}\}) = (x-r)(x-z)(x-\overline{z}) = (x-r)(x^2 + 2\text{Re}(z) + |z|^2).$$

Let $\{u, v+iw\}$ denote a tangent vector to $\{0, \pm i\}$ as in Definition 4.19. We may realize this tangent vector as the derivative at $t=0$ of the path $\{tu, i + t(v+iw)\}$. Thus we may compute the result of applying $\mathcal{R}^{-1}$ to this tangent vector by looking at the path $\mathcal{R}^{-1}(\{tu, i+t(v+iw)\})$ and taking its derivative in the space of coefficients. Performing this computation, we arrive at the path of polynomials

$$x^3 + (2tv - tu)x^2 + (1 - 2t^2uv + t^2v^2 + 2tw + t^2w^2)x - (tu + t^3uv^2 + 2t^2uw + t^3uw^2)$$

differentiating at $t=0$ gives the tangent vector, realized a quadratic polynomial as in Definition 4.20.

$$(2v-u)x^2 + 2wx - u$$

Having these two vectors on hand, the rest of the proof is a direct computation: we compute the norm of each with respect to the Riemannian metric on each respective space, and then see the results are equal. First, in the space of roots, using Definition 4.19 we see the norm square of $\{u, v \pm iw\}$ is simply its Euclidean value:

$$\|\{u, v \pm iw\}\|^2_{\text{Roots}} = u^2 + v^2 + w^2.$$

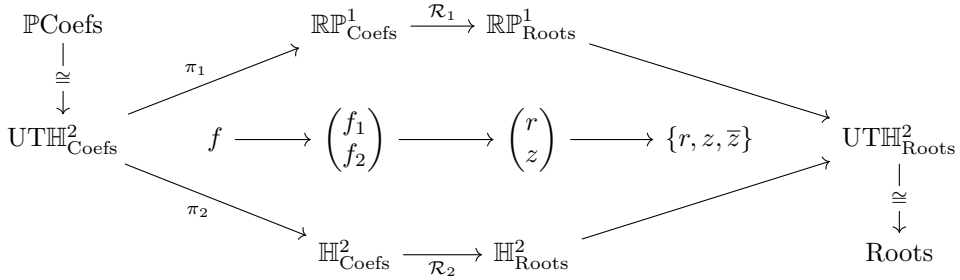
Performing the analogous computation in coefficient space, we take the coefficient vector $(-(u+2v), 2w, -u)$ of the quadratic above, and apply the norm squared of Definition 4.20:

$$\begin{aligned}\|((2v-u), 2w, -u)\|^2_{\text{Coefs}} &= \left(\frac{(2v-u)+u}{2}\right)^2 + \left(\frac{2w}{2}\right)^2 + (-u)^2 \\ &= v^2 + w^2 + u^2\end{aligned}$$

Thus $\mathcal{R}^{-1}$ induces an isometry between the tangent spaces $T_{\{0,\pm i\}}\text{UT}\mathbb{H}^2_{\text{Roots}}$ to the roots and $T_{x(x^2+1)}\text{UT}\mathbb{H}^2_{\text{Coefs}}$ to the coefficients at our chosen basepoint, as required. □

After the right preliminary work, the proof of our main theorem, (the analog of Theorem 4.9 but for cubics) reduced to checking a simple computation at the basepoint. But, we can say even more than this, and relate exactly how this isometry acts with respect to the foilations of the space of cubics by circles and hyperbolic planes (Observation 4.21).

Theorem 4.27. *The roots map $\mathcal{R}\colon \mathbb{P}\text{Coefs} \to \text{Roots}$, restricted to the space $\text{UT}\mathbb{H}^2_{\text{Coefs}}$ of real cubics with negative discriminant, factors as $\mathcal{R}_3 = (\mathcal{R}_1, \mathcal{R}_2) \circ (\pi_1, \pi_2)$ where $\pi = (\pi_1, \pi_2)$ is the trivialization of the unit tangent bundle given by the foliations by circles, and hyperbolic planes, respectively, in Observation 4.24, and $\mathcal{R}_1, \mathcal{R}_2$ are the root maps for linear and quadratic polynomials. This is best seen diagrammatically: compare the diagram below with Figure 31.*



Proof. This is not deep, and follows directly from our geometric interpretations of the space of quadratic and linear polynomials. As isometries of $\text{UT}\mathbb{H}^2$ preserve the fiber bundle structure, $\mathcal{R}$ preserves the $\mathbb{RP}^1$ foliation, and similarly the $\mathbb{H}^2$ foliation described in Observations 4.21 and 4.24. By Observation 4.21, we may identify these foliations with the preimages of projections π_1, π_2 onto the linear and irreducible quadratic factors respectively. As these are preserved by the roots map, $\mathcal{R}_3$ factors through the trivialization π to a pair of maps sending these factors to their roots. But these are already familiar: they are just the lower dimensional roots maps $\mathcal{R}_1$ and $\mathcal{R}_2$.

□

The bottom row of this diagram gives the formula for the complex root as in Theorem 4.14; the top row gives the analog returning the real root. Choosing to identify the base $\mathbb{H}^2_{\text{Coefs}}$ with one of the sections of Observation 4.24, one may see the solution to the cubic as a *process* as follows. (For simplicity of exposition, we identify it with the fiber of cubics with real root zero here.) Starting with a point $f \in \text{UT}\mathbb{H}^2_{\text{Coefs}}$ (recall Figure 30 as a visual aid here), we slide along the fiber through f until reaching the hyperbolic sheet specified by the section. The real root is given by the distance traveled along this fiber, and the complex root is given by taking the

resulting point, which is now a cubic of the form $(ax^2 + bx + c)x$ and applying the isometry from the parabola model of $\mathbb{H}^2$ in $[a : b : c]$ to the upper half plane.

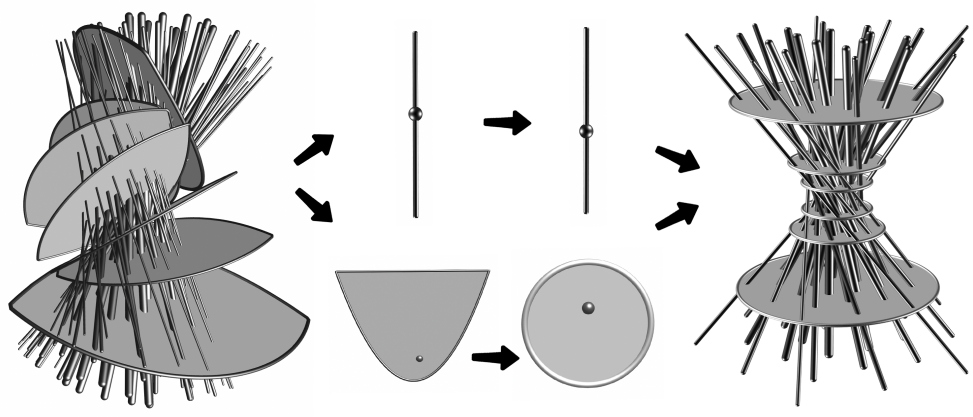


FIGURE 31. Geometric factoring of the roots map $\mathcal{R}_3 \colon \mathrm{UT}\mathbb{H}^2_{\mathrm{Coefs}} \to \mathrm{UT}\mathbb{H}^2_{\mathrm{Roots}}$ of Theorem 4.27, using the root maps $\mathcal{R}_1, \mathcal{R}_2$ of lower degrees. On the left, the space $\mathrm{UT}\mathbb{H}^2_{\mathrm{Coefs}} \subset \mathbb{RP}^3$ is depicted together with the trivialization determined by fixing a real or complex root. On the right is the space $\mathrm{UT}\mathbb{H}^2_{\mathrm{Roots}} \subset \mathrm{UT}\mathbb{CP}^1$ of unordered roots with the same trivialization. Compare with the diagram in Theorem 4.27.

Remark 4.28. One may derive an expression for the cubic formula from this procedure, similarly to what was done in Remark 4.10, using the parameterization of the $\mathbb{RP}^1$ and $\mathbb{H}^2$ factors in Observation 4.24 and attempting to invert their dependence of the coefficients $[a : b : c : d]$ on the parameters u, v. Of course, the inherent messiness of the cubic formula must make this challenging at some point, which we can now specify: it lies in giving explicit formulas for the projections of $\mathrm{UT}\mathbb{H}^2_{\mathrm{Coefs}}$ onto its foliations[38].

4.3.4. *Applications to Cubic Numbers.* When reasoning about cubic numbers, we often want to deal just with the points in $\mathbb{C}$ themselves, and not the abstract space of roots. Thus, from the perspective of polynomials, we are interested in the image of the *projection* $\mathrm{UT}\mathbb{H}^2_{\mathrm{Roots}} \to \mathbb{H}^2_{\mathrm{Roots}} \subset \mathbb{C}$, not the space of roots itself. This is yet another place where the circle and hyperbolic fibrations of the space of cubics are important, as they are the kernel and 1-eigenspaces of the differential of this projection, respectively.

One place this may arise is in trying to bound distances between cubic numbers. Thus, to control distances between cubic numbers in terms of their minimal polynomials, we do not need to use a full expression[39] for the unit tangent bundle metric on $\mathrm{UT}\mathbb{H}^2_{\mathrm{Coefs}}$ but rather just a means of measuring the hyperbolic distance between their projections to the base $\mathbb{H}^2$. While abstractly this is given exactly by the pull back of the metric on $\mathbb{H}^2_{\mathrm{Coefs}}$ by the bundle projection $\mathrm{UT}\mathbb{H}^2 \to \mathbb{H}^2$, any sufficiently simple expression for this would hopefully make some of the analysis of Section 6 in the quadratic case extendable to cubics.

[38] In light of Theorem 4.27 it cannot be anywhere else: as the only remaining portion to the cubic formula from this perspective is to solve the associated quadratic and linear equations: both of which have simple roots maps as we have seen before.

[39] While the Riemannian metric here is easy to describe by translating the standard euclidean metric on some tangent space to a point $p \in \mathrm{UT}\mathbb{H}^2_{\mathrm{Coefs}} \subset \mathbb{RP}^3$ around by action given by the representation ρ, its expression is complicated, making the computation of a distance function unwieldy.

Another place this arises is in the study of one and two parameter families of cubics. Let $F\colon \mathbb{R}^n \to \mathbb{P}\text{Coefs}$ be a smooth map for $n \in \{1,2\}$ tracing out a submanifold of the space of cubics. A natural question for the production of good images, is when does the result of drawing the complex root of each polynomial in the family produce a coherent image in $\mathbb{C}$: that is, when is the composition $\mathcal{R}_{\mathbb{C}} \circ F$ a homeomorphism onto its image? Given that on the space of roots, the projection onto the complex root is directly collapsing the $\mathbb{RP}^1_{\text{Roots}}$ factor, we recall $\mathcal{R}_{\mathbb{C}}$ is the projection along the $\mathbb{RP}^1$ fibers of Observation 4.24, followed by an isometry. This gives an explicit condition on F.

Observation 4.29. *Let $F\colon \mathbb{R}^n \to \mathrm{UTH}^2_{\text{Coefs}}$ be a family of real cubics in the space of projectivized coefficients, for $n \in \{1,2\}$. Then the map $\mathcal{R}_{\mathbb{C}} \circ F$ sending a cubic to its complex root is an embedding if F is everywhere transverse to the $\mathbb{RP}^1$ fibration of $\mathrm{UTH}^2_{\text{Coefs}}$.*

Restricting attention to planar starscapes, the relevant families $F\colon \mathbb{R}^2 \to \mathbb{RP}^3$ are those whose image is some affine plane $\mathcal{S} \subset \mathbb{RP}^3$ (that is, the projectivization of some $\mathbb{R}^3$ through the origin) intersected with the space of cubics of negative discriminant, as in Corollary 4.15 highlighted in the introduction to this section. Examples include the accompanying figure (25) as well as the gallery Figures 2a, 2b and 7a–7d. The fiber above any point $r \in \mathbb{C}$ is a line $(\mathcal{R}_2\pi_2)^{-1}(r) = L_r$ (parameterized as in Observation 4.24), thus any affine plane $\mathcal{S} \subset \mathrm{UTH}^2_{\text{Coefs}}$ containing L_r fails to embed under projection onto the complex root. As we may find an affine plane containing any projective line we like, this can happen at any point in the upper half plane. The fact that this can only happen at a single point in a single starscape is implied by the fact that the fibers of $\mathrm{UTH}^2_{\text{Coefs}}$ are pairwise skew lines in $\mathbb{RP}^3$, and thus no two are ever contained in a plane.

Corollary 4.30. *Let $\mathcal{S} \subset \mathrm{UTH}^2_{\text{Coefs}}$ be an affine 2-dimensional projective subspace, and $\mathcal{R}|_{\mathcal{S}}\colon \mathcal{S} \to \mathbb{C}$ be the projection onto the complex root. Then $\mathcal{R}|_{\mathcal{S}}$ is singular along at most one projective line $L \subset \mathcal{S}$. Furthermore, for a planar starscape (i.e. $\mathcal{S}$ has rational normal vector), this line, if it exists, corresponds to a rational polynomial with a linear factor over $\mathbb{Q}$ (i.e. a quadratic point in the complex plane). Finally, every such polynomial gives a singularity for $\mathcal{R}_{\mathcal{S}}$ for some planar starscape $\mathcal{S}$.*

To see the statement that the line lies over a quadratic point, suppose the projective line corresponds to roots $\{r, z, \overline{z}\}$, where r is the varying real root, while z and $\overline{z}$ are fixed. Then the projective line has an expression as

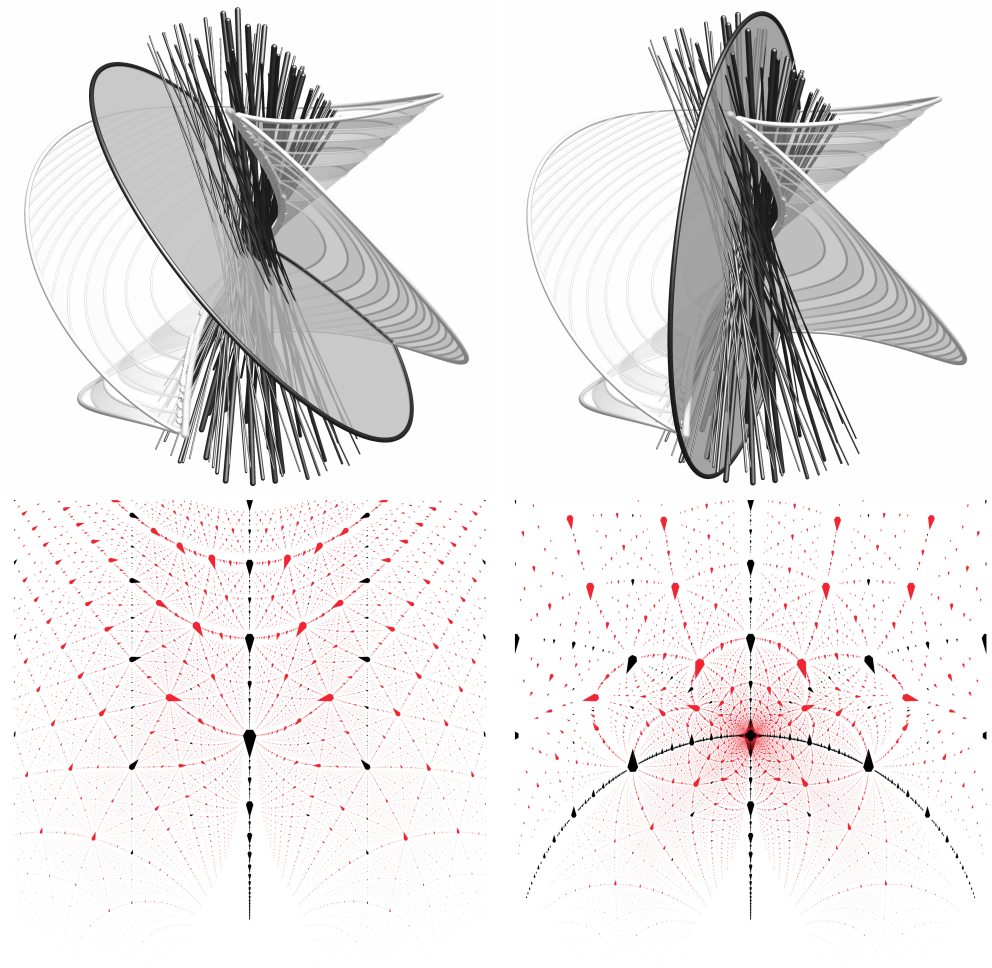
$$[1 : -z - \overline{z} - r : z\overline{z} + (z + \overline{z})r : -z\overline{z}r]. \tag{11}$$

If this lies in a planar starscape, then it lies in a projective plane with rational normal, i.e. there is a rational vector $[r : s : t : u]$ such that

$$\begin{pmatrix} 1 & -z-\overline{z} & z\overline{z} & 0 \\ 0 & 1 & -z-\overline{z} & z\overline{z} \end{pmatrix} \begin{pmatrix} r \\ s \\ t \\ u \end{pmatrix} = 0.$$

In particular, this implies 1, $z+\overline{z}$ and $z\overline{z}$ are $\mathbb{Q}$-linearly dependent. If they have rank two, then this implies that $[r : s : t] = [s : t : u]$, which implies $[r : s : t : u]$ has the form $[1 : x : x^2 : x^3]$; this further implies z is rational (as the projective line in coefficient space has a fixed rational root x). If they have rank one, then z lies on the intersection of two rational geodesics, i.e. is quadratic (see Observation 4.12).

For the final statement, suppose the projective line L is as in (11) above. Then it lies on infinitely many rational geodesics; taking any two will span an appropriate $\mathcal{S}$.



(A) $ax^3 + cx + d$ (B) $ax^3 + bx^2 + cx + b$

FIGURE 32. 2-parameter projective families of cubics, plotted in coefficient space together with their projection onto their complex root. The family on the left is transverse to the $\mathbb{S}^1$ fibers of $\mathrm{UT}\mathbb{H}^2_{\mathrm{Coefs}}$, and so the projection onto roots is an embedding. The family on the right is not transverse to the $\mathbb{S}^1$ foliation, and contains the fiber $(x^2+1)(px+q)$. Thus the projection onto the complex root is not an embedding, and collapses an entire curve above i.

5. Diophantine approximation

5.1. **Classical Diophantine approximation.** The study of Diophantine approximation is the study of the relative placement of real or complex numbers with regards to their arithmetic complexity. To illustrate, we consider the unit interval $[0, 1]$. We measure the complexity of a rational number p/q in lowest terms by its denominator, defining its *height* to be q. Then we observe a fundamental phenomenon one might call **repulsion**: distinct rationals of low height cannot be too close to one other. Explicitly,

$$\left|\frac{p_1}{q_1} - \frac{p_2}{q_2}\right| \geq \frac{1}{q_1 q_2}. \tag{12}$$

This leads one to consider the question of **good approximations**: fix $\alpha \in [0,1]$ and ask whether there are rational p/q which are surprisingly close to α, in terms of their height. One asks whether there are infinitely or finitely many p/q such that

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{q^k}. \tag{13}$$

The behaviour with regards to exponent $k = 2$ distinguishes rationals from irrationals.

Theorem 5.1 (Dirichlet [23])**.** *Let $\alpha \in \mathbb{R}$. Then α is irrational if and only if there exist infinitely many distinct $p/q \in \mathbb{Q}$ such that*

$$|\alpha - p/q| < 1/q^2.$$

In other words, rationals are "poorly approximable" and irrationals are "well approximable." This can be proven by a simple pigeonhole principle argument, which we include here for the sake of exposition, as later proofs will imitate the method.

Proof. Let α be irrational. Choose an integer $Q > 1$. Divide the unit interval $[0,1]$ into Q even subintervals. Then, among the real numbers $0, \alpha, 2\alpha, \ldots, Q\alpha$, there must be two, say $i\alpha$ and $j\alpha$, where $0 \leq i < j \leq Q$, whose fractional parts fall into the same subintervals. Then we have $|(j-i)\alpha - p| < 1/Q$ for some integer p. Letting $q = j - i$, observe that $q \leq Q$, and we obtain $|\alpha - p/q| < 1/qQ \leq 1/q^2$. As α is irrational, we may choose Q' to be such that $|q\alpha - p| > 1/Q'$, and run the argument again; by construction, we discover a new, distinct rational approximation. In this way, if α is irrational we discover infinitely many such approximations. By contrast, if α is rational, then (12) limits the ability to find good approximations. □

Dirichlet's Theorem is illustrated in Figure 14; if one places disks over each rational p/q with radius $1/q^2$, then the irrationals are covered by infinitely many disks, while rationals by only finitely many. We can create a more starscape-esque version by a constant scaling, in Figure 15. This latter version has a different feel, and more vividly illustrates the mutual repulsion of rational numbers.

The natural accompaniment to Dirichlet's elementary result is a deep one of Roth: if the exponent 2 is strengthened to $2+\epsilon$ for any positive ϵ, then all *algebraic* α fail to have infinitely many good approximations [54].

Theorem 5.2 (Roth [54])**.** *Let $\epsilon > 0$. Let $\alpha \in \mathbb{R}$ be algebraic of degree ≥ 2. Then there are only finitely many distinct $p/q \in \mathbb{Q}$ such that*

$$|\alpha - p/q| < 1/q^{2+\epsilon}.$$

This finiteness is in fact true for almost all real numbers α (in the sense of Lebesgue measure), a result due to Khintchine [36, Theorem 29].

However, one can construct real numbers which are well-approximable to all higher exponents, called *Liouville numbers* after Liouville's famous construction [42]. The key to constructing a Liouville number is to artificially build something incredibly close to a series of rational numbers. The simplest example is $\sum_{k=0}^{\infty} \frac{1}{10^{k!}}$: the partial sums form rational approximations that are too good to allow for a finiteness property like Theorem 5.2, even for any fixed positive exponent of q. These were the first explicit transcendental numbers.

Having studied the exponent k in approximation within $1/q^k$, we can turn to a finer question of constants. For example, does Dirichlet's Theorem hold if $1/q^2$ is replaced with $1/Cq^2$ for various increasing values of C? The theorem holds until $C = \sqrt{5}$, above which the golden ratio and certain of its relatives are no longer approximable by infinitely many rationals. This state of affairs continues until another tipping point, $C = 2\sqrt{2}$, above which $\sqrt{2}$ is poorly approximable, and so on. These tipping points form the beginning of the theory of the *Lagrange spectrum*.

For the rich theory of Diophantine approximation, including further historical context for these results, the reader may begin with [16, 33, 58].

The Diophantine approximation of the complex plane away from the real line is less well-studied. Here we may ask the approximants to come from number fields: in our context, it is natural to stratify these by degree, and ask: how well-approximable is a complex number by algebraic numbers of degree $\leq d$?

In order to do so, we need to generalize the notion of height, to measure the arithmetic complexity of algebraic numbers in general.

5.1.1. *Measuring arithmetic complexity with the naïve height of a polynomial.* The simplest notion of arithmetic complexity may be with reference to the coefficients of its minimal polynomial. Recall that an algebraic number has a unique *minimal polynomial*, most often taken to be the unique monic irreducible $f(x) \in \mathbb{Q}[x]$ for which it is a root. We may take this polynomial to be in $\mathbb{Z}[x]$ by scaling up the denominator, which causes us to lose the monic condition. We will write f_α for the unique scaling whose coefficients are integral but with no common factor, and positive leading coefficient, and refer to this as the *minimal polynomial*, following much of the literature of Diophantine approximation (for example [17]).

Consider any polynomial $f = a_d x^d + \cdots a_1 x + a_0 \in \mathbb{Z}[x]$. Then the *naïve height* of f is defined as

$$H(f) := \max_{0 \leq i \leq d} |a_i|.$$

This measure has the advantage of simplicity, and a close connection to the linear algebra of the lattice of polynomials of degree d in the space of coefficients. It is also clear that there are only finitely many polynomials of bounded degree and height.

We will write $H(f_\alpha)$ for the height of the minimal polynomial f_α of α, which measures the arithmetic complexity of α.

5.1.2. *Good approximations drawn from fixed degree.* To generalize Dirichlet's Theorem 5.1, one might ask to approximate by algebraic numbers of bounded degree d (so that Dirichlet's Theorem becomes the case of $d = 1$, i.e. approximation by rationals). One can define, following Koksma [37] the quantity $k_d(\alpha)$ to be the surpremum of all k such that there are infinitely many algebraic numbers β of degree $\leq d$ satisfying

$$|\alpha - \beta| < \frac{1}{H(f_\beta)^k}.$$

Dirichlet's Theorem 5.1, in this language, states that $k_1(\alpha) \leq 2$ for α rational and $k_1(\alpha) \geq 2$ for α real and irrational. Roth's Theorem 5.2 is that $k_1(\alpha) = 2$ for α real and algebraic.

As regards the real case, Wirsing conjectured that for transcendental $\alpha \in \mathbb{R}$, $k_d(\alpha) \geq d + 1$ [67]. This is known for $d = 1$ (from Dirichlet's Theorem) and for $d = 2$ [22].

For general degree, Sprindžuk gave an answer for almost all α (generalizing Khintchine's statement). We see a qualitative difference between real and non-real α.

Theorem 5.3 (Sprindžuk, [61])**.** *For almost all $\alpha \in \mathbb{R}$, $k_d(\alpha) = d+1$. For almost all $\alpha \in \mathbb{C}\backslash\mathbb{R}$, $k_d(\alpha) = (d+1)/2$.*

This gives us a better idea of the natural sizing for algebraic points in the complex plane: a sizing of $1/H(f_\alpha)^{(d+1)/2}$ would be the natural analogue of Figure 15. See Figure 35.

Next, we may consider the generalisation of Roth's Theorem 5.2 governing the approximability of algebraic numbers. Schmidt showed that for algebraic $\alpha \in \mathbb{R}$ of degree at least 2, $k_d(\alpha) = \min\{\deg(\alpha), d + 1\}$ [56]. For complex numbers away from the real line (our concern here), it is slightly more complicated.

Theorem 5.4 (Bugeaud, Evertse [17], Theorem 2.1, Corollary 2.4)**.** *For algebraic $\alpha \in \mathbb{C}\backslash\mathbb{R}$,*

$$k_d(\alpha) = \min\{\deg(\alpha)/2, (d + 1)/2\}$$

except in the case that $\deg(\alpha) \geq d+2$ *and* d *is even. In this case,* $k_d(\alpha) \in \{(d+1)/2, (d+2)/2\}$.

In particular, for $d = 2$*, we have* $k_2(\alpha) = 2$ *if and only if the quantities* 1*,* $\alpha\overline{\alpha}$ *and* $\alpha + \overline{\alpha}$ *are* $\mathbb{Q}$*-linearly dependent (and* $k_2(\alpha) = 3/2$ *otherwise).*

In degree $d > 2$, Bugeaud and Evertse give precise conditions for determining which of the two possibilities for $k_d(\alpha)$ is correct in most cases, but were not able to compute it in all cases.

Let us consider the dichotomy given for $d = 2$, where Theorem 5.4 says that some α are much more approximable than others, based on whether 1, $\alpha\overline{\alpha}$ and $\alpha + \overline{\alpha}$ are $\mathbb{Q}$-linearly dependent. To see why this is the case, we recast this characterization more geometrically: such α have the property that they lie on rational hyperbolic geodesics, i.e. those geodesics corresponding to rational planes in coefficient space. These are exactly the *rational geodesics* discussed in Observation 4.12.

This dichotomy is illustrated in Figure 33. In fact, we will show in Section 6 that this dichotomy holds even for non-algebraic α.

Note that cubics cannot lie on rational geodesics[40], an effect which is quite prominent in Figure 1c. This is why $k_2(\alpha) \in \{3/2, 2\}$ (i.e., we do not need to allow for $k_2(\alpha) = \deg(\alpha)/2$ separately).

Based on this geometric interpretation of the $d = 2$ case of Bugeaud and Evertse, one wonders if the exceptional cases all have similar geometric interpretations. We muse on this briefly in Section 7.

5.2. **Measuring approximation: heights and distances.**

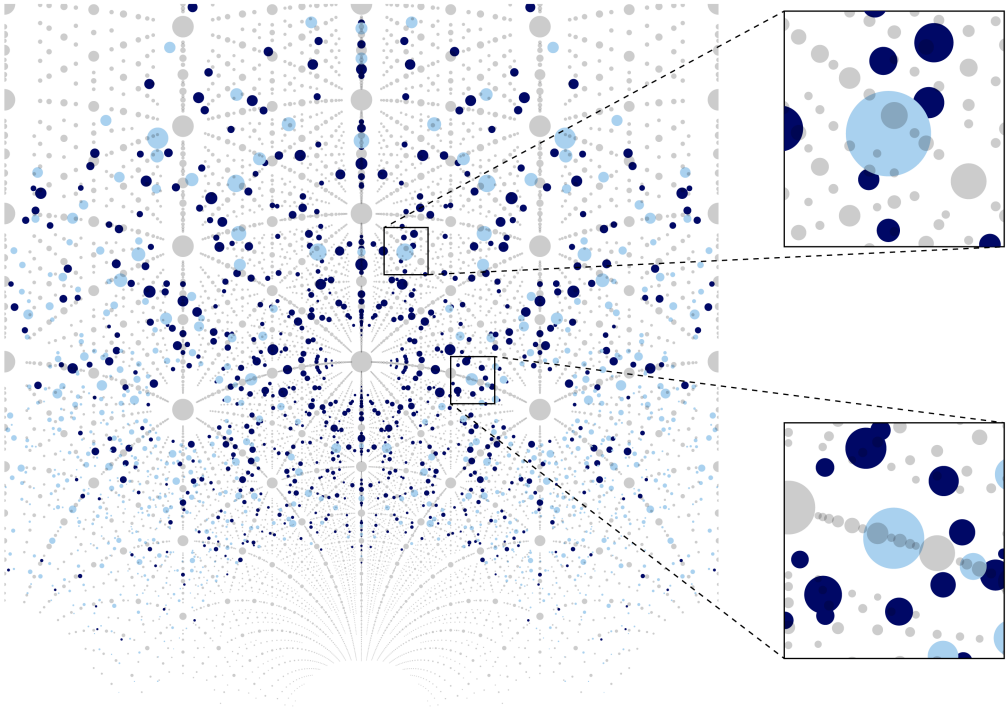
5.2.1. *The importance of* $\mathrm{PSL}(2;\mathbb{Z})$ *and the hyperbolic metric.* There is a natural symmetry of the algebraic numbers in the upper half plane: the action of $\mathrm{PSL}(2;\mathbb{Z})$. That is, the equivariant action of $\mathrm{PSL}(2;\mathbb{C})$ on coefficient and root space, restricted to those elements which preserve the lattice $\mathbb{Z}^3$ in the space $\mathbb{R}^3$ of coefficients (see Section 4.2.5). In this work, we extol the philosophy that, for Diophantine approximation away from the real line, this action should be built into our definitions. Hyperbolic distance and complex distance are conformally equivalent[41], and this symmetry respects the former. Therefore the relative hyperbolic positions of the algebraic numbers are preserved under $\mathrm{PSL}(2;\mathbb{Z})$. Several of the notions of arithmetic complexity in the literature partially respect this symmetry. If they do not, then, given any complex Diophantine approximation statement (such as those of Bugeaud and Evertse), it seems natural to translate by $\mathrm{PSL}(2;\mathbb{Z})$ until the statement is strongest (by which we mean, translate by $\mathrm{PSL}(2;\mathbb{Z})$, find the best approximations according to the theorem, and then transport the constellation back to the original region of interest, where perhaps a priori only worse approximations were guaranteed).

Perhaps our position is most simply stated in terms of our visualizations: the positions of the dots are periodic under $\mathrm{PSL}(2;\mathbb{Z})$ and therefore we argue the sizings of them should be too. We illustrate this in Figure 34, showing the transport of a constellation of algebraic numbers under $\mathrm{PSL}(2;\mathbb{Z})$, but shown in the euclidean metric and the height sizing. The varying sizes and distances of the dots illustrate the changing levels of approximation from the classical perspective.

While $H(f_\alpha) = H(f_{1/\alpha})$, unfortunately $H(f_\alpha) \neq H(f_{\alpha+1})$. Consequently, the naïve height is not invariant under the action of $\mathrm{PSL}(2;\mathbb{Z})$ on α. For non-real numbers, the height does,

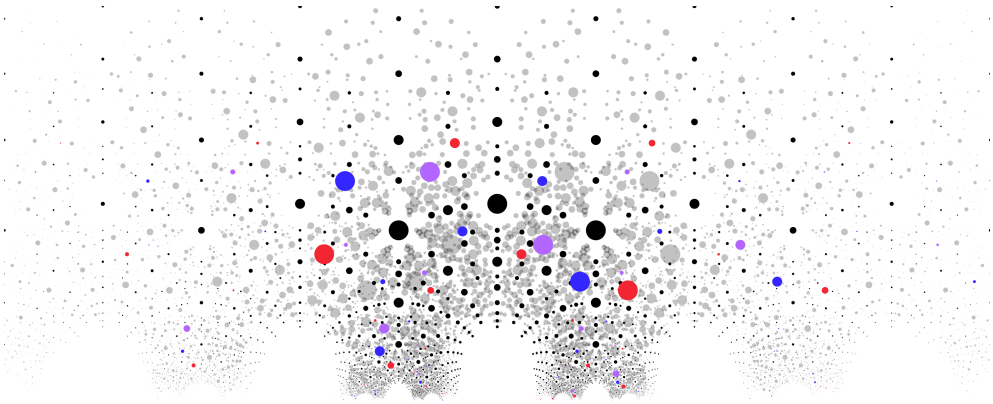
[40] One way to see this is to reduce to the unit circle; cubics cannot lie on the unit circle $\alpha\overline{\alpha} = 1$ unless their real root r is rational, since the constant coefficient of the minimal polynomial $\alpha\overline{\alpha}r$ is rational.

[41] That is, locally the hyperbolic metric and the euclidean metric are very nearly multiples of each other. This becomes exact at the level of tangent spaces for the Riemannian metric, where $ds^2_{\mathrm{Hyp}} = \frac{1}{\mathrm{Im}(z)} ds^2_{\mathrm{Euc}}$.



(A)

FIGURE 33. Quadratics shown in gray, quartics shown in light blue (if they have no real conjugates) and dark blue (if they do). The quartics which lie off such geodesics (such as the one highlighted in the top right) are approximated only by quadratics at some distance on these geodesics. In contrast quartics on a geodesic such as the unit circle have more quadratics near them as shown in the lower highlight.



(A)

FIGURE 34. Cubic algebraic numbers in the naïve height, using the maximum coefficient, plotted in the euclidean metric. $\mathrm{PSL}(2;\mathbb{Z})$ orbits are shown for the complex roots of $x^3 + x - 1$ (red), $x^3 + x^2 - 1$ (purple) and $x^3 + x^2 + x - 1$ (blue).

however, attain a minimum on its full $\mathrm{PSL}(2;\mathbb{Z})$ orbit. Therefore we define[42]

$$H_{\mathrm{PSL}}(f) = \min\{H(f \circ A) : A \in \mathrm{PSL}(2;\mathbb{Z})\}.$$

To compute this minimum, there is a recent algorithm due to Stoll-Cremona and Hutz-Stoll which may[43] be helpful [34, 62].

5.2.2. *Weil height.* A more nuanced generalisation of the notion of height is the Weil height, defined in terms of the absolute values of an ambient number field. For an algebraic number α contained in a number field K, the Weil height is defined[44] as

$$H(\alpha) = \prod_{v \in M_K} \max\{1, ||\alpha||_v\}, \tag{14}$$

where the product is over the set M_K of all *normalized absolute values* $||\alpha||_v = |\alpha|_v^{[K_v:\mathbb{Q}_v]/[K:\mathbb{Q}]}$ of K. Here, K_v and $\mathbb{Q}_v$ are the completions of K and $\mathbb{Q}$ at v. For further details, a nice introduction to this is available in [33, Section B.1–B.2]. This is actually independent of the choice of K containing α. This definition is a generalisation of the case $K = \mathbb{Q}$, namely

$$H(p/q) = \prod_{v \in M_{\mathbb{Q}}} \max\{1, |p/q|_v\}, \tag{15}$$

where $|\cdot|_v$ ranges over all p-adic absolute values, as well as the archimedean one. This case can be more simply and intuitively rewritten as

$$H(p/q) = \max\{|p|, |q|\},$$

if p and q are taken to be coprime and integral. Fortunately, the Weil height and naïve height of its minimal polynomial are closely related by a well known relationship in terms of the degree $d := [\mathbb{Q}(\alpha) : \mathbb{Q}]$ of α [16, Lemma A.2] (note that the Mahler measure satisfies $M(f_\alpha) = H(\alpha)^d$ [10, Proposition 1.6.6]):

$$\binom{d}{\lfloor d/2 \rfloor}^{-1} H(f_\alpha) \leq H(\alpha)^d \leq \sqrt{d+1} H(f_\alpha). \tag{16}$$

Here it is important that f_α is minimal in the sense of coprime integer coefficients.

Note that $H(\alpha) = H(1/\alpha)$. For the same reasons discussed in the previous section, it is natural to define

$$H_{\mathrm{PSL}}(\alpha) = \min\{H(A.\alpha) : A \in \mathrm{PSL}(2;\mathbb{Z})\}.$$

5.2.3. *Repulsion in the complex plane in terms of Weil height.* We can now state a generalization of the repulsion statement (12).

Proposition 5.5 ([10, Theorem 1.5.21])**.** *Suppose $\alpha \neq \beta$ are distinct algebraic numbers, and let $d \geq 1$ be the degree of a field containing both. Then*

$$|\alpha - \beta|^2 \geq \frac{1}{2^d H(\alpha)^d H(\beta)^d}. \tag{17}$$

The exponent 2 on the left can be removed if $\alpha, \beta \in \mathbb{R}$.

See [16, Section A.2] for a version in terms of the naïve height.

[42] Although the definition, and much of the discussion, works for $\mathbb{R}$ as well as $\mathbb{C}$, $\mathrm{PSL}(2;\mathbb{Z})$ doesn't preserve any meaningful metric on the real line, and all of $\mathbb{Q}$ falls into a single orbit: this definition would lose too much information.

[43] It generalizes the reduction theory of binary quadratic forms with respect to $\mathrm{PSL}(2;\mathbb{Z})$ to general binary forms. It is not known how small a height one is guaranteed under their algorithm, nor where the minimum is attained.

[44] We beg the reader's forgiveness for the use of H for both naïve height of a polynomial and Weil height of a number; they do *not* satisfy $H(f_\alpha) = H(\alpha)$, but the notation is standard in the literature.

5.2.4. *Discriminant as a measure of arithmetic complexity.* From the perspective of the geometry discussed in the previous section, and the images we've drawn, one might consider the measure of arithmetic complexity given by the discriminant.

For a polynomial $f = a_d x^d + \cdots + a_1 x + a_0 = a_d \prod_{i=1}^{d}(x - \alpha_i) \in \mathbb{Z}[x]$, let Δ_f denote the discriminant of f. Recall that the discriminant is a measure of the differences between the roots:

$$\Delta_f = a_d^{2d-2} \prod_{i<j} (\alpha_i - \alpha_j)^2.$$

We will refer to an algebraic number as having a discriminant, namely the discriminant of its minimal polynomial, and write $\Delta_\alpha := \Delta_{f_\alpha}$. This has the particular advantage of being invariant under $\mathrm{PSL}(2;\mathbb{Z})$.

How do the previously defined heights and the discriminant relate? Mahler proved a relationship in one direction [44], namely:

$$|\Delta_\alpha| \leq d^d H(\alpha)^{d(2d-2)}. \tag{18}$$

By using (16), we obtain the related inequality:

$$|\Delta_\alpha| \leq d^d (d+1)^{d-1} H(f_\alpha)^{2d-2}. \tag{19}$$

In general one doesn't expect a converse inequality, since the discriminant is invariant under $f(x) \mapsto f(x+1)$, while the Weil height would be expected to grow. Even within one fundamental region of the upper half plane, one doesn't expect a tight relationship. For example, it is possible to define a family of quadratic irrationalities $(\alpha_n)_{n\geq 1}$ such that $H(\alpha_n)^4/|\Delta_{\alpha_n}| \to \infty$ as $n \to \infty$. Namely, the polynomials $x^2 + n$ have upper-half-plane roots α_n approaching ∞ along the imaginary axis, and $4H(\alpha_n)^4/|\Delta_{\alpha_n}| = 4n^2/4n = n \to \infty$.

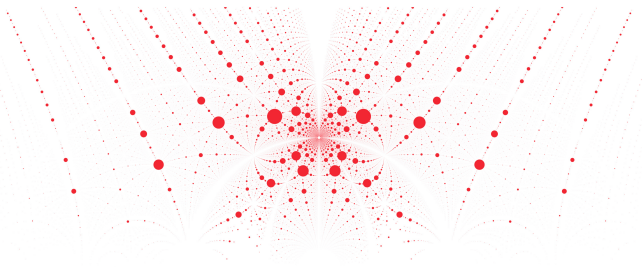
5.2.5. *Sizing in starscape images.* In light of the comparisons (16), (18), and (19), as well as Theorems 5.3 and 5.4, one might compare the naïve sizings in Figures 12 and 13 with slightly more nuanced versions given in Figure 35. These latter sizings are all chosen to match Theorem 5.3 just as Figure 15 matches Dirichlet's Theorem 5.1.

6. Diophantine approximation in the quadratics

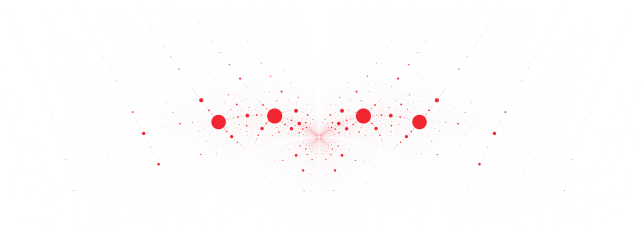
We now revisit the basic theory of Diophantine approximation by complex quadratic irrationalities, from the starscapes perspective: we use the *hyperbolic distance* to measure distance and the *discriminant* to measure arithmetic complexity. We give analogs of repulsion (as in (12) and (17)), Dirichlet's Theorem 5.1 guaranteeing infinitely many approximations, and Roth's Theorem 5.2 on the approximation of algebraic numbers, in this new situation. In this way, we recover the $d = 2$ case of Theorem 5.4, with some additional geometric nuance (we distinguish between approximations coming from different rational geodesics). The proofs are elementary (with the exception that they depend on Schmidt's Subspace Theorem), and are based on the geometry of Section 4. We work in the coefficient space and transport the results to the complex plane afterward.

This perspective offers a few benefits. First, it respects the natural $\mathrm{PSL}(2;\mathbb{Z})$ symmetry. Second, the geometry nicely explains the special cases that arise in Theorem 5.4, and we can now observe that the same dichotomy holds for non-algebraic complex numbers: those on rational geodesics are better approximable. In particular, Theorem 6.3 implies that $k_2(\alpha) \geq 2$ for any complex number on a rational geodesic, whereas Theorem 5.3 says we have $k_2(\alpha) = 3/2$ for almost all complex non-real numbers. Also, our method discriminates between approximations taken from fixed rational geodesics. The relationship between our hyperbolic/discriminant statements and the classical exponents $k_2(\alpha)$ is explained in Section 6.3.

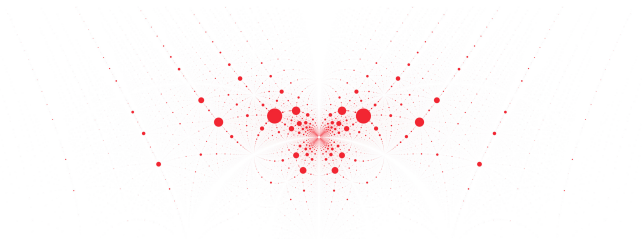
Throughout this section, it will be convenient to use Vinogradov notation: that is, $f \ll_\alpha g$ denotes that f is bounded above by a constant multiple of g, where the constant may depend on α.



(A) Discriminant: $|\Delta_\alpha|^{(d+1)/(4d-4)}$



(B) Naïve height: $H(f_\alpha)^{(d+1)/2}$



(C) Weil height or Mahler measure: $H(\alpha)^{d(d+1)/2} = M(f)^{(d+1)/2}$

FIGURE 35. Several natural choices of sizing by arithmetic complexity of the roots of cubics to match Theorem 5.3, scaled so that the complex root of x^3+x-1 is the same size. The roots plotted are the family ax^3+cx^2+bx+c. Compare with Figures 12 and 13.

6.0.1. *Repulsion amongst imaginary quadratics.* Recall that with the Weil height function we have a simple repulsion principle, (17):

$$|\alpha-\beta|^2 > \frac{1}{2^d H(\alpha)^d H(\beta)^d}.$$

Using the hyperbolic metric (denoted d_{hyp}) and the discriminant, we obtain the following version of repulsion.

Theorem 6.1. *Let $\alpha \neq \beta$ be two non-real quadratic irrationalities, of discriminants Δ_α and Δ_β respectively. Then the hyperbolic distance between α and β, considered in the upper half*

plane, is at least

$$d_{hyp}(\alpha,\beta) \geq \operatorname{acosh}\left(\sqrt{1+\frac{1}{\Delta_\alpha\Delta_\beta}}\right).$$

If $\Delta_\alpha = \Delta_\beta$*, then*

$$d_{hyp}(\alpha,\beta) \geq \operatorname{acosh}\left(1+\frac{1}{|\Delta_\alpha|}\right).$$

Proof. Suppose α and β are associated to some vectors f_α, f_β in coefficient space. By (5), the distance between them is

$$d_{\text{Coefs}}(f_\alpha,f_\beta) = \operatorname{acosh}\left(\frac{|\langle f_\alpha,f_\beta\rangle|}{\sqrt{|\Delta_\alpha\Delta_\beta|}}\right).$$

Since $\alpha \neq \beta$, this distance is greater than 0. However, it lies in

$$\operatorname{acosh}\left(\frac{\mathbb{Z}}{\sqrt{\Delta_\alpha\Delta_\beta}}\right),$$

which is a discrete set of values whose smallest positive value is

$$\operatorname{acosh}\left(\frac{n}{\sqrt{\Delta_\alpha\Delta_\beta}}\right),$$

where n is the smallest integer greater than $\sqrt{|\Delta_\alpha\Delta_\beta|}$. In particular, $n^2 - \Delta_\alpha\Delta_\beta \geq 1$. This implies that

$$\frac{n}{\sqrt{\Delta_\alpha\Delta_\beta}} \geq \sqrt{1+\frac{1}{\Delta_\alpha\Delta_\beta}}.$$

The special case is the case that $n = |\Delta_\alpha| + 1$. □

6.0.2. *Fundamental geometric lemma.* Fix $\alpha \in \mathbb{C}\backslash\mathbb{R}$, not quadratic. We will consider approximations by quadratic irrational β. We can give corresponding vectors in coefficient space:

$$f_\alpha = [\alpha_1 : 1 : \alpha_2], \quad f_\beta = [p_1 : n : p_2]$$

where $p_1, n, p_2 \in \mathbb{Z}$ and α_1, α_2 are not both rational. More precisely, we have $\alpha_1 = 1/(\alpha+\overline{\alpha})$ and $\alpha_2 = (\alpha\overline{\alpha})/(\alpha+\overline{\alpha})$.

We wish to compute the hyperbolic distance between f_α and f_β. Our proofs will rely on a fundamental lemma which relates this distance to linear forms, with coefficients depending on α, in f_β's coordinates.[45]

Lemma 6.2. *Suppose $\alpha \in \mathbb{C}\backslash\mathbb{R}$ is a fixed non-quadratic, with $f_\alpha = [\alpha_1 : 1 : \alpha_2]$.*
Let β be quadratic, with $f_\beta = [p_1 : n : p_2]$, where $n, p_1, p_2 \in \mathbb{Z}$.
Define the linear forms:

$$\begin{aligned} L_1 &:= L_1(p_1,n,p_2) = n\alpha_1 - p_1, \\ L_2 &:= L_2(p_1,n,p_2) = -n\alpha_2 + p_2, \\ L_3 &:= L_3(p_1,n,p_2) = \alpha_1 p_2 - \alpha_2 p_1. \end{aligned}$$

(1) *We have*

$$d_{hyp}(\alpha,\beta) \leq \operatorname{acosh}\left(1+\frac{\max\{|L_1L_2|, L_3^2\}}{|\Delta_\beta|}\right).$$

[45]This generalizes the rational approximation case, since there the angle θ between the projective lines $[\alpha : 1]$ and $[p : q]$ satisfies

$$\cos\theta = \frac{\langle(\alpha,1),(p,q)\rangle}{||(p,q)||||(\alpha,1)||} = \frac{|q\alpha - p|}{||(p,q)||||(\alpha,1)||}.$$

(2) *Suppose β is sufficiently close to α, namely $d_{hyp}(\alpha,\beta) < \operatorname{acosh} 2$. Then*

$$d_{hyp}(\alpha,\beta) \geq \operatorname{acosh}\left(1 + \frac{m_\alpha |L_1 L_2|}{|\Delta_\beta|}\right),$$

for some constant $m_\alpha > 0$ depending only on α.

Proof. Recall that the distance is invariant under scaling, so we will temporarily replace f_α with $f'_\alpha = nf_\alpha$, so the middle coordinates of the two vectors agree. We have

$$\begin{aligned}
&\frac{\langle f'_\alpha, f_\beta\rangle^2}{||f'_\alpha||^2||f_\beta||^2} - 1 \\
&= \frac{(n^2 - 2n\alpha_1 p_2 - 2n\alpha_2 p_1)^2 - (n^2 - 4n^2\alpha_1\alpha_2)(n^2 - 4p_1p_2)}{||f'_\alpha||^2||f_\beta||^2} \\
&= 4n^2 \frac{(\alpha_1 p_2 - \alpha_2 p_1)^2 + (n\alpha_1 - p_1)(n\alpha_2 - p_2)}{||f'_\alpha||^2||f_\beta||^2} \\
&= 4\frac{L_3^2 - L_1L_2}{||f_\alpha||^2||f_\beta||^2}
\end{aligned}$$

Evidently,

$$L_3^2 - L_1L_2 \leq 2\max\{|L_1L_2|, L_3^2\}.$$

Therefore

$$\frac{\langle f_\alpha, f_\beta\rangle^2}{||f_\alpha||^2||f_\beta||^2} \leq 1 + \frac{2\max\{|L_1L_2|, L_3^2\}}{||f_\beta||^2}.$$

Hence,

$$\frac{-\langle f_\alpha, f_\beta\rangle}{||f_\alpha||||f_\beta||} \leq 1 + \frac{\max\{|L_1L_2|, L_3^2\}}{||f_\beta||^2}.$$

Next we show

$$L_3^2 - L_1L_2 \gg_\alpha \min\{|L_1L_2|, L_3^2\}.$$

By the definitions of the L_i, we have

$$L_1\alpha_2 + L_2\alpha_1 = L_3.$$

In particular,

$$L_3^2 - L_1L_2 = (L_1\alpha_2 + L_2\alpha_1)^2 - L_1L_2 \geq (4\alpha_1\alpha_2 - 1)L_1L_2$$

by the arithmetic-geometric mean inequality. Note that $K_\alpha := 4\alpha_1\alpha_2 - 1 > 0$ since α is not real. Thus if L_1L_2 is positive, we are done. On the other hand, if it is negative, then

$$L_3^2 - L_1L_2 = L_3^2 + |L_1L_2| \geq |L_1L_2|.$$

We have shown that, for some constant $m'_\alpha > 0$ depending only on α,

$$\frac{\langle f_\alpha, f_\beta\rangle^2}{||f_\alpha||^2||f_\beta||^2} \geq 1 + \frac{m'_\alpha |L_1L_2|}{||f_\beta||^2}.$$

Therefore, for β sufficiently close to α (i.e., so that $\frac{\langle f_\alpha, f_\beta\rangle^2}{||f_\alpha||^2||f_\beta||^2} < 2$), this implies (taking $m_\alpha = (\sqrt{2}-1)m'_\alpha$) that

$$\frac{-\langle f_\alpha, f_\beta\rangle}{||f_\alpha||||f_\beta||} \geq 1 + \frac{m_\alpha |L_1L_2|}{||f_\beta||^2}.$$

□

6.1. **Quadratic Dirichlet's Theorem.**

6.1.1. *Quadratic Dirichlet's Theorem on a rational geodesic.* We first consider the question of Diophantine approximation on a single *rational geodesic*, i.e. the image of a rational plane in coefficient space (see Section 4.2.5). This is motivated by the observation that each such geodesic looks, in Figure 1a, like a copy of Figure 15, so we expect it to have Diophantine approximation properties similar to the rationals. It will turn out that points lying on such geodesics are better approximable than points elsewhere in $\mathbb{C}$: see Figure 33. Recall that, given $\alpha \in \mathbb{C}\backslash\mathbb{R}$, α lies on a rational geodesic if and only if 1, $\alpha + \overline{\alpha}$ and $\alpha\overline{\alpha}$ are $\mathbb{Q}$-linearly dependent (Observation 4.12).

We begin with an analogue to Dirichlet's Theorem 5.1, asserting the existence of infinitely many good approximations on a rational geodesic.

Theorem 6.3. *Let $\alpha \in \mathbb{C}\backslash\mathbb{R}$ not be quadratic irrational, but lying on a rational geodesic. Then there exists a constant $K_\alpha > 0$, depending only on the $\mathrm{PSL}(2;\mathbb{Z})$ orbit of α, such that there are infinitely many quadratic irrational β lying on that rational geodesic, with*

$$d_{hyp}(\alpha, \beta) \leq \operatorname{arcosh}\left(1 + \frac{K_\alpha}{|\Delta_\beta|^2}\right).$$

Proof. Suppose $\alpha \in \mathbb{C}\backslash\mathbb{R}$ is a fixed non-quadratic, with $f_\alpha = [\alpha_1 : 1 : \alpha_2]$. The theorem statement is invariant under $\mathrm{PSL}(2;\mathbb{Z})$ (i.e. replacing α and all candidate β with their images under the action of some element of $\mathrm{PSL}(2;\mathbb{Z})$, we preserve Δ_β and the hyperbolic distances). We assume α is on a rational geodesic, so $a\alpha_1 + b + c\alpha_2 = 0$. We may translate by $\mathrm{PSL}(2;\mathbb{Z})$ until $\max\{|a|, |b|, |c|\}$ is minimal; this is a constant depending only on the $\mathrm{PSL}(2;\mathbb{Z})$ orbit of α. We can actually choose a canonical geodesic amongst these finitely many in any way we wish; say the one of smallest radius, and amongst those, center nearest the origin but to its right. We set some such convention.

Let $Q > 0$ be an integer.

We use the classical method of proof of Dirichlet's Theorem 5.1 to find a solution $(n_0, q_0) \in \mathbb{Z}^2$ to

$$|n_0\alpha_1 - q_0| \leq 1/Q, \quad n_0 \leq Q.$$

Namely, we divide the unit interval into Q even subintervals, and the box principle guarantees some $i\alpha_1$ and $j\alpha_1$, for some $0 \leq i < j \leq Q$, lie in the same interval modulo $\mathbb{Z}$; we let $n_0 = j - i$. We have $an_0\alpha_1 = -bn_0 - cn_0\alpha_2$, so that

$$|cn_0\alpha_1 - cq_0| \leq |c|/Q, \quad |cn_0\alpha_2 + bn_0 + aq_0| = |-an_0\alpha_1 + aq_0| \leq |a|/Q.$$

We let $f_\beta = [p : n : q] = [cq_0 : cn_0 : -bn_0 - aq_0]$. The corresponding $\beta \in \mathbb{C}$ is a candidate good quadratic irrational approximation to α.

Increasing Q and finding a new approximation, we can in fact produce infinitely many such f_β which are linearly independent, and such that the resulting infinite sequence of distinct β approach α.

It remains to show that these β are indeed good approximations. We have the following observations:

$$n \ll_\alpha Q, \quad |L_1 L_2| \ll_\alpha 1/Q^2,$$
$$L_3^2 = |L_1\alpha_2 + L_2\alpha_1|^2 \ll_\alpha \max\{|L_1|^2, |L_2|^2\} \ll_\alpha 1/Q^2.$$

Here and for the entire proof, the constants in the Vinogradov notation depend on α, but this in the canonical choice of α within its $\mathrm{PSL}(2;\mathbb{Z})$ orbit, so the constants actually only depend on the orbit.

Combining these with Lemma 6.2,

$$\cosh d_{\mathrm{Coefs}}(f_n, f_\beta) - 1 \ll_\alpha \frac{1}{n^2|\Delta_\beta|}.$$

Let $C_\alpha = |\Re(\alpha)|/2|\Im(\alpha)|$. This is a positive constant depending only on α. Then for all β sufficiently close to α, $|\Re(\beta)| > C_\alpha|\Im(\beta)|$. We consider only those solutions β which are at least that close. Then $n/2p > C_\alpha|\Delta_\beta|^{1/2}/2p$, hence $n > C_\alpha|\Delta_\beta|^{1/2}$. Using this fact, we obtain

$$\cosh d_{\text{Coefs}}(f_n, f_\beta) - 1 \ll_\alpha \frac{1}{|\Delta_\beta|^2}.$$

This proves the theorem. □

It is interesting to note that the constant K_α depends on a particular ideal class of a real quadratic field, since by Observation 4.12, it can only lie on one rational geodesic and that geodesic is associated to such an ideal class. In the proof, the constant K_α directly depends on the class. It is natural to wonder about a Lagrange spectrum for α on rational geodesics.

Also interesting is that all α lying on rational geodesics are exceptions to Sprindžuk's Theorem 5.3. It is likely possible to prove using Khintchine's classic methods, that, within a single geodesic, almost all α have $k_2(\alpha) = 2$.

6.1.2. *Quadratic Dirichlet's Theorem in general.* In the general case, where α may not lie on a geodesic, we have weaker approximation guarantee, with a similar proof.

Theorem 6.4. *Let $\alpha \in \mathbb{C}\backslash\mathbb{R}$ not be a quadratic irrational. Let $K > 0$ be any constant. Then, there are infinitely many quadratic irrationalities β with*

$$d_{hyp}(\alpha, \beta) \leq \operatorname{acosh}\left(1 + \frac{K}{|\Delta_\beta|^{3/2}}\right).$$

Proof. Let $Q > 1$ be an integer.

Let $f_i = [i\alpha_1 : i : i\alpha_2] = if_\alpha$ for positive integers i. Divide the unit interval into Q equal subintervals, and consider the vectors f_i modulo $[\mathbb{Z} : \mathbb{Z} : \mathbb{Z}]$. Then, by the pigeonhole principle, there are some $0 \leq i < j \leq Q^2$ such that f_i and f_j have first and third coordinates lying in the same pair of subintervals modulo $\mathbb{Z}$. Let $n := j - i$. Note that $0 < n \leq Q^2$.

By construction, we have $|n\alpha_1 - p_1| < 1/Q$ and $|n\alpha_2 - p_2| < 1/Q$ for some integers p_1 and p_2. Defining $f_\beta = [p_1 : n : p_2]$, we have a corresponding $\beta \in \mathbb{C}$: this is the candidate good approximation we seek.

Choose Q' large enough such that $|n\alpha_1 - p_1|, |n\alpha_2 - p_2| > 1/Q'$ for all $n < Q^2$, $p_1, p_2 \in \mathbb{Z}$. Then, running this argument again with $Q = Q'$, we obtain a new solution f'_β that is linearly independent of f_β. By this method, there are infinitely many such solutions, with β approaching α.

It remains to show that these β are indeed good approximations. We have the following:

$$n \leq Q^2, \quad |L_1 L_2| \leq 1/Q^2, \quad L_3^2 = |L_1\alpha_2 + L_2\alpha_1|^2 \ll_\alpha \max\{|L_1|^2, |L_2|^2\} \leq 1/Q^2.$$

Combining these with Lemma 6.2,

$$\cosh d_{\text{Coefs}}(f_n, f_\beta) - 1 \ll_\alpha \frac{1}{n|\Delta_\beta|}.$$

Note that the theorem statement is invariant under the action of $\mathrm{PSL}(2;\mathbb{Z})$. Therefore, we may assume without loss of generality that $\Re(\alpha) > C\Im(\alpha)$ for any positive constant C, by translation by $\mathbb{Z}$. This implies the same fact about all β sufficiently close to α: that for any fixed positive C, we can guarantee $n/2p > C|\Delta_\beta|^{1/2}/2p$, hence $n > C|\Delta_\beta|^{1/2}$. Using this fact, we obtain

$$\cosh d_{\text{Coefs}}(f_n, f_\beta) - 1 \ll_\alpha \frac{1}{C|\Delta_\beta|^{3/2}}.$$

This proves the theorem. □

6.2. **Poor approximation of algebraic numbers by quadratics.** In this section, we will give a complementary result to Theorems 6.3 and 6.4, showing that algebraic numbers are not any better approximable than those theorems guarantee.

Most of the deep results in the Diophantine approximation of algebraic numbers, including Roth's Theorem 5.2 and Theorem 5.4 of Bugeaud and Evertse, are consequences of the following far-reaching result.

Theorem 6.5 (Schmidt's Subspace Theorem [57, 58])**.** *Let $n \geq 2$, and let $L_1, \ldots, L_n$ be linearly independent linear forms in n variables, with real algebraic coefficients. Let $\epsilon > 0$ be real. Then the solutions $\mathbf{x} \in \mathbb{Z}^n$ to*

$$|L_1(\mathbf{x})L_2(\mathbf{x}) \cdots L_n(\mathbf{x})| < \frac{1}{(\max\{1, |x_1|, \ldots, |x_n|\})^\epsilon}$$

lie in finitely many proper subspaces of $\mathbb{Q}^n$.

As an example, if one takes $\alpha \in \mathbb{R}$ to be algebraic, Roth's Theorem 5.2 can be reformulated as the statement that there are only finitely many solutions to

$$|q||q\alpha - p| < \frac{1}{q^\epsilon}. \tag{20}$$

To recover this assertion from Schmidt's Subspace Theorem, one can choose $L_1(p, q) = q$ and $L_2(p, q) = q\alpha - p$.

Theorem 5.4 of Bugeaud and Evertse uses Schmidt's Subspace Theorem on the space of coefficients. To approximate an algebraic number α, the coefficient vector f is subject to a linear form given by $|f(\alpha)|$ (so the coefficients of the linear form are the powers of α). In our case, the application is different: we work again on the space of coefficient vectors, but our linear form is given in terms of 1, $\alpha + \overline{\alpha}$ and $\alpha\overline{\alpha}$. That is, in terms of the coefficient vector associated to α, instead of a vector of its powers. We are, in effect, searching for approximations to the coefficient vector f_α of α within the coefficient space. It is interesting to ask whether this method extends to higher degree.

We will now use Schmidt's Subspace Theorem 6.5 to deduce the main result of this section.

Theorem 6.6. *Suppose that $\alpha \in \mathbb{C}\backslash\mathbb{R}$ is algebraic and non-quadratic. Let $\eta > 0$. If α lies on a rational geodesic, then there are only finitely many quadratic irrationals β on that geodesic such that*

$$d_{hyp}(\alpha, \beta) \leq \operatorname{acosh}\left(1 + \frac{1}{|\Delta_\beta|^{2+\eta}}\right). \tag{21}$$

Amongst quadratic irrationals β not sharing a rational geodesic with α, there are only finitely many such that

$$d_{hyp}(\alpha, \beta) \leq \operatorname{acosh}\left(1 + \frac{1}{|\Delta_\beta|^{3/2+\eta}}\right). \tag{22}$$

In particular, if α is not on any rational geodesic, then there are only finitely many quadratic irrational β satisfying (22) at all.

Proof. Note that, by Theorem 6.1, at most one β having $\Delta_\beta = \Delta$ can satisfy (21) or (22), for each Δ. Therefore, by throwing away at most finitely many approximations β, we can reduce to considering β having discriminant $|\Delta_\beta|$ exceeding any fixed bound, or, consequently, to considering β sufficiently close to α.

First, we will show that there are only finitely many β which are not on the same rational geodesic as α (where α may or may not be on any rational geodesic), and which satisfy

$$d_{\text{Coefs}}(f_\alpha, f_\beta) \leq \operatorname{acosh}\left(1 + \frac{1}{|\Delta_\beta|^{3/2+\eta}}\right).$$

For β sufficiently close to α, Lemma 6.2 implies that any such solution β satisfies

$$|L_1 L_2| \ll_\alpha 1/|\Delta_\beta|^{1/2+\eta}.$$

For β sufficiently close to α, we also have

$$|\Re(\beta)| \gg\ll_\alpha |\Im(\beta)|. \tag{23}$$

In particular,

$$n \gg\ll_\alpha |\Delta_\beta|^{1/2}.$$

Therefore (altering η),

$$|n||L_1(p_1, n, p_2)||L_2(p_1, n, p_2)| \ll_\alpha \frac{1}{|n|^\eta}.$$

These three linear forms are independent. So by Schmidt's Subspace Theorem 6.5, these solutions lie on finitely many proper subspaces, i.e. rational geodesics. Note that this finite collection of orbits depends only on α. Choose any one geodesic. In particular, assume that $ap_1 + bn + cp_2 = 0$. In particular, $|a|$, $|b|$, $|c|$ are bounded above by a constant depending only on α.

Then, what we have is actually

$$|n\alpha_1 - p_1||cn\alpha_2 + ap_1 + bn| = |n\alpha_1 - p_1||n\alpha_2 - p_2||c| \ll_\alpha \frac{1}{|n|^{1+\eta}} < \frac{1}{|n|^\eta}.$$

Since α is not on the geodesic, $a\alpha_1 + b + c\alpha_2 \neq 0$, which implies these are independent linear forms in two variables n and p_1. Again by Schmidt's Subspace Theorem 6.5, the solutions lie on finitely many lines in (non-projectivized) coefficient space. Hence there are finitely many solutions.

Next, we show that there are only finitely many solutions β which are on the same rational geodesic as α (thus we are in the case that α is on a rational geodesic), and which satisfy

$$d_{hyp}(\alpha, \beta) \leq \operatorname{acosh}\left(1 + \frac{1}{|\Delta_\beta|^{2+\eta}}\right).$$

For such β sufficiently close to α, by Lemma 6.2,

$$|L_1 L_2||\Delta_\beta| \ll_\alpha 1/|\Delta_\beta|^\eta.$$

By the same argument surrounding (23) (altering η),

$$|n|^2|n\alpha_1 - p_1||n\alpha_2 - p_2| \ll_\alpha \frac{1}{|n|^\eta}.$$

Assume that this geodesic is characterised by $ap_1 + b + cp_2 = 0$, and therefore $a\alpha_1 + b + c\alpha_2 = 0$. Therefore $|n\alpha_2 - p_2| = |a/c||n\alpha_1 - p_1|$ and so this becomes (altering η):

$$|n||n\alpha_1 - p_1| \ll_\alpha \frac{1}{|n|^\eta}.$$

These are independent linear forms in two variables, so by Schmidt's Subspace Theorem 6.5 on the two-dimensional space in n and p_1, we have solutions on only finitely many lines in coefficient space. This means there are only finitely many β. □

6.2.1. *Complex Liouville numbers.* We demonstrate that there are non-quadratic complex numbers which are extremely well-approximable by quadratic irrationals (in particular, so as to be necessarily non-algebraic, as a result of Theorem 6.6).

In analogy with the classical case, we will call $\alpha \in \mathbb{C}$ a *complex quadratic Liouville number* if, for every positive integer m, there are infinitely many quadratic irrational β such that

$$d_{hyp}(\alpha, \beta) \leq \operatorname{acosh}\left(1 + \frac{1}{|\Delta_\beta|^m}\right). \tag{24}$$

To accomplish this, we will construct a Cauchy sequence of quadratic irrationals β_k. Calling the limit α, we will show that the hyperbolic distance between α and β_k satisfies (24). The construction is simple: we only require at each stage that

$$d_{hyp}(\beta_k, \beta_{k+1}) < \operatorname{acosh}\left(1 + \frac{1}{|\Delta_{\beta_k}|^k}\right), \quad |\Delta_{\beta_k}| \geq 2. \tag{25}$$

As the quadratic irrationals of absolute discriminant ≥ 2 are dense, this is possible. To see that the sequence is Cauchy and the limit satisfies (24), we can compute, for $M > m$,

$$\begin{aligned} d_{hyp}(\beta_M, \beta_m) &\leq \sum_{k=m}^{M-1} d_{hyp}(\beta_k, \beta_{k+1}) \\ &\leq d_{hyp}(\beta_m, \beta_{m+1}) \sum_{k=m}^{M-1} \frac{d_{hyp}(\beta_k, \beta_{k+1})}{d_{hyp}(\beta_m, \beta_{m+1})} \\ &\ll d_{hyp}(\beta_m, \beta_{m+1}) \\ &\leq \operatorname{acosh}\left(1 + \frac{1}{|\Delta_{\beta_m}|^m}\right). \end{aligned}$$

Note that the step in which the sum symbol disappears follows from (25) and the series expansion as $x \to \infty$,

$$\operatorname{acosh}(1 + 1/x^2) = \frac{\sqrt{2}}{x} - \frac{1}{6\sqrt{2}x^3} + \frac{3}{80\sqrt{2}x^5} + \cdots. \tag{26}$$

This implies (24).

As this construction has a great deal of freedom, we can construct quadratic Liouville numbers on any fixed rational geodesic, for example. The countability of the rational geodesics also implies we can construct quadratic Liouville numbers not lying on any rational geodesic.[46]

6.3. **Comparison with classical results.** We wish to compare Theorems 6.4 and 6.6 with Theorem 5.4. In other words, to compare the classical approach to Diophantine approximation in the complex plane with the approach we consider here, using hyperbolic metric and discriminant sizing.

We begin with the following Lemma 6.7, relating the naïve height and the discriminant as measures of arithmetic complexity. In the case of quadratics, and taking into account $\mathrm{PSL}(2;\mathbb{Z})$ invariance, (19) becomes

$$|\Delta_\alpha| \leq 12 H_{\mathrm{PSL}}(f_\alpha)^2. \tag{27}$$

As discussed above, we don't expect an inequality in the other direction in general. However, if we take into account the $\mathrm{PSL}(2;\mathbb{Z})$ action, and loosen the exponents, then we can prove such a thing in the quadratic case.

Lemma 6.7. *Suppose β is of degree 2. Then*

$$|\Delta_\beta| \geq 3 H_{\mathrm{PSL}}(f_\beta), \tag{28}$$

and also,

$$|\Delta_\beta| \geq \frac{3 H_{\mathrm{PSL}}(f_\beta)^2}{N_{\mathrm{PSL}}(\beta)}, \tag{29}$$

where N_{PSL} denotes the norm of the element of its $\mathrm{PSL}(2;\mathbb{Z})$ orbit which lies in the standard fundamental region.

[46] To see this, count the geodesics and assign a tubular neighbourhood of width $1/n$ to the n-th such geodesic; require all terms β_k past the n-th to avoid the first n tubular neighbourhoods; each term has finitely many restrictions placed upon it. (To make this work, the geodesics must be ordered as the terms are created, i.e. the n-th geodesic is always chosen so that the closure of its neighbourhood does not include β_n.)

Proof. The quantities involved are all invariant under the action of $\mathrm{PSL}(2;\mathbb{Z})$, so it suffices to assume β lies in the usual $\mathrm{PSL}(2;\mathbb{Z})$ fundamental region. For β an imaginary quadratic irrationality satisfying $ax^2+bx+c=0$, lying in the usual fundamental region,

$$H(f_\beta) = \max\{|a|,|b|,|c|\}, \quad N(\beta) = |c/a| \geq 1, \quad |\Re(\beta)| = |b/2a| \leq 1/2.$$

From this we conclude that $|b| \leq |a| \leq |c|$ hence $H(f_\beta) = |c|$ and $b^2 < |ac|$. Then we have

$$|\Delta_\beta| = 4|ac| - b^2 \geq 3|ac| \geq 3|c| = 3H(f_\beta).$$

Alternately, we use $N(\beta) = |c/a|$ for the second inequality. □

Next, we recall that the hyperbolic metric and euclidean metric are locally conformally equivalent. More precisely, for β sufficiently close to α in either metric,

$$(1-\epsilon)\Im(\alpha)d_{hyp}(\alpha,\beta) < |\alpha-\beta| < (1+\epsilon)\Im(\alpha)d_{hyp}(\alpha,\beta). \tag{30}$$

Finally, we need the series expansion as $x \to \infty$ for the inverse hyperbolic cosine given in (26).

Collecting the above relationships, it is a simple computation to derive the following:

Lemma 6.8. (1) *Suppose*

$$d_{hyp}(\alpha,\beta) < \operatorname{acosh}\left(1 + \frac{C}{|\Delta_\beta|^k}\right).$$

Then

$$|\alpha-\beta| < \frac{(1+\epsilon)\Im(\alpha)N_{\mathrm{PSL}}(\beta)^{k/2}\sqrt{2C}}{3^{k/2}H_{\mathrm{PSL}}(f_\beta)^k}.$$

(2) *Suppose*

$$d_{hyp}(\alpha,\beta) > \operatorname{acosh}\left(1 + \frac{C}{|\Delta_\beta|^k}\right).$$

Then for $|\Delta_\beta|$ sufficiently large,

$$|\alpha-\beta| > \frac{(1-\epsilon)\Im(\alpha)\sqrt{2C}}{12^{k/2}H_{\mathrm{PSL}}(f_\beta)^k}.$$

In particular, Theorems 6.1, 6.3, 6.4, and 6.6 imply statements in the euclidean metric with the choice of the naïve height, which we collect here for completeness.

Theorem 6.9. (1) *Let $\alpha \neq \beta$ be two non-real quadratic irrationalities. Then for and positive ϵ, and for $|\Delta_\beta|$ sufficiently large,*

$$|\alpha-\beta| \geq \frac{(1-\epsilon)\Im(\alpha)\sqrt{2}}{2\sqrt{3}H_{\mathrm{PSL}}(f_\beta)^{1/2+\epsilon}H_{\mathrm{PSL}}(f_\alpha)^{1/2+\epsilon}}.$$

If $\Delta_\alpha = \Delta_\beta$ is sufficiently large in absolute value, then

$$|\alpha-\beta| \geq \frac{(1-\epsilon)\Im(\alpha)\sqrt{2}}{2\sqrt{3}\min\{H_{\mathrm{PSL}}(f_\beta), H_{\mathrm{PSL}}(f_\alpha)\}}.$$

(2) *Let $\alpha \in \mathbb{C}\backslash\mathbb{R}$ not be quadratic irrational, but lying on a rational geodesic. Then there exists a constant $K_\alpha > 0$, depending only on the $\mathrm{PSL}(2;\mathbb{Z})$ orbit of α, such that there are infinitely many quadratic irrational β lying on that rational geodesic, with*

$$|\alpha-\beta| < \frac{K_\alpha}{H_{\mathrm{PSL}}(f_\beta)^2}.$$

(3) *Let $\alpha \in \mathbb{C}\backslash\mathbb{R}$ not be a quadratic irrational. Let $K > 0$ be any constant. Then, there are infinitely many quadratic irrationalities β with*

$$|\alpha-\beta| < \frac{K}{H_{\mathrm{PSL}}(f_\beta)^{3/2}}.$$

(4) *Suppose that $\alpha \in \mathbb{C}\backslash\mathbb{R}$ is algebraic and non-quadratic. Let $\eta > 0$. If α lies on a rational geodesic, then there are only finitely many quadratic irrationals β on that geodesic such that*

$$|\alpha - \beta| < \frac{1}{H_{\mathrm{PSL}}(f_\beta)^{2+\eta}}. \tag{31}$$

Amongst quadratic irrationals β not sharing a rational geodesic with α, there are only finitely many such that

$$|\alpha - \beta| < \frac{1}{H_{\mathrm{PSL}}(f_\beta)^{3/2+\eta}}. \tag{32}$$

In particular, if α is not on any rational geodesic, then there are only finitely many quadratic irrational β satisfying (22) at all. Note that in (21) and (22), H_{PSL} can be weakened to H.

In particular, up to constants depending on α, our theorems recover the computation of $k_2(\alpha)$ given by Theorem 5.4 of Bugeaud and Evertse for the quadratic case.[47] Theorem 6.9 offers some refinement in terms of drawing a distinction between approximations on a geodesic containing α or not.

However, the constants are different. As the imaginary part of α approaches ∞, the constant K_α of item (2) of Theorem 6.9 weakens. By contrast item (1) of Theorem 6.9 becomes a stronger statement as $\alpha \to \infty$. It is natural to ask whether there is an analogue to the Lagrange spectrum for approximation by quadratic irrationals, at least for those lying on a geodesic. The preceding discussion demonstrates that knowledge of the spectrum in one of the classical or hyperbolic/discriminant settings would not imply the other, although there would be some relationships.

7. To boldly go where no one has gone before

The investigation of algebraic starscapes raises a wide variety of possible future research directions, many of which we intend to continue to investigate. We invite you to join us.

7.1. **The homogeneous geometry of higher degrees.** For higher degree polynomials, the beginnings of the geometric story remain relatively unchanged, but strong conclusions such as Theorems 4.3 and 4.27 become more complicated to draw, as the dimension of the space of polynomials grows.

In particular, fixing a degree $n \geq 1$, we have that $\mathbb{P}\mathrm{Coefs}_n \cong \mathbb{CP}^n$ identifies with complex projective space, and $\mathrm{Roots}_n \cong \mathrm{SP}^n(\mathbb{CP}^1)$ is the set of unordered n-tuples in the Riemann sphere. The roots map $\mathcal{R}_n \colon \mathbb{CP}^n \to \mathrm{SP}^n(\mathbb{CP}^1)$ is a homeomorphism, and is equivariant with respect to the natural $\mathrm{PSL}(2;\mathbb{C})$ actions on each side: given on the space of roots by precomposition with a Möbius transformation, and on the space of coefficients by the action of the unique irreducible representation $\mathrm{PSL}(2;\mathbb{C}) \to \mathrm{PSL}(n;\mathbb{C})$.

Restricting to real coefficients, the fact that $\mathcal{R}$ is a homeomorphism implies $\mathcal{R}_n \colon \mathbb{RP}^n \to \mathrm{SP}^n(\mathbb{CP}^1)$ is an embedding, equivariant with respect to the restricted $\mathrm{PSL}(2;\mathbb{R})$ action on each side. The orbits of this action decompose the space of degree n polynomials, whose nature depends on the degree. For $n \leq 3$, each component of the complement of the discriminant locus comprises an entire orbit itself, and thus comes equipped with the structure of a homogeneous geometry for $\mathrm{PSL}(2;\mathbb{R})$. For $n > 3$, the $\mathrm{PSL}(2;\mathbb{R})$ orbits foliate each component. Consider as an example the space of quartics. The complement of the discriminant locus has two components; those with two pairs of complex conjugate roots, and those with a pair of complex conjugate roots and two real roots. The $\mathrm{PSL}(2;\mathbb{R})$ action decomposes the latter into

[47] To see this for items (2) and (3) in Theorem 6.9 requires some consideration of the relationship between $H_{\mathrm{PSL}}(f_\alpha)$ and $H(f_\alpha)$; we need to know the quotient is bounded in terms of the height of the relevant element of $\mathrm{PSL}(2;\mathbb{Z})$, which is a constant in terms of α.

a family of codimension-1 hypersurfaces, which are generically[48] diffeomorphic to $\mathrm{PSL}(2;\mathbb{R})$ itself. This complicates the overall picture, and suggests generalisations of Theorems 4.3 and 4.27 will involve fiber bundles of homogeneous spaces, rather than just the spaces themselves.

Building a more robust geometric toolkit would not only allow the extension of these ideas to higher degree polynomials, but strong enough tools may provide a window into exploring their solvability. In particular, while the next case of interest, quartics, admits a solution by radicals, the quintics and beyond do not. It is an exciting prospect to try and understand this dichotomy geometrically along our journey.

7.2. **Starscape curves.** Revisit Figure 25. The images of rational planes (projective rational lines) in the coefficient space form delicate beaded necklaces (linear starscapes). Any two algebraic numbers lie on at least one common curve (more if the point has the non-transversal property discussed in Corollary 4.30). It is possible to give algebraic equations for these curves in general dimension, in terms of the two associated minimal polynomials f_1 and f_2, namely, viewing the complex plane as $\mathbb{R}^2$, the curve is $\Re(f_1(x+iy))\Im(f_2(x+iy)) = \Re(f_2(x+iy))\Im(f_1(x+iy))$. We will call these *starscape curves*. The intricacies of some of these curves suggest that the projective geometry of coefficient space is quite disguised by projecting onto complex roots. Determining whether these starscape curves are related to any natural geometric structures on the complex plane may allow us to develop further extensions of the material in Section 4 to higher degree.

Furthermore, starscape curves appear to have a repulsion effect all their own (see the white-space surrounding geodesics in Figure 1a, for example). Is it possible to quantify how well approximable some complex number α is by these curves? We might measure the distance between α and a curve. What is the *height* of a starscape curve?

7.3. **Planar starscapes.** The planar starscapes we have studied provide a two dimensional analogue to the curves described above, with the complex quadratics (Figure 1a) providing a primary example. As the degree of polynomials increases these families can get more complicated, but the projection to the upper half-plane by complex roots is always available (although each polynomial might be represented by increasing numbers of individual roots). Even in the cubic case where there is at most one complex pair we start to see new behaviour with the "blackhole"-like phenomena seen on the right of Figure 32b, where a quadratic point is the complex root of many polynomials. The starscapes shown in Figures 2d, 2e, and 2f show further intriguing behaviours that might be studied, such as seemingly denser regions and the isolated quadratic and cubic points at the top middle of Figure 2f.

7.4. **Algorithms to draw starscapes and Farey structure.** Some of the images here have taken minutes or even hours to compute. Can we develop faster, more intelligent algorithms? The current images are made using a brute force approach: generate a large number of polynomials, solve them, and then plot the results which fit into a desired region. The code is short and many of the hard subproblems, like solving polynomials, are already implemented in computer algebra environments. However, it is very inefficient: we don't effectively sieve for points which will end up in the desired region, and work is repeated for every point.

The geometry discussed above provides a path to improvements. Can we efficiently predict which polynomials need to be solved for a given region? Along each curve we see a recursive pattern resembling the rational numbers in Figure 15. An efficient way to generate rational numbers is to use the Farey (or Stern-Brocot) tree stucture, discussed in Section 4.1.2 and shown in Figure 18: in essence, use the mediant operation to fill in gaps, recursively. A variation on

[48]Indeed, by the action of $\mathrm{PSL}(2;\mathbb{R})$, the pair of complex roots can be moved to any point in $\mathbb{H}^2$. Fixing this point, the remaining degree of freedom of the $\mathrm{PSL}(2;\mathbb{R})$ action acts by rotation on the ideal boundary $\partial_\infty\mathbb{H}^2 = \mathbb{RP}^1$, and each $\mathrm{PSL}(2;\mathbb{R})$ orbit is determined by the *angle* between these two points, measured between the tangent vectors at the complex root pointing to the real roots, as in Equation 10.

this can be achieved by the naïve addition of polynomials as coefficient vectors. This produces a new polynomial with a root "between" the original two along the curve. In effect, we are asking about higher dimensional Stern-Brocot trees [40]. It should also be possible to make use of the $\mathrm{PSL}(2;\mathbb{Z})$ symmetry, at least in large-scale pictures.

7.5. **Continued fractions.** Do there exist continued fraction algorithms for approximation by algebraic numbers of fixed or bounded degree? The Farey structure of the rationals is, in some sense, the source of the continued fraction algorithm for real numbers approximated by rationals (the continued fraction algorithm can be viewed both as a traversal of a Cayley graph for $\mathrm{PSL}(2;\mathbb{Z})$ and as a geometric process of repeated mediants). Figure 1a wonders aloud, might it be possible to extend the continued fraction method to approximate complex numbers by quadratic irrationals? It is possible that existing multidimensional continued fraction algorithms, which have a long history, may provide hints; a recent article from which to enter the literature is [47].

7.6. **Higher degree and geometry-sensitive Diophantine approximation.** We have seen that the geometry of the roots map naturally classifies certain types of approximations. Our example is quadratic approximations from one rational geodesic, where we saw that approximations from a geodesic containing α can be better than approximations not on the geodesic. Moving to the cubic case, one might ask how well a quadratic is approximated by cubics from a particular planar starscape. If the quadratic is a singular point in the sense of Corollary 4.30 and Figure 32, then we conjecture the cubics from that planar starscape are better approximations of the quadratic than those from starscapes not having the property.

More generally, to what extent can the results of Section 6 be extended to higher degree? Could such an extension settle the oustanding cases in the work of Bugeaud and Evertse? In higher degree, are the exceptionally well-approximable algebraic numbers all living on starscape curves, or are there other reasons to be well-approximable? The work of Bugeaud and Evertse also indicates a difference in approximability based on how many real conjugates an algebraic number has. What pictures would one draw to see this effect?

7.7. **Mahler measure and Lehmer's Conjecture.** Lehmer's famous conjecture is related to another important measure of arithmetic complexity, namely the *Mahler measure* (which is not technically any type of measure). A polynomial $f = a_d x^d + \cdots a_1 x + a_0 = a_d \prod_{i=1}^{d} (x - \alpha_i) \in \mathbb{Z}[x]$ has Mahler measure

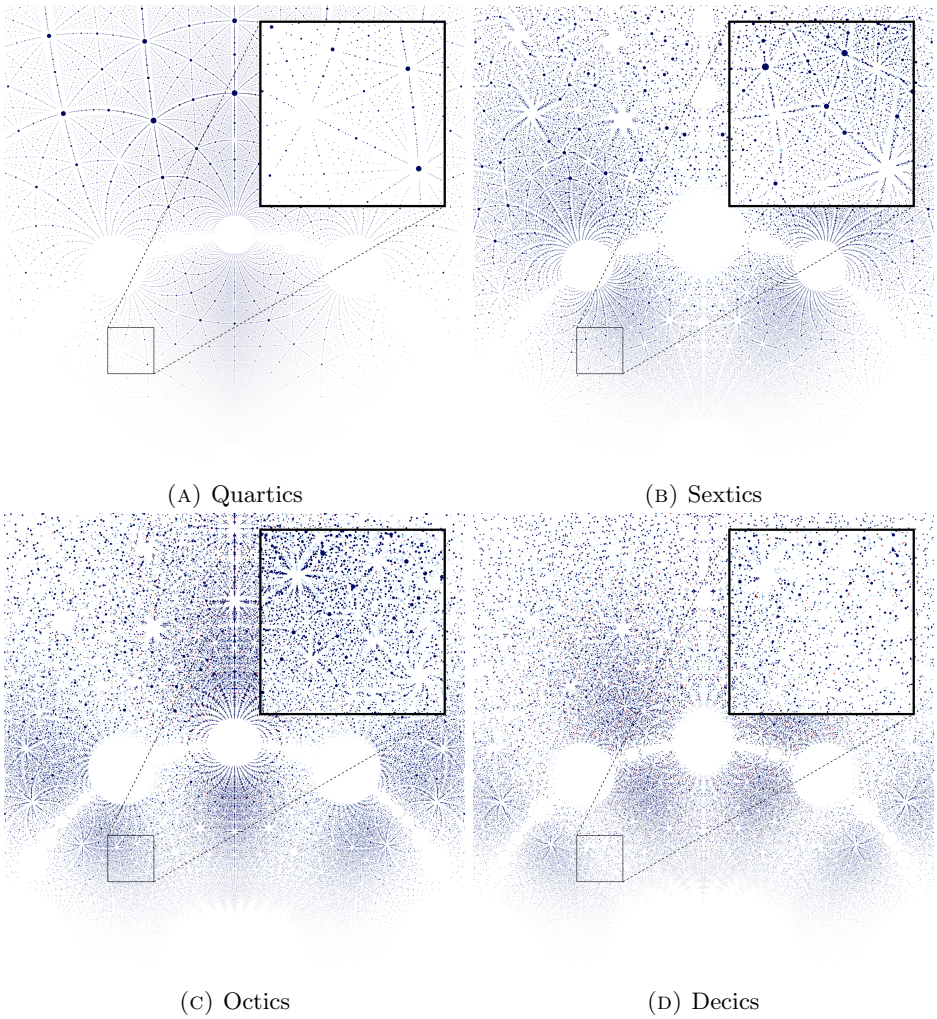
$$M(f) = |a_d| \prod_{i=0}^{d} \max\{1, |\alpha_i|\}.$$

The Weil height of an algebraic number is exactly related to the Mahler measure of its minimal polynomial: $M(f_\alpha) = H(\alpha)^d$.

Lehmer's conjecture states that there is a lower bound to $M(f)$ away from polynomials whose roots are roots of unity. The smallest known Mahler measure of a non-root-of-unity is called *Salem's number*, $1.176280818\ldots$ associated to the roots of the *Lehmer polynomial* $x^{10} + x^9 - x^7 - x^6 - x^5 - x^4 - x^3 + x + 1$.[49] Lehmer's Conjecture is known to hold for non-reciprocal polynomials (those whose coefficients are not palindromic) [60], and Voutier [65] gives an explicit lower bound in terms of the degree, implying the conjecture is true if fields are restricted by degree. Therefore to properly visualize Lehmer's conjecture would require a starscape in increasing degree, and some understanding of the real roots accompanying the complex roots.

For these various reasons, there is a sense in which Lehmer's conjecture does not properly belong to the complex plane and the algebraic starscapes, although it is natural to wonder

[49] The complex roots of Lehmer's polynomial all lie on the unit circle, and its Mahler measure is a function of its real roots.



(A) Quartics (B) Sextics

(C) Octics (D) Decics

FIGURE 36. Pictures of roots of reciprocal polynomials off the unit circle (the many roots on the unit circle are not plotted). For odd degrees the reciprocal polynomials always have the root -1. The points are coloured by the number of conjugate real roots with dark blue for 0, light blue for 2, dark red for 4 and light red for 6 (dark and light blue are swapped from Figure 33 to make the image clearer). Note the strong repulsion of the unit circle in all cases, especially close to i and third/sixth roots of unity. There are also denser regions in the direction of the d-th roots of unity for degree d.

if it gives rise to any interesting images. One related image which is particularly stunning is the reciprocal polynomials, shown in Figure 36. The image itself asks a variety of interesting questions.

REFERENCES

[1] Dan Anderson. https://twitter.com/dandersod/status/1264918722155876353, May 2020. Accessed: 2020-08-09.
[2] Dan Anderson. https://twitter.com/dandersod/status/1265007977964228609, May 2020. Accessed: 2020-08-09.

[3] J.W. Anderson. *Hyperbolic Geometry*. Springer undergraduate mathematics series. Springer, 1999.

[4] John Baez. Algebraic Numbers. https://blogs.ams.org/visualinsight/2013/09/01/algebraic-numbers/, September 2013. Accessed: 2020-08-09.

[5] John Carlos Baez. The Beauty of Roots. https://johncarlosbaez.wordpress.com/2011/12/11/the-beauty-of-roots/. Accessed: 2020-08-09.

[6] John Carlos Baez. The Beauty of Roots (Part 2). https://johncarlosbaez.wordpress.com/2012/01/07/the-beauty-of-roots-part-2/. Accessed: 2020-08-09.

[7] David H. Bailey, Jonathan M. Borwein, Neil J. Calkin, Roland Girgensohn, D. Russell Luke, and Victor H. Moll. *Experimental mathematics in action*. A K Peters, Ltd., Wellesley, MA, 2007.

[8] Andrej Bauer. TedEx "Zeroes". http://math.andrej.com/2014/10/16/tedx-zeroes/. Accessed: 2020-08-09.

[9] Jwalin Bhatt. Field of Algebraic numbers in C. https://www.youtube.com/watch?v=8NDfMxw49co, July 2018. Accessed: 2020-08-09.

[10] Enrico Bombieri and Walter Gubler. *Heights in Diophantine geometry*, volume 4 of *New Mathematical Monographs*. Cambridge University Press, Cambridge, 2006.

[11] Peter Borwein. *Computational excursions in analysis and number theory*, volume 10 of *CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC*. Springer-Verlag, New York, 2002.

[12] Peter Borwein, Tamás Erdélyi, and Friedrich Littmann. Polynomials with coefficients from a finite set. *Transactions of the American Mathematical Society*, 360:5145–5154, 2008.

[13] Peter Borwein and Loki Jörgenson. Visible structures in number theory. *Amer. Math. Monthly*, 108(10):897–910, 2001.

[14] Harrison Bray, Diana Davis, Kathryn Lindsey, and Chenxi Wu. The shape of thurston's master teapot, 2019.

[15] Stephen Brooks. Visualisation of the (countable) field of algebraic numbers in the complex plane. https://en.wikipedia.org/wiki/File:Leadingcoeff.png, March 2010. Accessed: 2020-08-09.

[16] Yann Bugeaud. *Approximation by algebraic numbers*, volume 160 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 2004.

[17] Yann Bugeaud and Jan-Hendrik Evertse. Approximation of complex algebraic numbers by algebraic numbers of bounded degree. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)*, 8(2):333–368, 2009.

[18] James W. Cannon, Floyd William J., Richard Kenyon, and Walter R. Parry. Hyperbolic geometry. *Flavors of Geometry MSRI Publications*, 31, 1996.

[19] Eunice Y.S. Chan, Robert M. Corless, Laureano Gonzalez-Vega, J. Rafael Sendra, Juana Sendra, and Steven E. Thornton. Upper hessenberg and toeplitz bohemians. *Linear Algebra and its Applications*, 601:72 – 100, 2020.

[20] Dan Christiansen. Plots of roots of polynomials with integer coefficients. http://jdc.math.uwo.ca/roots/. Accessed: 2020-08-09.

[21] David A. Cox. *Primes of the form $x^2 + ny^2$*. Pure and Applied Mathematics (Hoboken). John Wiley & Sons, Inc., Hoboken, NJ, second edition, 2013. Fermat, class field theory, and complex multiplication.

[22] H. Davenport and Wolfgang M. Schmidt. Approximation to real numbers by quadratic irrationals. *Acta Arith.*, 13:169–176, 1967/68.

[23] L. G. P. Dirichlet. Verallgemeinerung eines satzes aus der lehre von den kettenbrüchen nebst einige anwendungen auf die theorie der zahlen. *S.-B. Preuss. Akad. Wiss.*, pages 93–95, 1842.

[24] Blaženka Divjak, Zlatko Erjavec, Barnabás Szabolcs, and Brigitta Szilágyi. Geodesics and geodesic spheres in $\widetilde{\mathrm{SL}(2,\mathbb{R})}$ geometry. *Math. Commun.*, 14(2):413–424, 2009.

[25] David Dumas. *SL(View)*, 2020. https://dumas.io/slview/.

[26] Greg Egan. Littlewood applet. http://www.gregegan.net/SCIENCE/Littlewood/Littlewood.html. Accessed: 2020-08-09.

[27] Jordan S. Ellenberg. What do roots of random polynomials look like? https://quomodocumque.wordpress.com/2010/01/09/what-do-roots-of-random-polynomials-look-like/. Accessed: 2020-08-09.

[28] Bernat Espigulé. The beauty of roots. http://gallery.bridgesmathart.org/exhibitions/2019-icerm-illustrating-mathematics/geonat. Accessed: 2020-08-09.

[29] Fauxtographique. Algebraic Numbers of Degree 2. https://www.deviantart.com/fauxtographique/art/Algebraic-Numbers-of-Degree-2-384042962. Accessed: 2020-08-09.

[30] Matt (https://mathematica.stackexchange.com/users/1735/matt-groff) Groff. How can we plot the complex roots of an equation? Mathematica & Wolfram Language Stack Exchange. URL:https://mathematica.stackexchange.com/questions/63028/how-can-we-plot-the-complex-roots-of-an-equation/63064#63064 (version: 2020-06-12).

[31] Edmund Harriss. https://twitter.com/Gelada/status/1263599901784621056, May 2020. Accessed: 2020-08-09.

[32] Edmund Harriss, Pierre Arnoux, Kate Stange, and Steve Trettel. Algebraic starscapes. Art Exhibit at Gaukurinn, Reykjavik, Iceland, February 2020.

[33] Marc Hindry and Joseph H. Silverman. *Diophantine geometry*, volume 201 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2000. An introduction.
[34] Benjamin Hutz and Michael Stoll. Smallest representatives of $\mathrm{SL}(2,\mathbb{Z})$-orbits of binary forms and endomorphisms of $\mathbb{P}^1$. *Acta Arith.*, 189(3):283–308, 2019.
[35] iadvd (https://math.stackexchange.com/users/189215/iadvd). What methods are known to visualize patterns in the set of real roots of quadratic equations? Mathematics Stack Exchange. URL:https://math.stackexchange.com/q/1412264 (version: 2020-06-12).
[36] A. Ya. Khintchine. *Continued fractions*. Translated by Peter Wynn. P. Noordhoff, Ltd., Groningen, 1963.
[37] J. F. Koksma. Über die Mahlersche Klasseneinteilung der transzendenten Zahlen und die Approximation komplexer Zahlen durch algebraische Zahlen. *Monatsh. Math. Phys.*, 48:176–189, 1939.
[38] J.M. Lee. *Introduction to Riemannian Manifolds*. Graduate Texts in Mathematics. Springer International Publishing, 2019.
[39] Adrien-Marie Legendre. *Essai sur la théorie des nombres*. Cambridge Library Collection. Cambridge University Press, Cambridge, 2009. Reprint of the second (1808) edition.
[40] Hakan Lennerstad. The n-dimensional Stern-Brocot tree. *Int. J. Number Theory*, 15(6):1219–1236, 2019.
[41] Jonathan Lidbeck. Complex polynomial roots. https://www.flickr.com/photos/jondissed/37297981445. Accessed: 2020-08-09.
[42] Joseph Liouville. Remarques relatives à des classes très-étendues de quantités dont la valeur n'est ni algébrique, ni même réductible à des irrationnelles algébriques. *C. R. Acad. Sci. Paris*, 18:883–885, 1844.
[43] John E. Littlewood. *Some problems in real and complex analysis*. D. C. Heath and Co. Raytheon Education Co., Lexington, Mass., 1968.
[44] K. Mahler. An inequality for the discriminant of a polynomial. *Michigan Math. J.*, 11:257–262, 1964.
[45] David M. Marciel. Visualizing the patterns in the sets of complex and real roots of quadratic and cubic equations. http://hobbymaths.blogspot.com/2015/09/visualizing-patterns-in-sets-of-complex.html. Accessed: 2020-08-09.
[46] John Matson. Polynomial Plot: Simple Math Expressions Yield Intricate Visual Patterns [Slide Show]. https://www.scientificamerican.com/article/math-polynomial-roots/, December 2009. Accessed: 2020-08-09.
[47] Nadir Murru. Linear recurrence sequences and periodicity of multidimensional continued fractions. *Ramanujan J.*, 44(1):115–124, 2017.
[48] Paul Nylander. polynomial roots. https://nylander.wordpress.com/2008/12/29/polynomial-roots/. Accessed: 2020-08-09.
[49] A. M. Odlyzko and B. Poonen. Zeros of polynomials with $0, 1$ coefficients. *Enseign. Math. (2)*, 39(3-4):317–348, 1993.
[50] Vincent Pantaloni. Map of the complex roots. http://prof.pantaloni.free.fr/IMG/pdf/Root-map.pdf. Accessed: 2021-11-01.
[51] Michael Pershan. https://twitter.com/mpershan/status/1264931366275293190, May 2020. Accessed: 2020-08-09.
[52] Clifford A. Pickover. *The mathematics of Oz*. Cambridge University Press, Cambridge, 2002. Mental gymnastics from beyond the edge.
[53] J. Richter-Gebert. *Perspectives on Projective Geometry: A Guided Tour Through Real and Complex Geometry*. Springer Berlin Heidelberg, 2011.
[54] K. F. Roth. Rational approximations to algebraic numbers. *Mathematika*, 2:1–20; corrigendum, 168, 1955.
[55] M. S. cubics lead coef 3 more. https://www.flickr.com/photos/104348204@N05/11557321664/. Accessed: 2020-08-09.
[56] Wolfgang M. Schmidt. Simultaneous approximation to algebraic numbers by rationals. *Acta Math.*, 125:189–201, 1970.
[57] Wolfgang M. Schmidt. Norm form equations. *Ann. of Math. (2)*, 96:526–551, 1972.
[58] Wolfgang M. Schmidt. *Diophantine approximation*, volume 785 of *Lecture Notes in Mathematics*. Springer, Berlin, 1980.
[59] Pablo Shmerkin and Boris Solomyak. Zeros of {-1, 0, 1} power series and connectedness loci for self-affine sets. *Experimental Mathematics*, 15:499 – 511, 2006.
[60] C. J. Smyth. On the product of the conjugates outside the unit circle of an algebraic integer. *Bull. London Math. Soc.*, 3:169–175, 1971.
[61] V. G. Sprindžuk. *Mahler's problem in metric number theory*. Translated from the Russian by B. Volkmann. Translations of Mathematical Monographs, Vol. 25. American Mathematical Society, Providence, R.I., 1969.
[62] Michael Stoll and John E. Cremona. On the reduction theory of binary forms. *J. Reine Angew. Math.*, 565:79–99, 2003.
[63] The Sage Developers. *SageMath, the Sage Mathematics Software System (Version 9.1)*, 2020. https://www.sagemath.org.

[64] Steven Thornton and Rob Corless. Bohemian matrices. http://www.bohemianmatrices.com/. Accessed: 2021-11-01.
[65] Paul Voutier. An effective lower bound for the height of algebraic numbers. *Acta Arith.*, 74(1):81–95, 1996.
[66] Daniel Wiegreffe. Fractals from complex roots of polynomials. https://wiegreffe.info/home/fractals-from-complex-roots-of-polynomials. Accessed: 2020-08-09.
[67] Eduard Wirsing. Approximation mit algebraischen Zahlen beschränkten Grades. *J. Reine Angew. Math.*, 206:67–77, 1961.
[68] Enrique Zeleny. Algebraic Numbers In The Complex Plane. https://demonstrations.wolfram.com/AlgebraicNumbersInTheComplexPlane/. Accessed: 2020-08-09.

Department of Mathematics, University of Colorado, Campus Box 395, Boulder, Colorado 80309-0395