Equidistribution of shapes of complex cubic fields of fixed quadratic resolvent

Robert Harron

Abstract. We show that the shape of a complex cubic field lies on the geodesic of the modular surface defined by the field's trace-zero form. We also prove a general such statement for all orders in étale $\mathbf{Q}$-algebras. Applying a method of Manjul Bhargava and Piper H to results of Bhargava and Ariel Shnidman, we prove that the shapes lying on a fixed geodesic become equidistributed with respect to the hyperbolic measure as the discriminant of the complex cubic field goes to infinity. We also show that the shape of a complex cubic field is a complete invariant (within the family of all cubic fields).

Contents

1. Introduction

David Terr, in his PhD thesis [Ter97], introduced the notion of the *shape* of a number field K. This is a certain lattice (of rank $[K : \mathbf{Q}] - 1$) attached to K considered up to rotation, reflection, and scaling. In [Ter97], he proves that the shapes of cubic fields are equidistributed in the space of shapes of rank 2 lattices (as the discriminant of the cubic field goes to infinity). Manjul Bhargava and Piper H ([BH16, H16]) generalize this result to show the equidistribution of shapes of S_n-number fields of degree n for $n = 4$ and 5. These authors in fact conjecture that such equidistribution holds for all n reflecting the idea that a degree n number field with Galois group S_n is "random". The primary goal of this article is to investigate the distribution of more specific (read "less random") families of

Date: July 15, 2019.
2010 *Mathematics Subject Classification.* 11R16, 11R45, 11E12.
Key words and phrases. Cubic fields, lattices, equidistribution, geodesics, majorant space.
The author is partially supported by a Simons Collaboration Grant.

number fields. In [Ter97], Terr proves that C_3-cubic fields all have the same shape (hexagonal!), thus showing that restricting the number fields can impose strong constraints on the shapes. In [BS14], the shapes of real cubic fields with fixed quadratic resolvent field are shown to lie in a finite set and be (essentially) equidistributed in it. We consider the shapes of complex cubic fields of fixed quadratic resolvent and show that they are equidistributed on certain geodesics on the modular surface (see Fig. 1 for an example). Combined with [Har17], our main result suggests an intriguing explanation of a result of Cohen–Morra [CM11, Theorem 1.1] and [BS14, Theorem 6] that the number of cubic fields with quadratic resolvent $\mathbf{Q}(\sqrt{-3})$ has bigger growth rate than that of fields with other quadratic resolvents by a log factor. We prove some other interesting results along the way, including that the shape of a complex cubic determines that field within the family of all cubic fields, and that the shapes of degree n number fields with given trace-zero form lie on the majorant space of the trace-zero form.

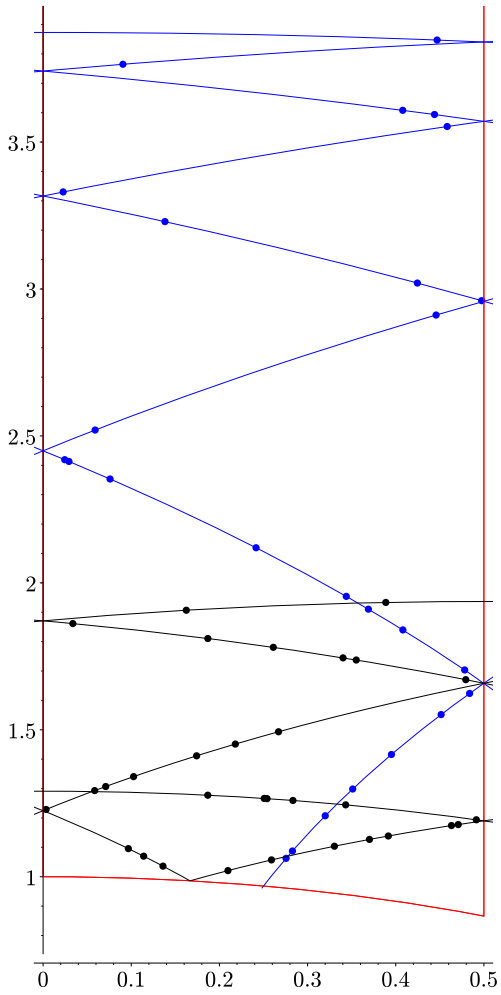

Figure 1. Shapes of all complex cubic fields K with quadratic resolvent $\mathbf{Q}(\sqrt{-20})$ and $|\Delta(K)| \leq 3,375,000$. The blue geodesic (i.e. the one that goes higher) corresponds to the class of $-x^2 + 6xy + 6y^2$, whereas the black geodesic corresponds to that of $-2x^2 + 6xy + 3y^2$. These represent the only two $\mathrm{GL}_2(\mathbf{Z})$-classes of indefinite integral binary quadratic forms of discriminant $60 = (-3) \cdot (-20)$.

We now describe in more detail our results and how they fit in to our current understanding.

1.1. **The shape as an invariant.** A fundamental use of attaching invariants to number fields (or to any objects in mathematics) is to help distinguish between them. The degree and the discriminant are the first one typically encounters and together they completely distinguish a quadratic number field from other number fields. Sadly, in every degree greater than 2, there are non-isomorphic fields with the same discriminant. A natural refinement of the discriminant of a degree n number field K, with ring of integers $\mathcal{O}_K$, is the isometry class of the *trace pairing*

$$\begin{array}{rcl} T_K : \mathcal{O}_K \times \mathcal{O}_K & \to & \mathbf{Z} \\ (a,b) & \mapsto & \mathrm{Tr}_{K/\mathbf{Q}}(ab) \end{array},$$

where $\mathrm{Tr}_{K/\mathbf{Q}} : K \to \mathbf{Q}$ denotes the usual field trace. Indeed, the discriminant is the determinant of this pairing (i.e. the determinant of a Gram matrix representing this bilinear form). Despite some goods news in the case of totally real fields (see e.g. [MSRG19]), the extra information of the trace form is no help for complex cubic fields: two complex cubic fields have isometric trace forms if and only if they have the same discriminant ([MS15, Theorem 3.3])

The *shape* of a number field stems from a similar refinement of the discriminant using Minkowski's geometry of numbers. One obtains a pairing

$$M_K : \mathcal{O}_K \times \mathcal{O}_K \to \mathbf{Z}$$

by first embedding K into $\mathbf{C}^n$ and taking the standard Hermitian inner product on $\mathbf{C}^n$. Unlike T_K whose signature depends on the number of real and complex places of K ([Tau68]), M_K is always positive-definite. In this article, we prove results that seem to indicate this feature allows it to retain more information about the number field than the trace form does. More specifically, we do not consider M_K, but rather its projection $M_K^\perp$ to the trace zero space, up to homotheties. In other words, we take the lattice corresponding to M_K, project it onto the space of elements of K of trace zero, and consider its equivalence class under scaling, rotations, and reflections. This is what is called the *shape* of K, denoted $\mathrm{sh}(K)$. In §5, we prove that the shape is a complete invariant in the family of complex cubic fields.

Theorem A. *If K is a complex cubic field and L is any other cubic field, then $\mathrm{sh}(K) \neq \mathrm{sh}(L)$. In particular, the shape is a complete invariant in the family of complex cubic fields.*

Since the shape of K is a rank 2 lattice up to homothety, it corresponds to a point in the upper-half plane. We in fact show that the field K is obtained by adjoining to $\mathbf{Q}$ the x-coordinate of its shape (or, when K is a pure cubic[1], its y-coordinate). This is then reminiscent of how a quadratic field is obtained by adjoining to $\mathbf{Q}$ the square root of its discriminant.

1.2. **The distribution of complex cubic shapes.** Once we have an invariant, we can ask how "random" it is. A general philosophy is that it should be as random as it can be! And this is what we show.

What constraints are there on the shape of a complex cubic field K? The following result gives an elegant answer involving the association of a geodesic on $\mathrm{GL}_2(\mathbf{Z})\backslash\mathfrak{H}$ to an indefinite binary quadratic form over $\mathbf{Z}$ (see §6 for some background about these geodesics and the beginning of §7 for a proof of this theorem).

Theorem B. *The shape of the complex cubic field K lies on the geodesic associated to the indefinite binary quadratic form $T_K^\perp$.*

Here, $T_K^\perp$ is the projection of T_K to the trace zero space.

In fact, in §8, we prove a generalization of this to all number fields. The shape of K is a point in the space of rank $n-1$ lattices and a construction of Siegel's associates to a quadratic form Q a subspace $\mathfrak{H}_Q$ of this space. We prove the following theorem.

Theorem C. *If K is a number field (of degree ≥ 3), then its shape lies on $\mathfrak{H}_{T_K^\perp}$.*

[1]Recall that a *pure* cubic field is one of the form $\mathbf{Q}(\sqrt[3]{m})$ for some non-cube $m \in \mathbf{Q}$.

When K is not totally real, $\mathfrak{H}_{T_K^\perp}$ is positive-dimensional, whereas $\mathrm{sh}(K)$ is a point. It is in this sense that we say that the positive-definiteness of $M_K^\perp$ seems to retain more information than $T_K^\perp$.

Considering $T_K^\perp$ as a quadratic form over $\mathbf{Z}$, we can study the associated primitive quadratic form $T_K^{\perp\prime}$ obtained by dividing by the gcd of the coefficients. We can characterize which (equivalence classes of) quadratic forms arise as $[T_K^{\perp\prime}]$ for some cubic field K (here we denote by $[Q]$ the equivalence class of the binary quadratic form Q under the action of $\mathrm{GL}_2(\mathbf{Z})$). The work of Bhargava and Shnidman [BS14, §5] tells us most of the story, though the following result, which we prove in §4, adds the finishing touch.

Theorem D. *If K is a cubic field such that the discriminant of its quadratic resolvent field is divisible by 3, then*

$$\mathrm{ord}_3 \Delta(T_K^{\perp\prime}) = \begin{cases} 0 & \text{if 3 is wild in } K, \\ 2 & \text{if 3 is tame in } K. \end{cases}$$

We refer to this phenomenon as the Tame–Wild Dichotomy. With this and [BS14, §5 and Theorem 4] in hand, we can say the following.

Corollary 1.1. *Let K be a cubic field whose quadratic resolvent field has discriminant d. Let*

$$D = \begin{cases} -3d & \text{if 3 is tame in } K \\ -d/3 & \text{if 3 is wild.} \end{cases} \tag{1.1}$$

Then, $T_K^{\perp\prime}$ has discriminant D. Furthermore, as K varies over fields with fixed d and fixed choice of whether 3 is tame or wild in K, the $T_K^{\perp\prime}$ of such K are equidistributed amongst the finitely many equivalence classes[2] of binary quadratic forms of discriminant D as the discriminant of K goes to infinity.

We note that, in [BS14], the authors refer to $T_K^{\perp\prime}$ as the "shape of K" even when K is not real. This is in contrast to the original definition of the shape in [Ter97]. In the real case, they are indeed discussing the shape of K and so the above corollary tells us about the distribution of shapes of real cubic fields with fixed quadratic resolvent.

After this result, what remains is to understand the distribution of shapes of complex cubic fields with fixed $[T_K^{\perp\prime}]$. We prove that they are equidistributed on the geodesic corresponding to $[T_K^{\perp\prime}]$ with respect to its natural hyperbolic measure as the discriminants of the fields go to infinity. For pure cubic fields, i.e. those fields whose $T_K^{\perp\prime}$ is (equivalent to) xy or $3xy - y$, this is proved in the author's previous article [Har17, Theorem C]. In that case, the corresponding geodesics have *infinite* length and the equidistribution is in a "regularized" sense as described in *ibid.* For non-pure complex cubic fields, the geodesics in question all have finite length. We will prove the following result in §7.

Theorem E. *Let $[Q]$ be a(n equivalence class of a) quadratic form that arises as the $[T_K^{\perp\prime}]$ of some non-pure complex cubic field K. Let $[\gamma_Q]$ denote the associated geodesic in $\mathrm{GL}_2(\mathbf{Z})\backslash\mathfrak{H}$ endowed with the measure μ_Q it inherits from the hyperbolic metric on $\mathfrak{H}$. Let W be a μ_Q-continuity set[3] of $[\gamma_Q]$ and let*

$$N(Q; X, W) = \{K \text{ complex cubic field} : [T_K^{\perp\prime}] = [Q], \mathrm{sh}(K) \in W, |\Delta(K)| < X\}.$$

Then,

$$\lim_{X\to\infty} \frac{N(Q; X, W)}{N(Q; X, [\gamma_Q])} = \frac{\mu_Q(W)}{\mu_Q([\gamma_Q])}, \tag{1.2}$$

i.e. the shapes are equidistributed as the discriminant of K goes to infinity. More specifically, there is a constant $C_Q > 0$ such that

$$N_Q(X; W) = C_Q \mu_Q(W) \sqrt{X} + o(\sqrt{X}). \tag{1.3}$$

[2]Because of ambiguous forms, to get equidistribution we must either count *oriented* cubic fields and use $\mathrm{SL}_2(\mathbf{Z})$-equivalence of binary quadratic forms, or weight the forms by the "size" of their automorphism group (ambiguous forms having automorphisms groups that are twice as big).

[3]Recall that a *continuity set* is a measurable subset whose boundary has measure 0.

In [Har17, Theorem C], what is proven is that when K is pure cubic, equation (1.3) holds for all *compact* μ_Q-continuity sets W. We prove analogues of Theorem E along the way for orders, as well as oriented fields and orders in Theorem 7.10, Corollary 7.12, and Theorem 7.13, respectively.

As an additional bit of information on the location of shapes of complex cubic fields, we have the following result that we prove in §5.1 and that we use in the proof of Theorem E.

Theorem F. *The only complex cubic fields whose shape lies on the boundary of the space of two-dimensional shapes (or, equivalently, whose lattice has extra automorphisms) are the wild pure cubic fields.*

1.3. **Shapes and log terms in asymptotics of counting number fields.** Aside from the inherent interest in understanding the distribution of shapes of number fields, the equidistribution result above together with that of [Har17] suggest an interesting explanation for the occurrence of log terms in the asymptotics of counting number fields.

A theorem of Henri Cohen and Anna Morra [CM11, Theorem 1.1], and independently Bhargava and Shnidman [BS14, Theorem 6], says that the number of cubic number fields K of fixed quadratic resolvent field K_2 (and discriminant bounded by X) grows like

$$\begin{cases} \sqrt{X} & \text{if } K_2 \neq \mathbf{Q}(\sqrt{-3}) \\ \sqrt{X}\log(X) & \text{if } K_2 = \mathbf{Q}(\sqrt{-3}). \end{cases}$$

What we show in this article and [Har17] is that for a fixed indefinite quadratic form Q (that occurs as $T_K^{\perp\prime}$ for some complex cubic field K), there is a constant $C_Q > 0$ such that for any *compact* continuity set W in $[\gamma_Q]$,

$$N_Q(X, W) = C_Q \mu_Q(W)\sqrt{X} + o(\sqrt{X}),$$

independent of the quadratic resolvent field of K. The quadratic resolvent field is $\mathbf{Q}(\sqrt{-3})$ if and only if the field is a pure cubic. In this case, the geodesics in question have infinite length. For instance, for a wild pure cubic field K, the shape is rectangular and is thus parametrized by what we call the *ratio* r_K of K that measures the ratio of the length of the sides of the rectangle. The natural measure on the space of rectangular lattices is $\dfrac{dr_K}{r_K}$, so that the measure of the set of rectangles with ratio in the interval $[1, R]$ is proportional to $\log(R)$. We can then think of the extra $\log(X)$ term coming from the measure of the sets W going to infinity.

The author has explored this phenomenon in joint work in other situations.

- In [HH19a] with Piper H, we study the shapes of V_4-quartic fields. Baily showed ([Bai80]) that there is a constant $C_{V_4} > 0$ such that the number of V_4-quartic fields with discriminant bounded by X is
$$C_{V_4}\sqrt{X}\log^2(X) + o(\sqrt{X}\log^2(X)).^4$$
We show that the shapes of V_4-quartic fields K are equidistributed (in a regularized sense) in a two-dimensional space of shapes of infinite measure. A bit more specifically, the shapes are given by rectangular prisms (so-called *orthorhombic* lattices) whose side lengths are in ratios given by the discriminants of the three quadratic subfields of K. We may therefore parametrize the shapes by two ratios $r_{K,1}$ and $r_{K,2}$ so that the relevant space of lattices has measure $\dfrac{dr_{K,1}dr_{K,2}}{r_{K,1}r_{K,2}}$. The number of K of discriminant bounded by X whose shape lies in a compact continuity set W is then shown to be proportional to
$$\mu(W)\sqrt{X} + o(\sqrt{X}).$$
Similarly to the case of pure cubic fields, the set of orthorhombic lattices with sides ratios in the box of side $R \times R$ is proportional to $\log^2(R)$ so that we now see two log terms arising. This result is generalized to cases of triquadratic fields in the upcoming PhD thesis of Jamal Hassan.

[4] Secondary terms have been determined since Baily's work.

- In joint work with Erik Holmes ([HH19b]), we study shapes of sextic fields containing a quadratic subfield. This case is of particular interest as Klüners' counterexample [Klü05] to Malle's conjecture is about fields of this form. Indeed, Klüners studies sextic fields K whose Galois group is $C_3 \wr C_2 (\cong S_3 \times C_3)$. These fields are exactly those non-Galois sextic fields containing some quadratic subfield K_2 over which they are Galois. Malle's original conjecture predicts that the number of such K of discriminant bounded by X grows like $\sqrt{X}$; however Klüners shows that even just the number of those with $K_2 = \mathbf{Q}(\omega)$ (where ω is a primitive cube root of unity) grows like $\sqrt{X}\log(X)$. In fact, Klüners shows that the number of K with $K_2 \neq \mathbf{Q}(\omega)$ grows like $\sqrt{X}$. In [HH19b], we consider such K with a fixed K_2 and study their "K_2-shapes", i.e. we take the Minkowski lattice attached to K and project it onto the orthogonal complement of K_2 (rather than just the orthogonal complement of $\mathbf{Q}$). We thus obtain rank 4 lattices. We show that, when $K_2 = \mathbf{Q}(\omega)$, the shapes live in a one-dimensional space parametrized by a ratio r_K with natural measure $\dfrac{dr_K}{r_K}$, that the number of fields with shape in a compact continuity set grows like $\sqrt{X}$, and that the full space has infinite measure. We also show that, when $K_2 \neq \mathbf{Q}(\omega)$ (and, for simplicity, when the field $K(\omega)$ has class number 1), the shapes lie in a one-dimensional space that is now parametrized by an *angle* θ_K. The measure of the full space then has finite measure and we show that the number of fields with shape in any continuity set grows like $\sqrt{X}$.

These results suggest that the study of shapes could be a fruitful avenue to understanding log terms in asymptotics of counting number fields. In particular, it would be interesting to produce heuristics for these counting functions that rely on understanding what kind of space the given shapes lie in.

2. Shapes and trace-zero forms of number fields

In this brief section, we collect the definitions and notation relevant to the discussion of shapes and trace-zero forms of a number field. Some justification for these definitions is provided in the introduction so this section will be pretty dry.

Let K be a degree n number field and let $\sigma_1, \dots, \sigma_n$ be the n distinct embeddings of K into $\mathbf{C}$. We may collect these together into what we refer to as the *Minkowski embedding of* K:

$$\begin{array}{rcl} j : K & \to & \mathbf{C}^n \\ \alpha & \mapsto & (\sigma_1(\alpha), \dots, \sigma_n(\alpha)). \end{array}$$

Let $K_{\mathbf{R}} := \mathrm{im}(j) \otimes_{\mathbf{Q}} \mathbf{R} \subseteq \mathbf{C}^n$. The restriction of the standard inner product $\langle \cdot\ ,\ \cdot \rangle$ on $\mathbf{C}^n$ to $K_{\mathbf{R}}$ turns $K_{\mathbf{R}}$ into a real Euclidean space (see e.g. [Neu99, §I.5] for details). We refer to this Euclidean space as the *Minkowski space of* K and call its inner product the *Minkowski inner product*. Given any order R in K (or even an ideal in an order), we obtain a full rank lattice $j(R)$ in $K_{\mathbf{R}}$. The *shape of* R is the equivalence class (under scaling, rotations, and reflections) of the projection of $j(R)$ onto the orthogonal complement of $j(1)$ in $K_{\mathbf{R}}$. The *shape of* K is the shape of the maximal order $\mathcal{O}_K$. Concretely, we introduce the "perp" map $R \to R$ defined by

$$\alpha^{\perp} := n\alpha - \mathrm{tr}(\alpha), \tag{2.1}$$

where $\mathrm{tr}(\alpha)$ is the usual trace map $\mathrm{tr} : K \to \mathbf{Q}$. Then, one may verify that $j(\alpha^{\perp})$ is n times the orthogonal projection of $j(\alpha)$ onto the orthogonal complement of $j(1)$: the key points being that

$$\langle j(1), j(\alpha) \rangle = \mathrm{tr}(\alpha) \quad \text{and} \quad \langle j(1), j(1) \rangle = n.$$

We let $R^{\perp}$ be the image of R under this perp map. Then, the shape of R is concretely the equivalence class under scaling, rotations, and reflections of $j(R^{\perp})$. Thus, if we begin with a $\mathbf{Z}$-basis of R, say $1, \alpha_1, \dots, \alpha_{n-1}$, the images of $\alpha_1^{\perp}, \dots, \alpha_{n-1}^{\perp}$ under j form a $\mathbf{Z}$-basis of $j(R^{\perp})$ of R. The knowledge of the Gram matrix of the Minkowski inner product with respect to an integral basis of R will then give us, by bilinearity, a Gram matrix representing the shape. The restriction of the Minkowski inner product to R is what we denote M_R and call the *Minkowski form of* R.[5] Although it does not take

[5]Technically, we are composing the Minkowski inner product with the restriction of j to R.

values in $\mathbf{Z}$, its values are algebraic integers. We denote by $M_R^\perp$ the induced pairing on $R^\perp$. The scaling by n present in the definition of the perp map is chosen so that $M_R^\perp$ again takes integral values. We may sometimes use $M_R^\perp$ to denote the positive-definite binary quadratic form associated to the pairing. We let M_K and $M_K^\perp$ denote the forms for $R = \mathcal{O}_K$. When $n = 3$, $M_K^\perp$ is a positive definite *binary* quadratic form $Q(x, y) = rx^2 + sxy + ty^2$ and thus has a point $z(Q) = a + bi$ associated to it in the upper-half plane $\mathfrak{H} = \{x + iy \in \mathbf{C} : y > 0\}$.

Lemma 2.1. *For a positive definite $Q(x, y) = rx^2 + sxy + ty^2$, we have that $z(Q) = a + bi$ where*

$$a = \frac{s}{2r} \quad \text{and} \quad b = \frac{\sqrt{4rt - s^2}}{2r}. \tag{2.2}$$

Proof. We can think of $Q(x, y)$ as being the quadratic form giving the norm squared of a vector $xv_1 + yv_2$ in a lattice spanned by two linearly independent vectors $v_1, v_2 \in \mathbf{C} \cong \mathbf{R}^2$. We also identify the upper-half plane with the set of lattices in $\mathbf{C}$ up to rotations, reflections, and scaling as usual, i.e. we identify the above lattice with v_2/v_1 (after a possible reflection, we may assume v_2/v_1 has positive imaginary part). The quadratic form does not change under orthogonal transformations, so we may assume v_1 lies on the positive real axis, in which case $v_1 = (\sqrt{r}, 0)$. Then, under the correspondence with $\mathfrak{H}$, we must have that $v_2 = (\sqrt{r}a, \sqrt{r}b)$. Taking inner products yields the result. □

We can do something similar by replacing the Minkowski inner product with the so-called trace form. The *trace form of R* is

$$\begin{array}{rccc} T_R : R \times R & \to & \mathbf{Z} \\ (\alpha, \beta) & \mapsto & \mathrm{tr}(\alpha\beta). \end{array}$$

This amounts to replacing the standard inner product on $\mathbf{C}^n$ with the dot product, i.e. we omit taking the complex conjugate of one of the vectors. The *trace-zero form of R* is the induced pairing $T_R^\perp$ on $R^\perp$. We call this the "trace-zero" form since the orthogonal complement of $j(1)$ in $K_{\mathbf{R}}$ is indeed the ($\mathbf{R}$-span of the image under j of the) space of elements of trace 0 in K. As with $M_R^\perp$, we may also denote by $T_R^\perp$ the binary quadratic form associated to the pairing. Since the form has $\mathbf{Z}$-coefficients, we may also consider the *primitive trace-zero form* $T_R^{\perp\prime}$, which is the associated primitive binary quadratic form (i.e. the one where we have divided by the greatest common divisor of the coefficients). Again, we use the subscript K for the case $R = \mathcal{O}_K$.

3. The Levi–Delone–Fadeev correspondence and shapes of cubic fields

The Levi–Delone–Fadeev [DF64, Lev14] correspondence provides a very useful bijection between isomorphism classes of cubic rings (that is (commutative, unital) rings R that are isomorphic to $\mathbf{Z}^3$ as $\mathbf{Z}$-modules) and $\mathrm{GL}_2(\mathbf{Z})$-equivalence classes of binary cubic forms. We collect in this section some formulas and features of this correspondence that will be of use to us in subsequent sections. We refer to [BST13], especially §2, for more/other details and unreferenced claims (another modern reference is [GGS02, §4]).

Let $F(x, y) = ax^3 + bx^2y + cxy^2 + dy^3$ be a binary cubic form with coefficients in $\mathbf{Z}$. Associated to it, we have a cubic ring R_F with a basis $1, \alpha, \beta$ such that

$$\alpha\beta = -ad, \tag{3.1}$$

$$\alpha^2 = -ac - b\alpha + a\beta, \tag{3.2}$$

$$\beta^2 = -bd - d\alpha + c\beta. \tag{3.3}$$

The discriminant of R_F equals that of F and is given by

$$b^2c^2 - 4ac^3 - 4b^3d - 27a^2d^2 + 18abcd. \tag{3.4}$$

Let $\mathrm{tr} : R \to \mathbf{Z}$ be the trace map (i.e. the map that sends $\gamma \in R$ to the trace of the $\mathbf{Z}$-linear multiplication-by-γ map on R).

Lemma 3.1. *The elements α and β satisfy*

$$f_\alpha(x) = a^2 F\left(\frac{x}{a}, 1\right) \quad \text{and} \quad f_\beta(x) = \frac{x^3}{d} \cdot F\left(\frac{-d}{x}, 1\right)$$

respectively. In particular,

$$\operatorname{tr}(\alpha) = -b \tag{3.5}$$

$$\operatorname{tr}(\beta) = c. \tag{3.6}$$

Proof. Using (3.1) and (3.2), one sees that

$$\begin{aligned} \alpha^3 = \alpha \cdot \alpha^2 &= -ac\alpha - b\alpha^2 + a\alpha\beta \\ &= -ac\alpha - b\alpha^2 - a^2 d. \end{aligned}$$

A similar calculation works for β. □

The action of $\mathrm{GL}_2(\mathbf{R})$ on $F(x, y)$ compatible with the correspondence is the so-called *twisted action*

$$g \cdot F(x, y) := \frac{1}{\det g} F((x, y)g),$$

where $(x, y)g$ is the linear change of variables given by the vector-matrix multiplication. Then, $\mathrm{GL}_2(\mathbf{Z})$-equivalence classes of forms correspond to isomorphism classes of rings. The forms themselves correspond to pairs consisting of a cubic ring R and a basis of $R/\mathbf{Z}$, namely F corresponds to $(R_F, (\alpha + \mathbf{Z}, \beta + \mathbf{Z}))$.

We have the following connections between F and ring-theoretic properties of R_F. Recall that a cubic ring R is said to be *maximal* at a prime p if there is no cubic ring R' such that $R' \supsetneq R$ and $p \mid [R' : R]$.

Proposition 3.2. *The cubic ring R_F is an integral domain if and only if $F(x, y)$ is irreducible over $\mathbf{Q}$. It is not maximal at p if and only if p divides all the coefficients of F or there is an element $g \in \mathrm{GL}_2(\mathbf{Z})$ such that p^2 divides the x^3-coefficient of $g \cdot F$ and p divides the x^2y-coefficient.*

Proof. This is [BST13, Proposition 12] and the comments after [BST13, Lemma 13]. □

Suppose R_F is an order in a complex cubic number field K; in particular, $F(x, y)$ is irreducible. Let η be a root of $F(x, 1)$ in K. Using Lemma 3.1, we get that

$$\alpha = a\eta \quad \text{and} \quad \beta = -d/\eta.$$

We now wish to say something about the *shape* of R_F. Let σ be the real embedding of K and let $\tau, \overline{\tau}$ be the pair of its non-real embeddings. Let $\theta = \sigma(\eta)$, $\xi = \tau(\eta)$, and let $j : K \hookrightarrow \mathbf{C}^3$ be the Minkowski embedding $\mu \mapsto (\sigma(\mu), \tau(\mu), \overline{\tau}(\mu))$. Then,

$$\begin{aligned} j(\eta) &= (\theta, \xi, \overline{\xi}), \\ j(\alpha) &= (a\theta, a\xi, a\overline{\xi}), \\ j(\beta) &= (-d/\theta, -d/\xi, -d/\overline{\xi}). \end{aligned}$$

As a first step towards determining the shape of R_F, we will determine the Gram matrix of the Minkowski form of R_F with respect to the basis $1, \alpha, \beta$. By construction, the entries of the Gram matrix will lie in the maximal real subfield of the Galois closure of K and so, following [Ter97, §9], we will use the $\mathbf{Q}$-basis $\theta^{-1}, 1, \theta$ for the field $\mathbf{Q}(\theta) \subseteq \mathbf{R}$.

Lemma 3.3. *The minimal polynomial of ξ over $\mathbf{Q}(\theta)$ is*

$$x^2 - t_\xi x + n_\xi,$$

where

$$t_\xi = -\frac{b}{a} - \theta$$
$$n_\xi = -\frac{d}{a}\theta^{-1}.$$

Proof. Note that over $\mathbf{Q}(\theta)$, we have

$$\begin{aligned} a(x-\theta)(x^2 - t_\xi x + n_\xi) &= ax^3 + (-a)(t_\xi + \theta)x^2 + a(n_\xi + \theta)x + (-a)\theta n_\xi \\ &= ax^3 + bx^2 + cx + d \\ &= F(x,1). \end{aligned}$$

□

Proposition 3.4. *The Gram matrix of M_{R_F} with respect to the basis $1, \alpha, \beta$ is*

$$\begin{pmatrix} 3 & -b & c \\ -b & -a(3d\theta^{-1} + c + b\theta) & -(bd\theta^{-1} + bc + ac\theta) \\ c & -(bd\theta^{-1} + bc + ac\theta) & -d(c\theta^{-1} + b + 3a\theta) \end{pmatrix}.$$

Proof. For any $\gamma \in K$, $\langle j(1), j(\gamma)\rangle = \mathrm{tr}(\gamma)$, so the first row (and column) follow from equations (3.5) and (3.6). The rest follows from the previous lemma and the fact that each of θ, ξ, and $\overline{\xi}$ satisfy $F(x,1)$. For instance, one may begin with

$$\begin{aligned} \langle j(\alpha), j(\beta)\rangle &= -ad\left(1 + \frac{\xi}{\overline{\xi}} + \frac{\overline{\xi}}{\xi}\right) \\ &= -ad\left(1 + \frac{\xi^2 + \overline{\xi}^2}{\xi\overline{\xi}}\right) \\ &= -ad\left(1 + \frac{(\xi + \overline{\xi})^2 - 2\xi\overline{\xi}}{\xi\overline{\xi}}\right) \\ &= -ad\left(1 + \frac{t_\xi^2 - 2n_\xi}{n_\xi}\right). \end{aligned}$$

□

Proposition 3.5. *The Gram matrix of $M_{R_F}^\perp$ with respect to the basis $\alpha^\perp, \beta^\perp$ is*

$$\begin{pmatrix} -3(9ad\theta^{-1} + (b^2 + 3ac) + 3ab\theta) & -3(3bd\theta^{-1} + 2bc + 3ac\theta) \\ -3(3bd\theta^{-1} + 2bc + 3ac\theta) & -3(3cd\theta^{-1} + (c^2 + 3bd) + 9ad\theta) \end{pmatrix}. \tag{3.7}$$

Proof. This follows from the previous proposition and bilinearity, using the formula (2.1) for the perp map. □

We also denote by $M_F^\perp$ the binary quadratic form associated to this Gram matrix. The map $F \mapsto M_F^\perp$ is $\mathrm{GL}_2(\mathbf{R})$-equivariant, where the action of $g \in \mathrm{GL}_2(\mathbf{R})$ on a binary quadratic form $Q(x,y)$ is $(g \cdot Q)(x,y) = Q((x,y)g)$. We denote by $\mathrm{sh}(R_F)$ the point $z(M_F^\perp) \in \mathfrak{H}$.

The *Hessian* of F is the binary quadratic form ([BS14, eq. (2)])

$$H_F(x,y) := (b^2 - 3ac)x^2 + (bc - 9ad)xy + (c^2 - 3bd)y^2. \tag{3.8}$$

It represents the trace-zero form of R_F ([BS14, Proposition 12]) (up to scaling). The association $F \mapsto H_F$ is $\mathrm{GL}_2(\mathbf{R})$-equivariant and we have that ([BS14, Proposition 11])

$$\Delta(H_F) = -3\Delta(F). \tag{3.9}$$

Note however that the Gram matrix of the trace-zero form of R_F with respect to $\alpha^\perp, \beta^\perp$ is

$$\begin{pmatrix} 6(b^2 - 3ac) & 3(bc - 9ad) \\ 3(bc - 9ad) & 6(c^2 - 3bd) \end{pmatrix}. \tag{3.10}$$

This is the Gram matrix of $6H_F$, accordingly we let $T_F^\perp := 6H_F$.

For technical reasons, in §7, we will need to work with *oriented* cubic rings as in [BS14, p. 55]. An *oriented cubic ring* is a pair (R, δ) where R is a cubic ring and δ is an isomorphism $\wedge^3 R \cong \mathbf{Z}$. This orientation has the effect of fixing an ordered basis of $R/\mathbf{Z}$ and as such the shape of an oriented ring is only determined up to $\mathrm{SL}_2(\mathbf{Z})$-equivalence, and therefore lives in $\mathrm{SL}_2(\mathbf{Z})\backslash\mathfrak{H}$. We will be able to translate results about shapes of oriented cubic rings into ones about plain old cubic rings.

4. The Tame-Wild Dichotomy

In this section, we prove Theorem D, which provides an interpretation for the discriminant of the primitive trace-zero form in terms of the tame versus wild ramification of 3 in the cubic field K.

We begin by remarking that if 3 is ramified in K, then $\mathrm{ord}_3(\Delta(K)) = 1, 3, 4$, or 5. Indeed, letting $\mathcal{D}(K)$ denote the different of K, if 3 factors as $\mathfrak{p}_1^2\mathfrak{p}_2$, then $\mathfrak{p}_1$ is tamely ramified, so that $\mathrm{ord}_{\mathfrak{p}_1}(\mathcal{D}(K)) = 1$, and hence $\mathrm{ord}_3(\Delta(K)) = 1$. Otherwise, $3 = \mathfrak{p}^3$ is wildly ramified, so that $3 \leq \mathrm{ord}_{\mathfrak{p}}(\mathcal{D}(K)) \leq 5$ by the standard inequality (from the standard reference [Ser68, §III.6, Proposition 13, and following remark]). Since the quadratic resolvent field of K is $\mathbf{Q}(\sqrt{\Delta(K)})$, the assumption that 3 divides the latter's discriminant is equivalent to $\mathrm{ord}_3(\Delta(K))$ being odd. Thus, we wish to prove the following result.

Proposition 4.1. *If K is a cubic field such that the discriminant of its quadratic resolvent field is divisible by 3, then*

$$\mathrm{ord}_3 \Delta\left(T_K^{\perp\prime}\right) = \begin{cases} 0 & \text{if } \mathrm{ord}_3 \Delta(K) = 3 \text{ or } 5, \\ 2 & \text{if } \mathrm{ord}_3 \Delta(K) = 1. \end{cases}$$

We will require the following lemma.

Lemma 4.2. *Suppose $F(x, y) = ax^3 + bx^2y + cxy^2 + dy^3$ is a binary cubic form corresponding to a maximal cubic ring.*

(a) *If F is a cube modulo 3, then it is $\mathrm{GL}_2(\mathbf{Z})$-equivalent to a form with $a \equiv b \equiv c \equiv 0 \pmod 3$ and $d \not\equiv 0 \pmod 3$.*
(b) *If F factors as a $G_1G_2^2$ modulo 3, where G_1 and G_2 are linear forms that aren't constant multiples of each other, then it is $\mathrm{GL}_2(\mathbf{Z})$-equivalent to a form with $a \equiv b \equiv d \equiv 0 \pmod 3$ and $c \not\equiv 0 \pmod 3$.*

Proof. If $F(x, y) \equiv (mx + ny)^3 \pmod 3$, and $n \equiv 0 \pmod 3$, then switching x and y does the trick. Otherwise, the change of variable $x' = x$, $y' = -n^{-1}mx + y$ (where n^{-1} is an inverse of n modulo 3) moves F to a form congruent to $ny^3 \pmod 3$. If this new coefficient of y^3 were divisible by 3, then F would correspond to a non-maximal ring by Proposition 3.2.

If $F \equiv (m_1x + n_1y)(m_2x + n_2y)^2 \pmod 3$, a similar change of variables allows us to assume that $m_2 \equiv n_1 \equiv 0 \pmod 3$. Similarly, $c \not\equiv 0 \pmod 3$ since F corresponds to a maximal ring. □

Proof of Proposition 4.1. Let $F(x, y) = ax^3 + bx^2y + cxy^2 + dy^3$ be a binary cubic form corresponding to the ring of integers of K. Write the Hessian of F as $H_F(x, y) = rx^2 + sxy + ty^2$ and recall that it is an integer multiple of $T_K^{\perp\prime}$. Suppose first that 3 is wild in K. Then, by [DH71, Lemma 11], $F(x, y)$ is a cube modulo 3, and so, by the previous lemma, we may assume that 3 divides each of a, b, and c, but not d. By (3.8), $3^2 \mid \gcd(r, s, t)$, so that

$$\mathrm{ord}_3 \Delta\left(T_K^{\perp\prime}\right) \leq \mathrm{ord}_3(\Delta(H_F)) - 4 = \mathrm{ord}_3(\Delta(K)) - 3.$$

This gives the desired result when $\mathrm{ord}_3(\Delta(K)) = 3$. When this valuation is 5, we note, by (3.4), that $\Delta(F) \equiv -4b^3d \pmod{3^4}$, so that $3^2 \mid b$. Then,

$$\Delta(F) \equiv -4ac^3 \pmod{3^5}.$$

Since F corresponds to a maximal ring, $3^2 \nmid a$, so that $3^2 \mid c$. This now shows that $3^3 \mid \gcd(r, s, t)$, so that

$$\mathrm{ord}_3 \Delta\left(T_K^{\perp\prime}\right) \leq \mathrm{ord}_3(\Delta(H_F)) - 6 = \mathrm{ord}_3(\Delta(K)) - 5 = 0,$$

as desired.

Now, suppose 3 is tamely ramified in K. Again by [DH71, Lemma 11] and the previous lemma, we may assume that now $a \equiv b \equiv d \equiv 0 \pmod 3$, but $c \not\equiv 0 \pmod 3$. Then, $3 \nmid t$. Thus,

$$\operatorname{ord}_3 \Delta\big(T_K^{\perp\prime}\big) = \operatorname{ord}_3(\Delta(H_F)) = \operatorname{ord}_3(\Delta(K)) + 1 = 2.$$

□

5. The shape as a complete invariant

In this section, we prove that the shape is a complete invariant in the family of complex cubic fields (Theorem A above). More precisely, as stated above, we prove the stronger fact that the shape of a given (isomorphism class of a) complex cubic field is distinct from the shape of any other cubic field. This comes down to an (ir)rationality argument, and is reminiscent of how the discriminant is a complete invariant of quadratic fields: the quadratic field of discriminant Δ *is* the field obtained by adjoining a square root of Δ to $\mathbf{Q}$. Similarly, we will see that the complex cubic field of shape $x + iy \in \mathfrak{H}$ *is* (isomorphic to) the field obtained by adjoining x (or, in the pure cubic case, y) to $\mathbf{Q}$.

First, note that if K is a real cubic field, then we have an equality of binary quadratic forms $M_K^\perp = T_K^\perp$ so that $M_K^\perp$ has coefficients in $\mathbf{Z}$. By (2.2), the x-coordinate of $\operatorname{sh}(K)$ is in $\mathbf{Q}$ and its y-coordinate is in a quadratic extension of $\mathbf{Q}$. Either way, these coordinates are not irrational cubic numbers. Theorem A was proved by the author when K is a pure cubic field, in [Har17, Theorem B], essentially by showing that (the image of the real embedding of) K is $\mathbf{Q}(\operatorname{Im}(\operatorname{sh}(K))$. As noted above, $\operatorname{Re}(\operatorname{sh}(K))$ lies in the image of K under its real embedding. It thus suffices to prove the following proposition.

Proposition 5.1. *If K is a non-pure, complex cubic field, then* $\operatorname{Re}(\operatorname{sh}(K))$ *is irrational.*

Proof. Let $F(x, y) = ax^3 + bx^2y + cxy^2 + dy^3$ be a binary cubic form with integer coefficients corresponding to the ring of integers of K. The formula for the shape in Proposition 3.5 (and (2.2)) shows that

$$\operatorname{Re}(\operatorname{sh}(K)) = \frac{3bd\theta^{-1} + 2bc + 3ac\theta}{9ad\theta^{-1} + (b^2 + 3ac) + 3ab\theta} \in \mathbf{Q}(\theta). \tag{5.1}$$

Introduce three unknown rational numbers c_1, c_2, and c_3 such that $c_1\theta^{-1} + c_2 + c_3\theta$ is the inverse of the denominator $9ad\theta^{-1} + (b^2 + 3ac) + 3ab\theta$. Assume for the sake of contradiction that $\operatorname{Re}(\operatorname{sh}(K)) = \rho \in \mathbf{Q}$. Then, the two equations

$$\begin{aligned} 1 &= (9ad\theta^{-1} + (b^2 + 3ac) + 3ab\theta) \cdot (c_1\theta^{-1} + c_2 + c_3\theta) \\ \rho &= (3bd\theta^{-1} + 2bc + 3ac\theta) \cdot (c_1\theta^{-1} + c_2 + c_3\theta) \end{aligned}$$

give a system of six linear equations over $\mathbf{Q}$ in the three unknowns c_1, c_2, and c_3 represented by the following matrix

$$M = \begin{pmatrix} b^2 - 6ac & 9ad & -3bd & 0 \\ -6ab & b^2 + 3ac & 9ad - 3bc & 1 \\ -9a^2 & 3ab & -2b^2 + 3ac & 0 \\ -bc & 3bd & -3cd & 0 \\ -3b^2 + 3ac & 2bc & -3c^2 + 3bd & \rho \\ -3ab & 3ac & -bc & 0 \end{pmatrix}.$$

We will show that this system has no solutions under the hypotheses of this proposition. We note the following consequences of the hypotheses and the rows of M. First, since $F(x, y)$ corresponds to an order in an S_3-cubic field, it is irreducible, so both a and d are non-zero. Also, b and c can't both be zero: if they are, then $K = \mathbf{Q}(\sqrt[3]{-d/a})$, which is a pure cubic. If $b = 0$, then the fourth row of M implies that $c_3 = 0$, which in turn implies, from row three, that $c_1 = 0$. The first row then implies that $c_2 = 0$, so that the inverse of the denominator in (5.1) is zero: clearly a contradiction. A similar argument shows that c can't be zero either. Finally, the expression $b^2 - 3ac$ is non-zero: otherwise,

the Hessian of $F(x,y)$ has square discriminant so that $F(x,y)$ corresponds to a pure cubic field [BS14, Lemma 33]. Given these non-vanishing statements, we can use the matrix

$$E = \begin{pmatrix} \frac{b}{b^2-3ac} & 0 & 0 & \frac{-3a}{b^2-3ac} & 0 & 0 \\ \frac{b(bc-9ad)}{-3d(b^2-3ac)} & 1 & 0 & \frac{\frac{1}{3}b^3-2abc+9a^2d}{-d(b^2-3ac)} & 0 & -1 \\ \frac{9a^2b}{b^2-3ac} & 0 & b & \frac{-27a^3}{b^2-3ac} & 0 & 0 \\ \frac{bc}{3d(b^2-3ac)} & 0 & 0 & \frac{b^2-6ac}{3d(b^2-3ac)} & 0 & 0 \\ 3ab & 0 & -\frac{2}{3}bc & -9a^2 & ab & 0 \\ \frac{3ab}{b^2-3ac} & 0 & 0 & \frac{-9a^2}{b^2-3ac} & 0 & 1 \end{pmatrix}$$

to effect a row reduction of M yielding

$$EM = \begin{pmatrix} b & 0 & -3d & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 3ab^2 & -2b^3+3abc-27a^2d & 0 \\ 0 & b & -2c & 0 \\ 0 & 0 & \frac{4}{3}b^3c-5abc^2-6ab^2d+27a^2cd & ab\rho \\ 0 & 3ac & -bc-9ad & 0 \end{pmatrix}$$

whose second row shows this system is inconsistent, as desired. □

Note that the proof did not require that $F(x,y)$ correspond to a maximal order, therefore this irrationality result still holds for orders in non-pure complex cubic fields.

5.1. **The shapes avoid the boundary.** In this brief section, we will prove Theorem F as a corollary of Proposition 5.1. The standard Gauss fundamental domain for $\mathrm{GL}_2(\mathbf{Z})\backslash\mathfrak{H}$ is

$$\mathcal{G} = \left\{ x+iy : 0 \le x \le \frac{1}{2}, x^2+y^2 \ge 1 \right\}.$$

The boundary of this set consists of two vertical rays

$$\{iy : y \ge 1\} \quad \text{and} \quad \left\{ \frac{1+iy\sqrt{3}}{2} : y \ge 1 \right\}$$

and the circular arc

$$\{\cos(\theta) + i\sin(\theta) : \pi/3 \le \theta \le \pi/2\}.$$

If R is an order in a non-pure complex cubic field, then Proposition 5.1 above shows that its shape cannot lie on the two vertical boundary components. To prove Theorem F, it therefore suffices to show that the shape cannot lie on the circular arc of the boundary. The matrix

$$\begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix} \in \mathrm{SL}_2(\mathbf{Z})$$

maps the unit semi-circle $\{\cos(\theta)+i\sin(\theta) : 0 < \theta < \pi\}$ to the vertical ray $\left\{\frac{1+iy}{2} : y > 0\right\}$ by fractional linear transformation. Thus, if R corresponds to some binary cubic form F such that $M_F^{\perp}$ gives a point on the circular arc of the boundary, then F is $\mathrm{SL}_2(\mathbf{Z})$-equivalent to a form F' with $M_{F'}^{\perp}$ giving a point whose real part is $1/2$. This latter option is impossible by Proposition 5.1, so we have proved Theorem F.

6. Indefinite binary quadratic forms and geodesics on the modular surface

The material in this section is only needed for §7. We collect some facts and constructions regarding the correspondence between indefinite integral binary quadratic forms and certain geodesics on the modular surface. We refer the reader to [Sar07] and [EW11, Ch. 9] for more details.

Let $Q(x,y) = rx^2+sxy+ty^2$ be a real binary quadratic form and let $D = s^2-4rt$ be its discriminant. There is an action of $\mathrm{GL}_2(\mathbf{R})$ on binary quadratic forms given by $(g\cdot Q)(x,y) = Q((x,y)g)$. The *connected component of the identity of the orthogonal similitude group of* Q is

$$\mathrm{GO}_Q^0(\mathbf{R}) = \{g \in \mathrm{GL}_2(\mathbf{R}) : g\cdot Q = c_g Q, c_g > 0, \det(g) > 0\}.$$

Let $Q_0(x,y) = xy$. A straightforward calculation shows that $\mathrm{GO}_{Q_0}^0(\mathbf{R})$ is the group of diagonal matrices with positive determinant. There is a group isomorphism

$$\begin{aligned}\mathbf{R}_{>0}\times\mathbf{R}^\times &\xrightarrow{\sim} \mathrm{GO}_{Q_0}^0(\mathbf{R})\\ (\lambda,\alpha) &\longmapsto g(\lambda,\alpha),\end{aligned}$$

where

$$g(\lambda,\alpha) := \begin{pmatrix}\lambda\alpha^{-1} & \\ & \lambda\alpha\end{pmatrix}.$$

Suppose $Q(x,y) = rx^2+sxy+ty^2$ is indefinite (i.e. $D>0$) with $t\neq 0$. Let

$$\theta_\pm := \frac{s\pm\sqrt{D}}{2t}$$

and let

$$P = \sqrt{t}\begin{pmatrix}\theta_+ & \theta_-\\ 1 & 1\end{pmatrix}.$$

Then, $P\cdot Q_0 = Q$ so that $\mathrm{GO}_Q^0(\mathbf{R}) = P\mathrm{GO}_{Q_0}^0(\mathbf{R})P^{-1}$. Suppose now that Q is an indefinite *integral* binary quadratic form of non-square determinant (in particular $t\neq 0$). Let

$$\mathrm{GO}_Q^0(\mathbf{Z}) := \mathrm{GO}_Q^0(\mathbf{R})\cap\mathrm{GL}_2(\mathbf{Z}) = \mathrm{GO}_Q^0(\mathbf{R})\cap\mathrm{SL}_2(\mathbf{Z}) =: \mathrm{SO}_Q(\mathbf{Z}).$$

The set of integer solutions $(U,W)\in\mathbf{Z}^2$ to the Pellian equation $u^2-Dw^2 = 4$ forms a group isomorphic to $\{\pm1\}\times\mathbf{Z}$. There is an injective homomorphism of this group into $\mathbf{Q}(\sqrt{D})^\times$ sending (U,W) to $\frac{1}{2}(U+W\sqrt{D})$. There are then four elements $\epsilon_0 = \frac{1}{2}(U_0+W_0\sqrt{D})$ in the image of this homomorphism such that the image is $\{\pm\epsilon_0^m : m\in\mathbf{Z}\}$. We denote by ϵ_0 the unique such element that is >1. Then, $\langle\pm1,\epsilon_0\rangle$ is isomorphic to $\mathrm{SO}_Q(\mathbf{Z})$ via

$$\frac{1}{2}(U+W\sqrt{D}) \mapsto M(U,W) := \begin{pmatrix}\frac{1}{2}(U-sW) & rW\\ -tW & \frac{1}{2}(U+sW)\end{pmatrix}.$$

It can be verified that

$$P\begin{pmatrix}\epsilon_0^{-1} & \\ & \epsilon_0\end{pmatrix}P^{-1} = M(U_0,W_0),$$

so that

$$\mathrm{SO}_Q(\mathbf{Z}) = P\left\{\pm\begin{pmatrix}\epsilon_0^{-1} & \\ & \epsilon_0\end{pmatrix}^m : m\in\mathbf{Z}\right\}P^{-1}.$$

If $Q(x,y)$ is any (real) indefinite binary quadratic form, it has two "roots" $\rho_\pm\in\mathbf{P}^1(\mathbf{R})$, i.e. $\rho_\pm = [\rho_{\pm,x}:\rho_{\pm,y}]\in\mathbf{P}^1(\mathbf{R})$ such that

$$Q(x,y) = r(\rho_{+,y}x-\rho_{+,x}y)(\rho_{-,y}x-\rho_{i,x}y).$$

This allows one to associate to Q a geodesic γ_Q in $\mathfrak{H}$. Namely, if neither of $\rho_\pm$ is infinite, γ_Q is the semicircle whose diameter is the line connecting ρ_+ and ρ_-; otherwise, γ_Q is a vertical line connecting ∞ to whichever of $\rho_\pm$ is not infinite. Note that γ_{Q_0} is the positive imaginary axis $\{iy : y>0\}$. There is an action of $\mathrm{GL}_2(\mathbf{R})$ on $\overline{\mathfrak{H}} := \mathfrak{H}\cup\mathbf{P}^1(\mathbf{R})$ by fractional linear transformation given by

$$g\cdot z = \begin{pmatrix}a & b\\ c & d\end{pmatrix}\cdot z = \begin{cases}\dfrac{az+b}{cz+d} & \det g>0\\ \dfrac{a\overline{z}+b}{c\overline{z}+d} & \det g<0.\end{cases}$$

Letting $g * z := (g^{-1})^T \cdot z$ for $g \in \mathrm{GL}_2(\mathbf{R})$, we have that the association $Q \mapsto \gamma_Q$ is $\mathrm{GL}_2(\mathbf{R})$-equivariant for this action $*$ on $\overline{\mathfrak{H}}$. Furthermore, if $Q(x,y)$ is any (real) definite binary quadratic form, then the association $Q \mapsto z(Q)$ (from Lemma 2.1) is also $\mathrm{GL}_2(\mathbf{R})$-equivariant for the action $*$ on $\mathfrak{H}$. We denote by $[Q]$ (resp. $[Q]_1$) the equivalence class of Q under $\mathrm{GL}_2(\mathbf{Z})$ (resp. $\mathrm{SL}_2(\mathbf{Z})$), similarly for $[z(Q)] \in \mathrm{GL}_2(\mathbf{Z})\backslash\mathfrak{H}$ (resp. $[z(Q)]_1 \in \mathrm{SL}_2(\mathbf{Z})\backslash\mathfrak{H}$). We will identify $\mathrm{GL}_2(\mathbf{Z})\backslash\mathfrak{H}$ (resp. $\mathrm{SL}_2(\mathbf{Z})\backslash\mathfrak{H}$) with the Gauss fundamental domain $\mathcal{G} \subseteq \mathfrak{H}$ given by

$$\mathcal{G} = \{x + iy : 0 \leq x \leq \frac{1}{2}, x^2 + y^2 \geq 1\} \tag{6.1}$$

and

$$\mathcal{G}_1 = \{x + iy : |x| \leq \frac{1}{2}, x^2 + y^2 \geq 1\}, \tag{6.2}$$

respectively. We denote by $[\gamma_Q] \subseteq \mathrm{GL}_2(\mathbf{Z})\backslash\mathfrak{H}$ (resp. $[\gamma_Q]_1 \subseteq \mathrm{SL}_2(\mathbf{Z})\backslash\mathfrak{H}$) the set of points in $\mathcal{G}$ that are $\mathrm{GL}_2(\mathbf{Z})$-equivalent (resp. in $\mathcal{G}_1$ that are $\mathrm{SL}_2(\mathbf{Z})$-equivalent) to points on γ_Q. This only depends on $[Q]$ (resp. $[Q]_1$). Then, $[\gamma_{Q_0}] = [\gamma_{Q_0}]_1 = \{iy : y \geq 1\}$. The matrix

$$g_\alpha = \begin{pmatrix} \alpha^{-1} & \\ & \alpha \end{pmatrix}, \alpha \in \mathbf{R}_{>0}$$

flows the point $i \in \gamma_{Q_0}$ along the geodesic to $g_\alpha * i = \alpha^2 i$. Thus, $P g_\alpha P^{-1}$ flows $P * i$ along the geodesic γ_Q.

Suppose from now on that Q is an indefinite integral binary quadratic form with non-square discriminant. As α varies on the interval $(\epsilon_0^{-1}, 1]$, the flow by $P g_\alpha P^{-1}$ gives an n_Q-fold cover of $[\gamma_Q]_1$ for some positive integer n_Q.[6] Specifically, let

$$\mathcal{T} := \left\{ \begin{pmatrix} \alpha^{-1} & \\ & \alpha \end{pmatrix} : \alpha \geq 1 \right\},$$

$$\mathcal{T}_{>\beta} := \left\{ \begin{pmatrix} \alpha^{-1} & \\ & \alpha \end{pmatrix} : 1 \geq \alpha > \beta \right\},$$

and

$$\mathcal{T}_Q := \mathcal{T}_{>\epsilon_0}.$$

Then,

$$\begin{aligned} \pi_Q : \mathcal{T}_Q &\longrightarrow [\gamma_Q]_1 \\ g_\alpha &\longmapsto [P g_\alpha * i]_1 \end{aligned} \tag{6.3}$$

is this n_Q-fold cover. The hyperbolic measure on $[\gamma_{Q_0}]_1$ is $d\mu_{Q_0}(iy) = d^\times y = \dfrac{dy}{y}$. It is invariant under the action of the diagonal subgroup of $\mathrm{SL}_2(\mathbf{R})$. If $f : [\gamma_{Q_0}]_1 \to \mathbf{C}$ is an L^1 function, then

$$\int_{[\gamma_{Q_0}]_1} f(z) d\mu_{Q_0}(z) = 2 \int_{\mathcal{T}} f(g_\alpha * i) d^\times \alpha.$$

The hyperbolic measure μ_Q on $[\gamma_Q]_1$ is obtained by pushing μ_{Q_0} forward using P, i.e. if $f : [\gamma_Q]_1 \to \mathbf{C}$ is an L^1 function, then

$$\int_{[\gamma_Q]_1} f(z) d\mu_Q(z) = \frac{2}{n_Q} \int_{\mathcal{T}_Q} f(\pi_Q(g_\alpha)) d^\times \alpha. \tag{6.4}$$

[6] This will never be a simple cover when, for instance, Q is reciprocal.

7. Equidistribution on geodesics

In this section, we prove Theorems B and E on the equidistribution of shapes of complex cubic fields on certain geodesics of the modular surface $\mathrm{GL}_2(\mathbf{Z})\backslash\mathfrak{H}$ building on the correspondences of the previous section, [BH16, H16], and [BS14].

To begin with, once we realize that the shape of a complex cubic field K lies on the geodesic determined by its trace-zero form, we can easily show it. Indeed, take your favourite complex cubic field, or a straightforward one like $\mathbf{Q}(\sqrt[3]{2})$, and compute its shape and its trace-zero form. For $\mathbf{Q}(\sqrt[3]{2})$, this gives

$$\mathrm{sh}(\mathbf{Q}(\sqrt[3]{2})) = i2^{1/3} \quad \text{and} \quad T^{\perp\prime}_{\mathbf{Q}(\sqrt[3]{2})}(x,y) = xy = Q_0,$$

respectively. Verify that the shape lies on the corresponding geodesic, as $i2^{1/3}$ lies on the geodesic γ_{Q_0} connecting 0 to ∞. Now, note that the formation of the shape, the trace-zero form, and the associated geodesic are all $\mathrm{GL}_2(\mathbf{R})$-equivariant. Since the set of (binary cubic forms corresponding to) complex cubic fields (or, more generally, complex cubic orders) is contained in a unique $\mathrm{GL}_2(\mathbf{R})$ orbit, the shape of every complex cubic field lies on the geodesic associated to its trace-zero form.

For the matter of equidistribution, we apply the method of [BH16, H16] to the work of [BS14]. For technical reasons, we will follow [BS14] and use oriented cubic rings in our proofs.

Let $[Q]_1$ be the $\mathrm{SL}_2(\mathbf{Z})$-equivalence class of a primitive indefinite binary quadratic form

$$Q(x,y) = rx^2 + sxy + ty^2$$

of nonsquare discriminant $D = s^2 - 4rt > 0$. We may assume that $t > 0$. We introduce some relevant objects from [BS14]. Let

$$Q'(x,y) = tx^2 - sxy + ry^2$$

be the adjoint quadratic form of Q so that

$$Q'(x,y) = t(x - \theta_+ y)(x - \theta_- y).$$

Let

$$V_{\mathbf{R}} = \mathbf{R}^2, \quad V_{\mathbf{Z}} = \mathbf{Z}^2,$$
$$V_{\mathbf{R}}^{(Q)} = \left\{(x,y) \in V_{\mathbf{R}} : \frac{Q'(x,y)}{rt} > 0\right\}$$

and

$$V_{\mathbf{Z}}^{(Q)} = \left\{(x,y) \in V_{\mathbf{Z}} : \frac{Q'(x,y)}{rt} > 0, sb \equiv rc \,(\mathrm{mod}\ 3t), sc \equiv tb \,(\mathrm{mod}\ 3r)\right\}.$$

For $(b,c) \in V_{\mathbf{R}}^{(Q)}$, define

$$\Delta(b,c) := -\frac{Q'(b,c)^2 \Delta(Q)}{3r^2t^2}.$$

If $F(x,y) = ax^3 + bx^2y + cxy^2 + dy^3$ is a binary cubic form whose Hessian is a positive multiple of Q, let $v_F := (b,c) \in V_{\mathbf{R}}$. Following [BS14], we define a *twisted cubic* action $*$ of $\mathrm{GO}_Q^0(\mathbf{R})$ on $V_{\mathbf{R}}$ via

$$g * \begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{\det g}\, g^3 \begin{pmatrix} x \\ y \end{pmatrix},$$

where the latter denotes the usual matrix-vector multiplication. As in [BS14], one can verify that for $g \in \mathrm{GO}_Q^0(\mathbf{R})$ and F whose Hessian is a positive multiple of Q,

$$v_{g\cdot F} = g * v_F.$$

A crucial ingredient to our proof is the following correspondence of Bhargava–Shnidman.

Theorem 7.1 (Theorem 21 of [BS14])**.** *The isomorphism classes of oriented cubic rings whose primitive trace-zero form is Q are in natural bijection with $\mathrm{SO}_Q(\mathbf{Z})$-orbits (with respect to the action $*$) of pairs $(b,c) \in V_{\mathbf{Z}}^{(Q)}$. Under this bijection, R is the ring corresponding to the binary cubic form*

$$ax^3 + bx^2y + cxy^2 + dy^3,$$

where

$$a = \frac{sb - rc}{3t} \quad \text{and} \quad d = \frac{sc - tb}{3r}.$$

Furthermore,

$$\Delta(R) = \Delta(b, c).$$

We port over the definition of irreducible to the space $V_{\mathbf{Z}}^{(Q)}$ and say that an element $v \in V_{\mathbf{Z}}^{(Q)}$ is *irreducible* if its associated binary cubic form is irreducible. By Proposition 3.2, this is equivalent to the associated cubic ring being an integral domain, i.e. an order in a cubic field.

In order to link [BS14] to [BH16], we relate constructions in the former to the group $\mathrm{GO}_Q^0(\mathbf{R})$. Note that $\mathrm{GO}_Q^0(\mathbf{R}) \cap \mathrm{GL}_2(\mathbf{Z}) = \mathrm{SO}_Q(\mathbf{Z})$.

Lemma 7.2. *$V_{\mathbf{R}}^{(Q)}$ is a $\mathrm{GO}_Q^0(\mathbf{R})$-orbit.*

Proof. Note that for $v = (x, y) \in \mathbf{R}^2$,

$$P^{-1} \cdot \begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{\sqrt{tD}} \begin{pmatrix} x - \theta_- y \\ -(x - \theta_+ y) \end{pmatrix}.$$

For each of $+$ and $-$, let

$$V_{\mathbf{R}}^{(Q_0),\pm} = \{(x, y) \in \mathbf{R}^2 : \pm Q_0'(x, y) = \mp xy > 0\}.$$

Let $(x', y') := (x - \theta_- y, -(x - \theta_+ y))$. Then,

$$\begin{aligned} v = (x, y) \in V_{\mathbf{R}} \quad &\text{if and only if} \quad -\frac{x'y'}{r} > 0 \\ &\text{if and only if} \quad P^{-1} \cdot v \in V_{\mathbf{R}}^{(Q_0),\mathrm{sgn}(r)}. \end{aligned}$$

The lemma thus comes down to showing that each of $V_{\mathbf{R}}^{(Q_0),\pm}$ is a $\mathrm{GO}_{Q_0}^0(\mathbf{R})$-orbit.

So, let $v_0 = (1, \mp 1) \in V_{\mathbf{R}}^{(Q_0),\pm}$, then

$$g(\lambda, \alpha) * v_0 = g(\lambda^3, \alpha^3)v_0 = (\lambda^3\alpha^{-3}, \mp\lambda^3\alpha^3) \in V_{\mathbf{R}}^{(Q_0),\pm}$$

so that $V_{\mathbf{R}}^{(Q_0),\pm}$ contains an orbit. On the other hand, if $(x, y) \in V_{\mathbf{R}}^{(Q_0),\pm}$, then $(x, y) = g(\lambda, \alpha) * v_0$ for $\lambda = \sqrt{\mp xy}$ and $\alpha = \sqrt{\mp y/x}$. □

Using this lemma, and the equivariance of all related constructions, we may define the shape of an element $v \in V_{\mathbf{R}}^{(Q)}$, as follows. There is some $v_1 \in V_{\mathbf{Z}}^{(Q)}$ that corresponds to some complex cubic order R_1. There is then some $g \in \mathrm{GO}_Q^0(\mathbf{R})$ such that $g * v_1 = v$. Define $\mathrm{sh}(v) := g * \mathrm{sh}(R_1)$. This gives a well-defined $\mathrm{GO}_Q^0(\mathbf{R})$-equivariant surjective map $\mathrm{sh} : V_{\mathbf{R}}^{(Q)} \to [\gamma_Q]_1$. Note that for $g \in \mathrm{GO}_Q^0(\mathbf{R})$,

$$\Delta(g * (b, c)) = \det(g)^2 \Delta(b, c). \tag{7.1}$$

For $v = \begin{pmatrix} x \\ y \end{pmatrix} \in V_{\mathbf{R}}^{(Q)}$, define its *ratio* to be

$$\mathrm{ratio}(v) := \frac{x - \theta_- y}{x - \theta_+ y}.$$

Note that $\mathrm{ratio}(-v) = \mathrm{ratio}(v)$. For $Pg(\lambda, \alpha)P^{-1} \in \mathrm{GO}_Q^0(\mathbf{R})$, define

$$\mathrm{ratio}(Pg(\lambda, \alpha)P^{-1}) := \alpha^{-2}.$$

A calculation shows that

$$\mathrm{ratio}(Pg(\lambda, \alpha)P^{-1} \cdot v) = \mathrm{ratio}(Pg(\lambda, \alpha)P^{-1}) \cdot \mathrm{ratio}(v),$$

i.e.

$$\mathrm{ratio}(Pg(\lambda, \alpha)P^{-1} * v) = \mathrm{ratio}(Pg(\lambda, \alpha)P^{-1})^3 \cdot \mathrm{ratio}(v). \tag{7.2}$$

Therefore, every $\mathrm{SO}_Q(\mathbf{Z})$ orbit in $V_{\mathbf{Z}}^{(Q)}$ contains a unique element v such that

$$1 \le \mathrm{ratio}(\pm v) < \epsilon_0^6.$$

Every such orbit thus contains a unique element such that

$$1 \le \mathrm{ratio}(v) < \epsilon_0^6 \quad \text{and} \quad x - \theta_+ y > 0.$$

By (7.1) and (7.2), there is an element $v^{(Q)} = \begin{pmatrix} x^{(Q)} \\ y^{(Q)} \end{pmatrix} \in V_{\mathbf{R}}^{(Q)}$ such that $\Delta(v^{(Q)}) = \mathrm{ratio}(v^{(Q)}) = 1$. After possibly acting by $g(1,-1)$, we may further assume that $x^{(Q)} - \theta_+ y^{(Q)} > 0$.

Let

$$\mathcal{F}_0^{(Q)} := \{g(\lambda,\alpha) : \epsilon_0^{-1} < \alpha \le 1\}$$

and

$$\mathcal{F}^{(Q)} := P\mathcal{F}_0^{(Q)}P^{-1}.$$

We leave the proof of the following straightforward result to the reader.

Lemma 7.3. *The set $\mathcal{F}_0^{(Q)}$ is a fundamental domain for the action of $\langle \pm 1, g(1,\epsilon_0)\rangle$ on $\mathrm{GO}_{Q_0}^0(\mathbf{R})$ and so $\mathcal{F}^{(Q)}$ is a fundamental domain for the action of $\mathrm{SO}_Q(\mathbf{Z})$ on $\mathrm{GO}_Q^0(\mathbf{R})$.*

For $X > 0$, let

$$\mathcal{R}_X := \{v \in \mathcal{F}^{(Q)} * v^{(Q)} : |\Delta(v)| < X\}$$

and, for $W \subseteq [\gamma_Q]_1$ a μ_Q-continuity set, let

$$\mathcal{R}_{X,W} := \{v \in \mathcal{F}^{(Q)} * v^{(Q)} : |\Delta(v)| < X, [\mathrm{sh}(v)]_1 \in W\}.$$

For a (Lebesgue) measurable subset $T \subseteq V_{\mathbf{R}}$, let $\mathrm{Vol}(T)$ denote its Euclidean volume (i.e. its usual Lebesgue measure). As in [BH16], our counts of cubic rings and fields will be estimated by counts of lattice points in $\mathcal{R}_{X,W}$ by relating these counts to the volume of $\mathcal{R}_{X,W}$. We will require the following lemma.

Lemma 7.4. *For $X \in \mathbf{R}_{>0}$ and $W \subseteq [\gamma_Q]_1$ a μ_Q-continuity set,*

$$\mathrm{Vol}(\mathcal{R}_{X,W}) = \mathrm{Vol}(\mathcal{R}_{1,W})X^{1/2}.$$

Proof. By (7.1), the element $Pg(\lambda,1)P^{-1} = g(\lambda,1)$ acting on $v \in V_{\mathbf{R}}^{(Q)}$ scales the discriminant by λ^4. This matrix also does not affect the shape of v, so $\mathcal{R}_{X,W} = g(X^{1/4},1) * \mathcal{R}_{1,W}$. Since $\mathcal{R}_{X,W}$ is a nice (i.e. semi-algebraic) subset of a $\mathbf{R}^2$, scaling it by $X^{1/4}$ scales its volume by $X^{1/2}$, as claimed. □

We also have the following lemma that we leave to the reader.

Lemma 7.5. *As a set*

$$\mathcal{F}^{(Q)} * v^{(Q)} = \left\{ v = \begin{pmatrix} x \\ y \end{pmatrix} \in V_{\mathbf{R}}^{(Q)} : 1 \le \mathrm{ratio}(v) < \epsilon_0^6,\ x - \theta_+ y > 0 \right\}.$$

7.1. **Ratio-of-volumes calculation.** In this section, we prove a crucial technical lemma that relates the Euclidean volume $\mathrm{Vol}(\mathcal{R}_{1,W})$ to the hyperbolic measure $\mu_Q(W)$ for any μ_Q-continuity set in $[\gamma_Q]_1$. The key idea, following [BH16, §6], is to pass through the group $\mathrm{GO}_Q^0(\mathbf{R})$ which is linked to both.

Lemma 7.6. *For any μ_Q-continuity set W in $[\gamma_Q]_1$, we have that*

$$\frac{\mathrm{Vol}(\mathcal{R}_{1,W})}{\mathrm{Vol}(\mathcal{R}_1)} = \frac{\mu_Q(W)}{\mu_Q([\gamma_Q]_1)}.$$

Proof. Let $\chi_{\mathcal{R}_{1,W}}$ denote the characteristic function of $\mathcal{R}_{1,W}$. Then,

$$\frac{\mathrm{Vol}(\mathcal{R}_{1,W})}{\mathrm{Vol}(\mathcal{R}_1)} = \frac{\displaystyle\int_{g \in \mathcal{F}^{(Q)}} \chi_{\mathcal{R}_{1,W}}(g * v^{(Q)}) J(g) dg}{\displaystyle\int_{g \in \mathcal{F}^{(Q)}} \chi_{\mathcal{R}_1}(g * v^{(Q)}) J(g) dg},$$

where $J(g)$ is an appropriate function accounting for the Jacobian determinant of the map $\mathcal{F}^{(Q)} \to \mathcal{R}_\infty$, and dg denotes the Haar measure on $\mathrm{GO}_Q^0(\mathbf{R})$. Recall that

$$\mathrm{GO}_Q^0(\mathbf{R}) = P\{g(\lambda,\alpha) : \lambda \in \mathbf{R}_{>0}, \alpha \in \mathbf{R}^\times\}P^{-1} \cong \mathbf{R}_{>0} \times \mathbf{R}^\times,$$

so that $dg = d^\times\lambda d^\times\alpha$. The Jacobian of multiplication by a (constant) matrix is the matrix itself, the Jacobian of an inverse function is the inverse of the Jacobian of the function, and the Jacobian of a composition is the product of the Jacobians, so that the Jacobian determinants of P and P^{-1} cancel. To determine $J(g)$, it thus suffices to determine $\left|\frac{\partial(x,y)}{\partial(\lambda,\alpha)}\right|$, where

$$x = \lambda\alpha^{-3}x^{(Q)}$$
$$y = \lambda\alpha^3 y^{(Q)}.$$

A straightforward calculation shows that

$$\left|\frac{\partial(x,y)}{\partial(\lambda,\alpha)}\right| = 6x^{(Q)}y^{(Q)}\frac{\lambda}{\alpha},$$

which implies that $J(g) = \det(g)$. With this is hand, we have that

$$\frac{\mathrm{Vol}(\mathcal{R}_{1,W})}{\mathrm{Vol}(\mathcal{R}_1)} = \frac{\displaystyle\int_{g(\lambda,\alpha)\in\mathcal{F}_0^{(Q)}} \chi_{\mathcal{R}_{1,W}}(Pg(\lambda,\alpha)P^{-1} * v^{(Q)})\lambda^2 d^\times\lambda d^\times\alpha}{\displaystyle\int_{g(\lambda,\alpha)\in\mathcal{F}_0^{(Q)}} \chi_{\mathcal{R}_1}(Pg(\lambda,\alpha)P^{-1} * v^{(Q)})\lambda^2 d^\times\lambda d^\times\alpha}.$$

By (7.1), $\Delta(Pg(\lambda,\alpha)P^{-1} * v^{(Q)}) = \lambda^4\Delta(v^{(Q)}) = \lambda^4$. The characteristic functions therefore impose the condition $0 < \lambda \leq 1$. We also know that the λ factor does not affect the shape, so

$$\begin{aligned}\frac{\mathrm{Vol}(\mathcal{R}_{1,W})}{\mathrm{Vol}(\mathcal{R}_1)} &= \frac{\displaystyle\int_0^1 \lambda^2 d^\times\lambda \int_{\epsilon_0^{-1}}^1 \chi_{\mathcal{R}_{1,W}}(Pg(1,\alpha)P^{-1} * v^{(Q)})d^\times\alpha}{\displaystyle\int_0^1 \lambda^2 d^\times\lambda \int_{\epsilon_0^{-1}}^1 \chi_{\mathcal{R}_1}(Pg(1,\alpha)P^{-1} * v^{(Q)})d^\times\alpha} \\ &= \frac{\displaystyle\int_{\epsilon_0^{-1}}^1 \chi_{\mathcal{R}_{1,W}}(Pg(1,\alpha)P^{-1} * v^{(Q)})d^\times\alpha}{\displaystyle\int_{\epsilon_0^{-1}}^1 \chi_{\mathcal{R}_1}(Pg(1,\alpha)P^{-1} * v^{(Q)})d^\times\alpha}.\end{aligned}$$

By Lemma 7.5, the value of $\chi_{\mathcal{R}_1}$ is constant on the region of integration here. On the other hand, the value of $\chi_{\mathcal{R}_{1,W}}(Pg(1,\alpha)P^{-1} * v^{(Q)})$ simply depends on whether the shape of $Pg(1,\alpha)P^{-1} * v^{(Q)}$ is in W. The shape of $v^{(Q)}$ is some element $z^{(Q)} \in [\gamma_Q]_1$. There is therefore some α_0 such that $z^{(Q)} = Pg(1,\alpha_0) * i$. Let $\widetilde{W} \subseteq \mathcal{T}_Q$ be the inverse of image of W under the map π_Q from (6.3). Since

$$[\mathrm{sh}(Pg(1,\alpha)P^{-1} * v^{(Q)})]_1 = \pi_Q(g(1,\alpha)g(1,\alpha_0)),$$

we have that

$$\begin{aligned}[\mathrm{sh}(Pg(1,\alpha)P^{-1} * v^{(Q)})]_1 \in W &\text{ if and only if } g(1,\alpha)g(1,\alpha_0) \in \widetilde{W} \\ &\text{if and only if } g(1,\alpha) \in g(1,\alpha_0^{-1})\widetilde{W}\end{aligned}$$

(where we are identifying $\mathcal{T}_Q$ with $\mathcal{T}/\langle g(1,\epsilon_0^{-1})\rangle$). We now have that

$$\frac{\mathrm{Vol}(\mathcal{R}_{1,W})}{\mathrm{Vol}(\mathcal{R}_1)} = \frac{\displaystyle\int_{\epsilon_0^{-1}}^1 \chi_{g(1,\alpha_0^{-1})\widetilde{W}}(g(1,\alpha))d^\times\alpha}{\displaystyle\int_{\epsilon_0^{-1}}^1 d^\times\alpha}.$$

The measure on $\mathcal{T}$ is invariant under multiplication, so we may replace $\chi_{g(1,\alpha_0^{-1})\widetilde{W}}$ with $\chi_{\widetilde{W}}$. Note that $\chi_{\widetilde{W}}(g(1,\alpha)) = \chi_W(\pi_Q(g(1,\alpha))$. Using (6.4), we obtain

$$\begin{aligned}\frac{\mathrm{Vol}(\mathcal{R}_{1,W})}{\mathrm{Vol}(\mathcal{R}_1)} &= \frac{\frac{2}{n_Q}\int_{[\gamma_Q]_1}\chi_W(z)d\mu_Q(z)}{\frac{2}{n_Q}\int_{[\gamma_Q]_1}d\mu_Q(z)}\\ &= \frac{\mu_Q(W)}{\mu_Q([\gamma_Q]_1)},\end{aligned}$$

as desired. □

7.2. **Equidistribution for oriented complex cubic orders.** The next step is to count the number of oriented complex cubic orders of bounded discriminant and shape in some region by relating it to the volume of $\mathcal{R}_{X,W}$. Together with Lemma 7.6, this will prove equidistribution for shapes of (oriented) complex cubic orders. The result for fields will then follow, in the next section, from a sieve.

For an $\mathrm{SO}_Q(\mathbf{Z})$-stable subset S of $V_{\mathbf{Z}}^{(Q)}$ and $X > 0$, let

$$N^{\mathrm{Or}}(S;X) := \#\{[v]_1 \in \mathrm{SO}_Q(\mathbf{Z})\backslash S : v \text{ is irreducible}, |\Delta(v)| < X\}$$

and for $W \subseteq [\gamma_Q]_1$ a μ_Q-continuity set, let

$$N^{\mathrm{Or}}(S;X,W) := \#\{[v]_1 \in \mathrm{SO}_Q(\mathbf{Z})\backslash S : v \text{ is irreducible}, |\Delta(v)| < X, [\mathrm{sh}(v)]_1 \in W\}.$$

For a subset $S \subseteq \mathbf{Z}^2 \subseteq V_{\mathbf{R}}$, let $\mu_p(S)$ denote its p-adic density, i.e. the measure of its closure in $\mathbf{Z}_p$ (where $\mathbf{Z}_p$ is given its usual Haar probability measure). In [BS14, p. 74], Bhargava and Shnidman prove[7] the analogue of [BH16, Theorem 8], namely the following.

Theorem 7.7 ([BS14])**.** *If $S \subseteq V_{\mathbf{Z}}^{(Q)}$ is any $\mathrm{SO}_Q(\mathbf{Z})$-stable subset defined by finitely many congruence conditions, then*

$$\begin{aligned}N^{\mathrm{Or}}(S;X) &= \prod_p \mu_p(S)\cdot\mathrm{Vol}(\mathcal{R}_X) + O(X^{1/4})\\ &= \prod_p \mu_p(S)\cdot\mathrm{Vol}(\mathcal{R}_1)X^{1/2} + O(X^{1/4}).\end{aligned}$$

The volume calculation starting on p. 73 of [BS14] shows that

$$\mathrm{Vol}(\mathcal{R}_X) = \frac{3\sqrt{3}rt\log\epsilon_0}{D}X^{1/2}$$

and [BS14, Lemma 26] shows that

$$\prod_p \mu_p(V_{\mathbf{Z}}^{(Q)}) = \frac{1}{3^{\alpha_D}rt},$$

where

$$\alpha_D := \begin{cases}1 & 3 \mid D\\ 2 & 3 \nmid D.\end{cases}$$

Define

$$C_D := \frac{3\sqrt{3}\log\epsilon_0}{3^{\alpha_D}D}. \tag{7.3}$$

We therefore have the following consequence which is [BS14, Theorem 29]: the number of oriented complex cubic orders (up to isomorphism) whose primitive trace-zero form is $\mathrm{SO}_Q(\mathbf{Z})$-equivalent to Q is

$$N^{\mathrm{Or}}(V_{\mathbf{Z}}^{(Q)};X) = C_D X^{1/2} + O(X^{1/4}).$$

[7] This proof is in some sense implicit in their discussion and is given as an analogue of their Lemma 24 and Theorem 25.

To obtain an analogue of this result for cubic orders whose shape lies in some W, we will use what we call the "W–$\overline{W}$ trick" from [BH16, §3]. We will require the following lemma.

Lemma 7.8. *Let H be a bounded measurable subset of $V_{\mathbf{R}}^{(Q)}$ and let $S \subseteq V_{\mathbf{Z}}^{(Q)}$ be given by finitely many congruence conditions. Then, for any $z \in \mathbf{R}_{>0}$, the number of irreducible lattice points in $S \cap zH$ is*

$$\left(\prod_p \mu_p(S)\right) \mathrm{Vol}(zH) + O(z).$$

Proof. As in [BH16, Lemma 6 and Lemma 9], what is required is to show that the number of *reducible* lattices points in $S \cap zH$ is negligible. Bhargava–Shnidman prove that the number of $\mathrm{SL}_2(\mathbf{Z})$-equivalence classes of binary cubic forms F whose Hessian is a (positive) integer multiple of Q with $|\Delta(F)| < X$ is $O(X^{1/4})$.[8] Since the discriminant of $(b, c) \in V_{\mathbf{R}}^{(Q)}$ is homogeneous of degree 4 in b and c, for any bounded, measurable subset H in $V_{\mathbf{R}}$, the number of irreducible lattice points in zH is $\mathrm{Vol}(zH) + O(z)$, as claimed. □

Proposition 7.9. *Let $W \subseteq [\gamma_Q]_1$ be a μ_Q-continuity set. Suppose that $S \subseteq V_{\mathbf{Z}}^{(Q)}$ is given by finitely many congruence conditions, then*

$$N^{\mathrm{Or}}(S; X, W) = \left(\prod_p \mu_p(S)\right) \cdot \mathrm{Vol}(\mathcal{R}_{1,W}) X^{1/2} + O(X^{1/4}).$$

Proof. Let $\epsilon > 0$ and let $\mathcal{R}'_{1,W}$ be a bounded, measurable subset of $\mathcal{R}_{1,W}$ such that

$$\mathrm{Vol}(\mathcal{R}'_{1,W}) \geq \mathrm{Vol}(\mathcal{R}_{1,W}) - \epsilon.$$

Define $R'_{X,W} := X^{1/4} R'_{1,W}$. By Lemma 7.8, the number of irreducible points in $S \cap \mathcal{R}'_{X,W}$ is

$$\left(\prod_p \mu_p(S)\right) \mathrm{Vol}(R'_{1,W}) X^{1/2} + O(X^{1/4}).$$

Since $\mathcal{R}_{X,W} \supseteq \mathcal{R}'_{X,W}$, we have that

$$N^{\mathrm{Or}}(S; X, W) \geq \left(\prod_p \mu_p(S)\right) (\mathrm{Vol}(\mathcal{R}_{1,W}) - \epsilon) X^{1/2} + O(X^{1/4}).$$

This is true for all $\epsilon > 0$, so that, in fact,

$$N^{\mathrm{Or}}(S; X, W) \geq \left(\prod_p \mu_p(S)\right) \mathrm{Vol}(\mathcal{R}_{1,W}) X^{1/2} + O(X^{1/4}). \tag{7.4}$$

Let $\overline{W} := [\gamma_Q]_1 \setminus W$ be the complement of W. We similarly obtain that

$$N^{\mathrm{Or}}(S; X, \overline{W}) \geq \left(\prod_p \mu_p(S)\right) \mathrm{Vol}(\mathcal{R}_{1,\overline{W}}) X^{1/2} + O(X^{1/4}). \tag{7.5}$$

[8] This is the analogue of [BS14, Lemma 24] mentioned on page 74 of *ibid.*

Then,

$$\begin{aligned}\left(\prod_p \mu_p(S)\right)\mathrm{Vol}(\mathcal{R}_1)X^{1/2}+O(X^{1/4}) &= N(S;X)\\ &= N^{\mathrm{Or}}(S;X,W)+N^{\mathrm{Or}}(S;X,\overline{W})\\ &\geq \mathrm{Vol}(\mathcal{R}_{1,W})X^{1/2}+\mathrm{Vol}(\mathcal{R}_{1,\overline{W}})X^{1/2}+O(X^{1/4})\\ &= \left(\prod_p \mu_p(S)\right)\mathrm{Vol}(\mathcal{R}_1)X^{1/2}+O(X^{1/4}).\end{aligned} \tag{7.6}$$

Thus, the inequality (7.6), and hence also (7.4) and (7.5), are all equalities. □

As a corollary, we obtain the equidistribution for shapes of oriented complex cubic orders. Indeed, for $X>0$ and a μ_Q-continuity set $W\subseteq[\gamma_Q]_1$, let $N^{\mathrm{Or}}_{\mathrm{rings}}(Q;X,W)$ denote the number of oriented complex cubic orders R (up to isomorphism) with $[T_R^{\perp\prime}]_1=[Q]_1$, $|\Delta(R)|<X$, and $[\mathrm{sh}(R)]_1\in W$. Then, $N^{\mathrm{Or}}_{\mathrm{rings}}(Q;X,W)=N^{\mathrm{Or}}(V_{\mathbf{Z}}^{(Q)};X,W)$. We thus obtain the following.

Theorem 7.10. *For every μ_Q-continuity set $W\subseteq[\gamma_Q]_1$,*

$$\lim_{X\to\infty}\frac{N^{\mathrm{Or}}_{\mathrm{rings}}(Q;X,W)}{N^{\mathrm{Or}}_{\mathrm{rings}}(Q;X)}=\frac{\mu_Q(W)}{\mu_Q([\gamma_Q]_1)}. \tag{7.7}$$

7.3. **Equidistribution for oriented complex cubic fields.** Turning (7.7) into a statement about *maximal* cubic orders requires a sieve adapted from [DH71, §5], as in [BH16, §5]. According to [BS14], the indefinite integral binary quadratic forms arising as the primitive trace-zero forms of some cubic field are exactly those with discriminant D such that D or $-D/3$ is a fundamental discriminant. Those whose D is not square correspond to closed geodesics, while $D=1$ or 9 correspond to vertical non-closed geodesics. We exclude the cases $D=1$ and $D=9$ as they correspond to the pure cubic fields and have been dealt with in [Har17] (and must, in fact, be dealt with separately since the geodesics have infinite measure). We therefore assume Q is such that D or $-D/3$ is fundamental and D is non-square.

Let $N^{\mathrm{Or}}(Q;X,W)$ denote the number of oriented complex cubic fields K (up to isomorphism) with $[T_K^{\perp\prime}]_1=[Q]_1$, $|\Delta(K)|<X$, and $[\mathrm{sh}(K)]_1\in W$. As described in [BS14, §4.3], for a prime p, those $v=(b,c)\in V_{\mathbf{Z}}^{(Q)}$ that are p-maximal are defined by congruences modulo p^2. Let $S_p\subseteq V_{\mathbf{Z}}^{(Q)}$ denote the set of elements that correspond to p-maximal rings. For $Y>2$, let

$$S_{<Y}:=\bigcap_{\substack{p\text{ prime}\\ p<Y}}S_p$$

and let

$$S_{\max}:=\bigcap_{p\text{ prime}}S_p.$$

Then, $N^{\mathrm{Or}}(Q;X,W)=N^{\mathrm{Or}}(S_{\max};X,W)$. Since $S_{\max}$ is given by infinitely many congruence conditions, Proposition 7.9 does not apply. However, for each Y, $S_{<Y}$ is given by finitely many congruences conditions, so we have that

$$N^{\mathrm{Or}}(S_{<Y};X,W)=\left(\prod_p\mu_p(S_{<Y})\right)\cdot\mathrm{Vol}(\mathcal{R}_{1,W})X^{1/2}+O(X^{1/4}).$$

Let $\overline{S}_p:=V_{\mathbf{Z}}^{(Q)}\setminus S_p$ be the complement of S_p. By the discussion on page 78 of [BS14],

$$N^{\mathrm{Or}}(\overline{S}_p;X)=O(X^{1/2}/p^2).$$

We may now follow the sieve of [DH71, §5].

Theorem 7.11. *For any μ_Q-continuity set $W \subseteq [\gamma_Q]_1$,*

$$N^{\mathrm{Or}}(Q; X, W) = \left(\prod_p \mu_p(S^{max}) \right) \mathrm{Vol}(\mathcal{R}_{1,W}) X^{1/2} + o(X^{1/2}).$$

Proof. For any positive integer Y, $N(S^{\max}; X, W) \leq N(S_{<Y}; X, W)$ so that

$$\begin{aligned} \limsup_{X \to \infty} \frac{N(S^{\max}; X, W)}{X^{1/2}} &\leq \lim_{Y \to \infty} \lim_{X \to \infty} \frac{N(S_{<Y}; X, W)}{X^{1/2}} \\ &\leq \left(\prod_p \mu_p(S^{\max}) \right) \mathrm{Vol}(\mathcal{R}_{1,W}). \end{aligned}$$

Conversely,

$$S_{<Y} \cap \mathcal{R}_{X,W} \subseteq (S^{\max} \cap \mathcal{R}_{X,W}) \cup \bigcup_{p \geq Y} \overline{S}_p$$

so that

$$\begin{aligned} \liminf_{X \to \infty} \frac{N(S^{\max}; X, W)}{X^{1/2}} &\geq \lim_{Y \to \infty} \lim_{X \to \infty} \frac{N(S_{<Y}; X, W)}{X^{1/2}} - O\left(\sum_{p \geq Y} \frac{N^{\mathrm{Or}}(\overline{S}_p; X)}{X^{1/2}} \right) \\ &\geq \lim_{Y \to \infty} \lim_{X \to \infty} \frac{N(S_{<Y}; X, W)}{X^{1/2}} - O\left(\sum_{p \geq Y} \frac{1}{p^2} \right) \\ &\geq \left(\prod_p \mu_p(S^{\max}) \right) \mathrm{Vol}(\mathcal{R}_{1,W}) \end{aligned}$$

since $\sum_{p \geq Y} p^{-2}$ is the tail of a convergent series. □

This, together with Lemma 7.6, implies the equidistribution of shapes of oriented complex cubic fields on $[\gamma_Q]_1$.

Corollary 7.12. *For any μ_Q-continuity set $W \subseteq [\gamma_Q]_1$,*

$$\lim_{X \to \infty} \frac{N^{\mathrm{Or}}(Q; X, W)}{N^{\mathrm{Or}}(Q; X)} = \frac{\mu_Q(W)}{\mu_Q([\gamma_Q]_1)}.$$

7.4. **Equidistribution for non-oriented cubic fields.** Finally, we can use the result of the previous section to prove Theorem E, i.e. equidistribution for (non-oriented) cubic fields and $\mathrm{GL}_2(\mathbf{Z})$-equivalence, expanding upon the discussion of [BS14, pp. 55–56].

For concreteness in this discussion, we view $[\gamma_Q]$ and $[\gamma_Q]_1$ as subsets of the Gauss fundamental domains $\mathcal{G}$ and $\mathcal{G}_1$ of (6.1) and (6.2), respectively. For $W \subseteq [\gamma_Q]$ a μ_Q-continuity set[9] and $X > 0$, let $N(Q; X, W)$ denote the number of (isomorphism classes of) complex cubic fields K with $[T_K^{\perp\prime}] = [Q]$, $|\Delta(K)| < X$, and $[\mathrm{sh}(K)] \in W$. Similarly, for $N_{\mathrm{rings}}(Q; X, W)$ and orders in complex cubic fields.

First note that each isomorphism class of complex cubic orders corresponds to two isomorphism classes of oriented complex cubic orders (because of the two possible orderings of a basis). If Q is ambiguous, then $[Q]_1 = [Q]$. Otherwise, $[Q]$ is the union of two $\mathrm{SL}_2(\mathbf{Z})$-equivalence classes $[Q]_1$ and $[wQ]_1$, where

$$w := \begin{pmatrix} -1 & \\ & 1 \end{pmatrix}.$$

[9] By abuse of notation, we use μ_Q to denote the hyperbolic measures on both $[\gamma_Q]_1$ and $[\gamma_Q]$.

Geometrically, when Q is ambiguous $[\gamma_Q]_1$ is the mirror image of itself across the imaginary axis, so that the map $[z]_1 \mapsto [z]$ gives a double cover of $[\gamma_Q]$ by $[\gamma_Q]_1$. Therefore,

$$\mu_Q([\gamma_Q]) = \frac{\mu_Q([\gamma_Q]_1)}{2}. \tag{7.8}$$

When Q is not ambiguous, this is not the case; instead $[\gamma_Q]_1$ is the mirror image of $[\gamma_{wQ}]_1$. The map $[z]_1 \mapsto [z]$ then gives a double cover from $[\gamma_Q]_1 \cup [\gamma_{wQ}]_1$ to $[\gamma_Q]$. Therefore,

$$\mu_Q([\gamma_Q]) = \frac{\mu_Q([\gamma_Q]_1) + \mu_{wQ}([\gamma_{wQ}]_1)}{2} = \mu_Q([\gamma_Q]_1). \tag{7.9}$$

To deal with the translation between oriented and non-oriented orders, we must therefore deal with two cases depending on whether Q is ambiguous.

Suppose first that Q is not ambiguous. Let R be a complex cubic order contained in a non-pure cubic field and suppose $[T_R^{\perp\prime}] = [Q]$. Then, R corresponds to two isomorphism classes of oriented cubic orders (R, δ_1) and (R, δ_2). Let us denote their shapes $[\mathrm{sh}_1]_1$ and $[\mathrm{sh}_2]_1$, respectively. Then, without loss of generality, $[\mathrm{sh}_1] \in [\gamma_Q]_1$ and $[\mathrm{sh}_2] \in [\gamma_{wQ}]_1$ (and $[\mathrm{sh}\,1] = [\mathrm{sh}_2] = [\mathrm{sh}(R)]$). Suppose $W \subseteq [\gamma_Q]$ is a μ_Q-continuity set. There are continuity sets $W_1 \subseteq [\gamma_Q]_1$ and $W_2 \subseteq [\gamma_{wQ}]_1$ such that $\widetilde{W} := W_1 \cup W_2 \subseteq [\gamma_Q]_1 \cup [\gamma_{wQ}]_1$ is the double-cover of W under the map $[z]_1 \mapsto [z]$ (and note that, as subsets of $\mathcal{G}$, $W_2 = w * W_1$, so that $\mu_Q(W_1) = \mu_{wQ}(W_2)$). Then,

$$\mu_Q(W) = \frac{\mu_Q(W_1) + \mu_{wQ}(W_2)}{2} = \mu_Q(W_1). \tag{7.10}$$

Note that

$$[\mathrm{sh}(R)] \in W \quad \text{if and only if} \quad [\mathrm{sh}_1]_1 \in W_1 \text{ and } [\mathrm{sh}_2]_1 \in W_2$$

so that

$$N(Q; X, W) = N^{\mathrm{Or}}(Q; X, W_1) = N^{\mathrm{Or}}(wQ; X, W_2)$$

(and similarly for $N_{\mathrm{rings}}(Q; X, W)$). Then,

$$\frac{N(Q; X, W)}{N(Q; X)} = \frac{N^{\mathrm{Or}}(Q; X, W_1)}{N^{\mathrm{Or}}(Q; X)}. \tag{7.11}$$

Now, suppose that Q is ambiguous. As in the previous paragraph, we get (R, δ_i) and $[\mathrm{sh}_i]_1$, for $i = 1, 2$. Without loss of generality, suppose $[\mathrm{sh}_1]_1 \in \mathcal{G}$. Let $\widetilde{W}$ be the inverse image of W under the double-cover $[\gamma_Q]_1 \to [\gamma_Q]$. Then, $\widetilde{W} = W \cup w * W$, so that

$$\mu_Q(W) = \frac{\mu_Q(\widetilde{W})}{2}. \tag{7.12}$$

Now,

$$[\mathrm{sh}(R)] \in W \quad \text{if and only if} \quad [\mathrm{sh}_1]_1 \in W \text{ and } [\mathrm{sh}_2]_1 \in w * W.$$

So, it certainly looks like we are double-counting everything when considering oriented rings and $\mathrm{SL}_2(\mathbf{Z})$-equivalence. It is however possible that sometimes $[\mathrm{sh}_1]_1 = [\mathrm{sh}_2]_1$. Luckily, this only happens when $[\mathrm{sh}(R)]$ lies on the boundary of $\mathcal{G}$ and Theorem F shows that this does not occur for shapes of orders in non-pure complex cubic fields! We thus have that

$$N(Q; X, W) = \frac{1}{2} N^{\mathrm{Or}}(Q; X, \widetilde{W}) \tag{7.13}$$

(and similarly for $N_{\mathrm{rings}}(Q; X, W)$).

Therefore, when Q is not ambiguous, combining (7.9), (7.10), and (7.11) with Corollary 7.12, we have that

$$\lim_{X \to \infty} \frac{N(Q; X, W)}{N(Q; X)} = \lim_{X \to \infty} \frac{N^{\mathrm{Or}}(Q; X, W_1)}{N^{\mathrm{Or}}(Q; X)} = \frac{\mu_Q(W_1)}{\mu_Q([\gamma_Q]_1)} = \frac{\mu_Q(W)}{\mu_Q([\gamma_Q])}.$$

And when Q is ambiguous, we can similarly combine (7.8), (7.12), and (7.13) with Corollary 7.12 to obtain that

$$\lim_{X\to\infty}\frac{N(Q;X,W)}{N(Q;X)}=\lim_{X\to\infty}\frac{\frac{1}{2}N^{\mathrm{Or}}(Q;X,\widetilde{W})}{\frac{1}{2}N^{\mathrm{Or}}(Q;X)}=\frac{\frac{1}{2}\mu_Q(\widetilde{W})}{\frac{1}{2}\mu_Q([\gamma_Q]_1)}=\frac{\mu_Q(W)}{\mu_Q([\gamma_Q])}.$$

We have therefore proved Theorem E, as well as its analogue for orders which he state here.

Theorem 7.13. *For every μ_Q-continuity set $W\subseteq[\gamma_Q]$,*

$$\lim_{X\to\infty}\frac{N_{\mathrm{rings}}(Q;X,W)}{N_{\mathrm{rings}}(Q;X)}=\frac{\mu_Q(W)}{\mu_Q([\gamma_Q])}.$$

8. Generalization to higher degree

In this section, we prove Theorem C, or rather a slight generalization: the shape of any order $\mathcal{O}$ in a finite étale $\mathbf{Q}$-algebra lies on the majorant space of the trace-zero form of $\mathcal{O}$. This gives a vast generalization of Theorem B.

Let $\mathcal{A}$ be an n-dimensional étale $\mathbf{Q}$-algebra. More concretely, $\mathcal{A}\cong K_1\times\cdots\times K_r$, where the K_i are algebraic number fields. Let $d_i:=\deg K_i$, so that $n=\sum_{i=1}^{r}d_i$. Each K_i has d_i embeddings into $\mathbf{C}$ which we will denote $\sigma_{i,1},\cdots,\sigma_{i,d_i}$. Then, as in the case of number fields, we have an embedding

$$\begin{array}{rcl} j:\mathcal{A} & \to & \mathbf{C} \\ (a_1,\dots,a_r) & \mapsto & (\sigma_{i,k}(a_i))_{i,k}\end{array}$$

that embeds $\mathcal{A}$ into an n-dimensional Euclidean space $\mathcal{A}_{\mathbf{R}}$. Let $\mathcal{O}$ be an order in $\mathcal{A}$. We can then once again speak of the shape $M_{\mathcal{O}}^{\perp}$ and the trace-zero form $T_{\mathcal{O}}^{\perp}$ of $\mathcal{O}$. Let r_1 be the number of real embeddings of $\mathcal{A}$ into $\mathbf{C}$ and r_2 be the number of pairs of complex embeddings. A result of Olga Taussky-Todd [Tau68] shows that the trace form of $\mathcal{O}$ has signature (r_1+r_2,r_2). The trace-zero form $T_{\mathcal{O}}^{\perp}$ then has signature (r_1+r_2-1,r_2). Let $O(T_{\mathcal{O}}^{\perp})$ be the orthogonal group of $T_{\mathcal{O}}^{\perp}$. Let T be the Gram matrix of $T_{\mathcal{O}}^{\perp}$ (with respect to some integral basis of $\mathcal{O}^{\perp}$). Siegel [Sie67, §3.2] defines the *majorant space*[10] $\mathfrak{H}_{\mathcal{O}}$ of T as the set of positive definite matrices M such that

$$MT^{-1}M=T.$$

This is a model for the symmetric space of $O(T_{\mathcal{O}}^{\perp})$. A direct generalization of Theorem B would say that the shape of $\mathcal{O}$ "lies on" $\mathfrak{H}_{\mathcal{O}}$. One can verify "by hand" that this equality holds for the Gram matrices (3.7) and (3.10) in the cubic case, thus providing another proof of Theorem B. By very slightly modifying the work of [Tau68], we show this holds in full generality.

Theorem 8.1. *Let T and M be the Gram matrices of $T_{\mathcal{O}}^{\perp}$ and $M_{\mathcal{O}}^{\perp}$, respectively, with respect to some integral basis of $\mathcal{O}^{\perp}$. Then,*

$$MT^{-1}M=T.$$

Proof. In fact, we will prove this equality holds even before taking orthogonal projections. Let $\alpha_0,\alpha_1,\dots,\alpha_{n-1}$ be an integral basis of $\mathcal{O}$ and let $\widetilde{T}$ and $\widetilde{M}$ be the Gram matrices of the trace form and the Minkowski inner product, respectively, with respect to this basis. Let $v_i=j(\alpha_i)\in\mathbf{C}^n$. We can think of the v_i as column vectors and let A be the $n\times n$-matrix

$$A=\begin{pmatrix} v_0 & v_1 & \cdots & v_{n-1}\end{pmatrix}.$$

Then,

$$\widetilde{M}=\overline{A}^TA\quad\text{and}\quad\widetilde{T}=A^TA. \tag{8.1}$$

First, note that the equality we wish to prove is independent of the choice of basis. Indeed, if $P\in\mathrm{GL}_n(\mathbf{R})$, $\widetilde{M}_1=P^T\widetilde{M}P$, $\widetilde{T}_1=P^T\widetilde{T}P$, and $\widetilde{M}_1\widetilde{T}_1^{-1}\widetilde{M}_1=\widetilde{T}_1$, then

$$P^T\widetilde{T}P=\widetilde{T}_1=\widetilde{M}_1\widetilde{T}_1^{-1}\widetilde{M}_1=P^T\widetilde{M}\widetilde{T}^{-1}\widetilde{M}P$$

[10] Siegel refers to it as the $\mathfrak{H}$-space.

so that $\widetilde{T} = \widetilde{M}\widetilde{T}^{-1}\widetilde{M}$ since P is invertible.

Second, note that what we just explained shows that the identity being true for $\widetilde{M}$ and $\widetilde{T}$ implies its truth for M and T. Indeed, let us choose $\alpha_0 = 1$ (which is always possible), then the matrix bringing $v_0, v_1, \dots, v_{n-1}$ to $v_0, v_1^{\perp}, \dots, v_{n-1}^{\perp}$ is some $P \in \mathrm{GL}_n(\mathbf{R})$, and $P^T\widetilde{M}P = 1 \oplus M$ and $P^T\widetilde{T}P = 1 \oplus T$.

Now, we proceed as in [Tau68], including the details here for the convenience of the reader. The matrix A has r_1 real rows and r_2 pairs of complex conjugate rows. There is therefore an explicit $R \in \mathrm{GL}_n(\mathbf{R})$ such that RA has the same real rows and each pair of complex conjugate rows with entries $a_k + ib_k, a_k - ib_k$, respectively, becomes a pair of rows with entries a_k, ib_k, respectively. After possibly reordering the rows (which amounts to replacing R with another matrix in $\mathrm{GL}_n(\mathbf{R})$), we may assume that the first $r_1 + r_2$ rows of RA are real and that the last r_2 are purely imaginary. Let D be the diagonal matrix whose first $r_1 + r_2$ entries are 1 and whose remaining r_2 entries are i. Then, DRA is some invertible real matrix whose inverse we will denote by P. Write

$$R^{-T}R^{-1} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$$

where B_{11} is an $(r_1 + r_2) \times (r_1 + r_2)$-matrix and B_{22} is an $r_2 \times r_2$-matrix. Then,

$$D^{-T}R^{-T}R^{-1}D^{-1} = \begin{pmatrix} B_{11} & -iB_{12} \\ -iB_{21} & -B_{22} \end{pmatrix}$$

so that B_{12} and B_{21} are 0, since $D^{-T}R^{-T}R^{-1}D^{-1} = P^TA^TAP$, which is a real matrix. Since $R^{-T}R^{-1}$ is positive definite, so are the B_{kk}. Let $\widetilde{T}_1 = P^T\widetilde{T}P$ and $\widetilde{M}_1 = P^T\widetilde{M}P$. Then, by (8.1) and what we have just showed,

$$\begin{aligned} \widetilde{M}_1\widetilde{T}_1^{-1}\widetilde{M}_1 &= \begin{pmatrix} B_{11} & \\ & B_{22} \end{pmatrix} \cdot \begin{pmatrix} B_{11}^{-1} & \\ & -B_{22}^{-1} \end{pmatrix} \cdot \begin{pmatrix} B_{11} & \\ & B_{22} \end{pmatrix} \\ &= \begin{pmatrix} B_{11} & \\ & -B_{22} \end{pmatrix} \\ &= \widetilde{T}_1, \end{aligned}$$

as desired. □

This leads to the following natural question: fix a primitive quadratic form Q over $\mathbf{Z}$ in $n-1$ variables (that arises as $T_K^{\perp\prime}$ for some number field K) of signature $(r_1 + r_2 - 1, r_2)$ (where $n = r_1 + 2r_2$), are the shapes of the degree n number fields K with signature (r_1, r_2) whose $T_K^{\perp\prime}$ is (equivalent to) Q equidistributed on $\mathfrak{H}_Q$?

Acknowledgments. The author would like to thank Manjul Bhargava, Asaf Hadari, Piper H, and Akshay Venkatesh for some helpful conversations.

References

[Bai80] Andrew Marc Baily, *On the density of discriminants of quartic fields*, J. Reine Angew. Math. **315** (1980), 190–210. MR 564533

[BH16] Manjul Bhargava and Piper H, *The equidistribution of lattice shapes of rings of integers in cubic, quartic, and quintic number fields*, Compositio Mathematica **152** (2016), no. 6, 1111–1120. MR 3518306

[BS14] Manjul Bhargava and Ariel Shnidman, *On the number of cubic orders of bounded discriminant having automorphism group C_3, and related problems*, Algebra & Number Theory **8** (2014), no. 1, 53–88. MR 3207579

[BST13] Manjul Bhargava, Arul Shankar, and Jacob Tsimerman, *On the Davenport–Heilbronn theorems and second order terms*, Inventiones Mathematicae **193** (2013), no. 2, 439–499. MR 3090184

[CM11] Henri Cohen and Anna Morra, *Counting cubic extensions with given quadratic resolvent*, Journal of Algebra **325** (2011), 461–478. MR 2745550 (2012b:11168)

[DF64] B. N. Delone and D. K. Faddeev, *The theory of irrationalities of the third degree*, Translations of Mathematical Monographs, Vol. 10, American Mathematical Society, Providence, R.I., 1964. MR 0160744

[DH71] H. Davenport and H. Heilbronn, *On the density of discriminants of cubic fields. II*, Proc. Roy. Soc. London Ser. A **322** (1971), no. 1551, 405–420. MR 0491593

[EW11] Manfred Einsiedler and Thomas Ward, *Ergodic theory with a view towards number theory*, Graduate Texts in Mathematics, vol. 259, Springer-Verlag London, Ltd., London, 2011. MR 2723325

[GGS02] Wee Teck Gan, Benedict Gross, and Gordan Savin, *Fourier coefficients of modular forms on G_2*, Duke Math. J. **115** (2002), no. 1, 105–169. MR 1932327

[H16] Piper H, *The equidistribution of lattice shapes of rings of integers of cubic, quartic, and quintic number fields: an artist's rendering*, Ph.D. thesis, Princeton University, 2016, p. 130.

[Har17] Robert Harron, *The shapes of pure cubic fields*, Proc. Amer. Math. Soc. **145** (2017), no. 2, 509–524. MR 3577857

[HH19a] Piper H and Robert Harron, *The shapes of galois quartic fields*, 2019, preprint.

[HH19b] Robert Harron and Erik Holmes, *Shapes of sextic fields and log-terms in Malle's connjecture*, 2019, in progress.

[Klü05] Jürgen Klüners, *A counterexample to Malle's conjecture on the asymptotics of discriminants*, C. R. Math. Acad. Sci. Paris **340** (2005), no. 6, 411–414. MR 2135320

[Lev14] Friedrich Levi, *Kubische Zahlkörper und binäre kubische Formenklassen.*, Berichte über die Verhandlungen der Kniglich Schsischen Gesellschaft der Wissenschaften zu Leipzig. Mathematisch-Physische Klasse **66** (1914), no. I, 26–37.

[MS15] Guillermo Mantilla-Soler, *On the arithmetic determination of the trace*, J. Algebra **444** (2015), 272–283. MR 3406177

[MSRG19] Guillermo Mantilla-Soler and Carlos Rivera-Guaca, *An introduction to Casimir pairings and some arithmetic applications*, 2019, available at arXiv:1812.03133v3 [math.NT].

[Neu99] Jürgen Neukirch, *Algebraic number theory*, Grundlehren der Mathematischen Wissenschaften, vol. 322, Springer-Verlag, Berlin, 1999, Translated from the 1992 German original by Norbert Schappacher. MR 1697859 (2000m:11104)

[Sar07] Peter Sarnak, *Reciprocal geodesics*, Analytic number theory, Clay Math. Proc., vol. 7, Amer. Math. Soc., Providence, RI, 2007, pp. 217–237. MR 2362203

[Ser68] Jean-Pierre Serre, *Corps locaux*, Hermann, Paris, 1968, Deuxième édition, Publications de l'Université de Nancago, No. VIII. MR 0354618

[Sie67] C. L. Siegel, *Lectures on quadratic forms*, Notes by K. G. Ramanathan. Tata Institute of Fundamental Research Lectures on Mathematics, No. 7, Tata Institute of Fundamental Research, Bombay, 1967. MR 0271028

[Tau68] Olga Taussky, *The discriminant matrices of an algebraic number field*, J. London Math. Soc. **43** (1968), 152–154. MR 0228473

[Ter97] David C. Terr, *The distribution of shapes of cubic orders*, Ph.D. thesis, University of California, Berkeley, 1997, p. 137. MR 2697241

Department of Mathematics, Keller Hall, University of Hawai'i at Mānoa, Honolulu, HI 96822, USA
E-mail address: rharron@math.hawaii.edu