

NUMBER THEORETICAL PROPERTIES OF ROMIK'S DYNAMICAL SYSTEM

BYUNGCHUL CHA AND DONG HAN KIM

ABSTRACT. We study a dynamical system that was originally defined by Romik in 2008 using an old theorem of Berggren concerning Pythagorean triples. Romik's system is closely related to the Farey map on the unit interval which generates an additive continued fraction algorithm. We explore some number theoretical properties of the Romik system. In particular, we prove an analogue of Lagrange's theorem in the case of the Romik system on the unit quarter circle, which states that a point possesses an eventually periodic digit expansion if and only if the point is defined over a real quadratic extension field of rationals.

1. Introduction

One of the oldest and most classical theorems in number theory is perhaps the infinitude of *primitive Pythagorean triples*, that is, positive integer triples (x, y, z) without common factor satisfying $x^2 + y^2 = z^2$. Less known is the fact that the set of all such triples can be equipped with a certain tree-like structure. To explain, let

$$(1.1) \quad M_1 = \begin{pmatrix} -1 & 2 & 2 \\ -2 & 1 & 2 \\ -2 & 2 & 3 \end{pmatrix}, \quad M_2 = \begin{pmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 3 \end{pmatrix}, \quad M_3 = \begin{pmatrix} 1 & -2 & 2 \\ 2 & -1 & 2 \\ 2 & -2 & 3 \end{pmatrix}.$$

Then the repeated actions of M_1, M_2, M_3 via left-multiplication on (the column vectors) $(3, 4, 5)$ and $(4, 3, 5)$ generate all primitive Pythagorean triples and each primitive Pythagorean triple shows up in the trees exactly once, as pictured in Figure 1. As far as we are aware, the oldest literature containing this theorem is the paper [3] by Berggren. We will refer the trees in Figure 1 as the *Berggren trees*. See [1] and [2] for the proofs of this theorem and related discussion.

Received February 10, 2019; Accepted May 30, 2019.

2010 *Mathematics Subject Classification.* Primary 11J70; Secondary 11A55.

Key words and phrases. Pythagorean triple, continued fraction, Berggren theorem, Romik system.

Research supported by the National Research Foundation of Korea (NRF-2018R1A2B6001624).

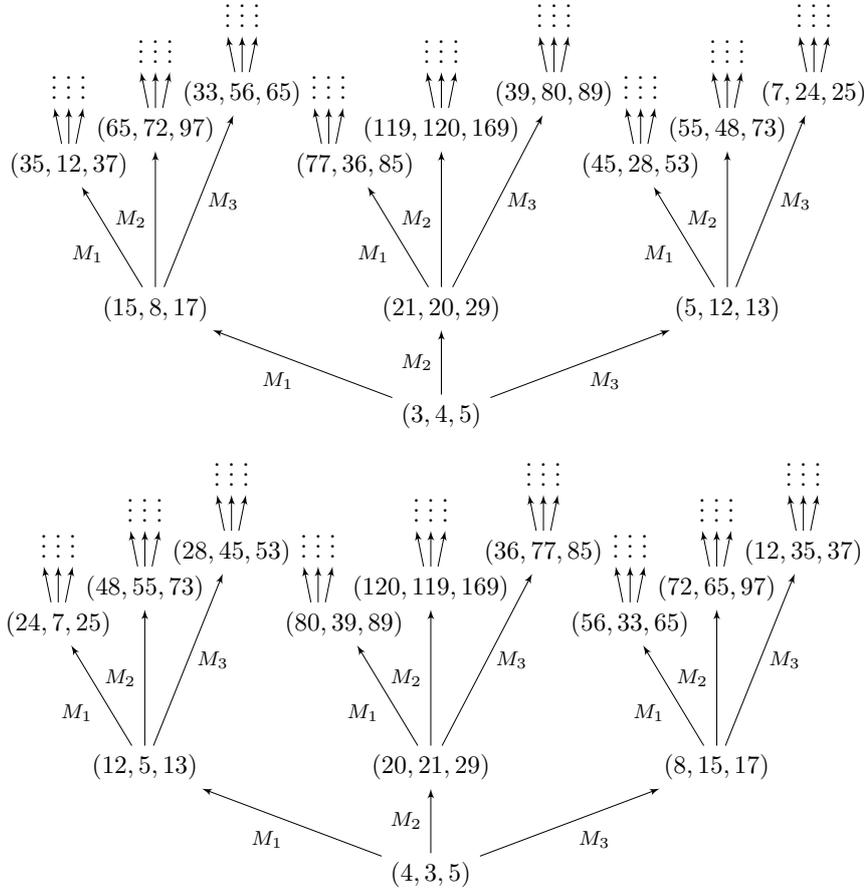


FIGURE 1. Berggren trees for Pythagorean triples

Also, the paper [14] by Romik contains an extensive list of bibliography on Berggren’s theorem.

In the same paper, Romik initiates the investigation of a dynamical system $T : \mathcal{Q} \rightarrow \mathcal{Q}$, where \mathcal{Q} is the (closed) unit quarter circle

$$\mathcal{Q} = \{(x, y) \in \mathbb{R}^2 \mid x \geq 0, y \geq 0 \text{ and } x^2 + y^2 = 1\},$$

and $T(x, y)$ is defined to be

$$T(x, y) = \left(\frac{|2 - x - 2y|}{3 - 2x - 2y}, \frac{|2 - 2x - y|}{3 - 2x - 2y} \right).$$

This dynamical system naturally arises from the Berggren trees in Figure 1. In fact, T is the unique continuous map from \mathcal{Q} to itself, which is “parent-finding”

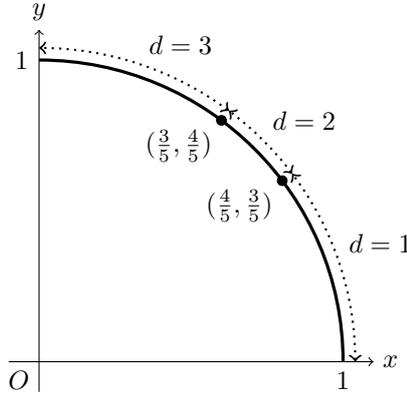


FIGURE 2. Digits of the points on \mathcal{Q}

in the following sense. If $(\frac{a}{c}, \frac{b}{c})$ is a rational point on \mathcal{Q} represented by a primitive Pythagorean triple (a, b, c) , then $T(\frac{a}{c}, \frac{b}{c}) = (\frac{a'}{c'}, \frac{b'}{c'})$, finding the *parent* Pythagorean triple (a', b', c') of (a, b, c) in the Berggren trees. Additionally, Romik assigns a digit $d = d(x, y) \in \{1, 2, 3\}$ for each $(x, y) \in \mathcal{Q}$ according to

$$(1.2) \quad d(x, y) = \begin{cases} 1 & \text{if } \frac{4}{5} < x \leq 1, \\ 2 & \text{if } \frac{3}{5} < x < \frac{4}{5}, \\ 3 & \text{if } 0 \leq x < \frac{3}{5}, \end{cases}$$

as in Figure 2. Then the digit expansion of (x, y) is defined to be

$$(1.3) \quad (x, y) = [d_1, d_2, \dots]_{\mathcal{Q}},$$

where $d_j = d(T^{j-1}(x, y))$ for $j = 1, 2, \dots$. We call this the *Romik digit expansion* of (x, y) . For the “boundary cases” $x = \frac{3}{5}$ and $x = \frac{4}{5}$, we allow both adjacent digits to be valid, instead of using “oe” and “eo” as in [14]. Namely,

$$d(\frac{4}{5}, \frac{3}{5}) = 1 \text{ or } 2, \quad d(\frac{3}{5}, \frac{4}{5}) = 2 \text{ or } 3.$$

Note that $T(\frac{3}{5}, \frac{4}{5}) = (1, 0)$ and $T(\frac{4}{5}, \frac{3}{5}) = (0, 1)$ are the only fixed points of $T(x, y)$, and their Romik digit expansions are

$$(1, 0) = [1^\infty]_{\mathcal{Q}} := [1, 1, \dots]_{\mathcal{Q}}, \text{ and } (0, 1) = [3^\infty]_{\mathcal{Q}} := [3, 3, \dots]_{\mathcal{Q}}.$$

As a consequence, Theorem 2 in [14] says that a point $(x, y) \in \mathcal{Q}$ is rational if and only if its Romik digit expansion ends with either 1^∞ or 3^∞ . Moreover, except for $(1, 0)$ and $(0, 1)$, every rational point (x, y) has *two* endings in its Romik digit expansions

$$(x, y) = [d_1, d_2, \dots, 2, 1^\infty]_{\mathcal{Q}} \text{ and } [d_1, d_2, \dots, 3, 1^\infty]_{\mathcal{Q}},$$

or

$$(x, y) = [d_1, d_2, \dots, 2, 3^\infty]_{\mathcal{Q}} \text{ and } [d_1, d_2, \dots, 1, 3^\infty]_{\mathcal{Q}}.$$

The Romik digit expansion of a point on \mathcal{Q} resembles classical (simple) continued fraction expansions of real numbers. In fact, as Romik explains in [14], the dynamical system (\mathcal{Q}, T) is conjugate to a dynamical system on the unit interval $[0, 1]$ constructed from, what he calls, a modified Euclidean algorithm.

Does (\mathcal{Q}, T) exhibit any number theoretical property shared by the dynamical system associated with the classical continued fraction? For example, Romik asks in [14] if an analogue of Lagrange's theorem for (\mathcal{Q}, T) is true. That is, is it true that a point $(x, y) \in \mathcal{Q}$ has an eventually periodic Romik digit expansion if and only if (x, y) is defined over a (real) quadratic extension of \mathbb{Q} ? The main result in our paper answers this question affirmatively. Our proof leverages the structure of a *quadratic space* $(\mathbb{R}^3, Q(\mathbf{x}))$ where $Q(\mathbf{x}) = x_1^2 + x_2^2 - x_3^2$ is the Pythagorean quadratic form. In this sense, the present paper is a natural continuation of the work [4] by Cha, Nguyen and Tauber.

Perhaps not surprisingly, our proof of the Lagrange theorem is reminiscent of Lagrange's original proof of his theorem for the usual continued fraction expansion. The heart of Lagrange's classical proof is to show that the discriminants of defining equations of the irrationals with the same tails in their continued fraction expansions lie in a bounded subset of \mathbb{R} . In our case, a key step is to show that the *Q-cross products* (see §3.1) arising from the vectors in the same Romik T -orbit lie in a bounded subset of \mathbb{R}^3 .

We point out here that one can deduce Lagrange's theorem for the Romik system as a corollary from another general theorem, namely, Panti's theorem in [11]. Panti proves a version of Lagrange's theorem which is applicable to many systems arising from unimodular partitions of the unit interval. The aforementioned dynamical system on the unit interval to which (\mathcal{Q}, T) is conjugate is an example of such a system. Thus, Panti's theorem in [11] can be applied in this context to prove the Lagrange theorem for (\mathcal{Q}, T) . After a version of this paper had been completed, we recently learned from [12] that Panti proved in the Lagrange and Galois theorems for, what he calls, all billiard maps based on unimodular partitions.

Even though Panti's theorem in [11] gives a simpler proof for the Lagrange theorem for (\mathcal{Q}, T) it seems to us that the tools we develop in the present paper are well-suited to the study of other number theoretical properties regarding (\mathcal{Q}, T) . In particular, our set-up for the Romik system can provide flexible tools in the study of *intrinsic Diophantine approximation*, following Kleinbock, Merrill, and their collaborators in [10] and [8]. The authors intend to pursue this in a follow-up work in the near future.

The rest of this paper is organized as follows. In §2, we set up notations and recall some background materials on quadratic spaces, as well as prior results mainly from [6] and [4]. We review the geometric construction from [4] and provide a self-contained proof of Berggren's theorem, which is essentially due to Conrad [6]. There are already many proofs available in the literature for

Berggren's theorem. However, our presentation provides a convenient starting point for our discussion later. The main result of the paper is contained in §3, where we complete the proofs of Lagrange's and Galois' theorems for the Romik system. Some open questions and potential future developments are described in §4.

2. Background materials and review on the Romik system

2.1. Notational convention and quadratic spaces

We will use bold-faced letters, such as \mathbf{v}, \mathbf{w} , etc., to denote vectors in \mathbb{R}^3 . They are regarded as column vectors, so that a 3×3 matrix acts on them via left-multiplication. Points on \mathbb{R}^2 are usually denoted by capital letters, such as P . We write x_1, x_2, x_3 for the standard coordinate functions in \mathbb{R}^3 and we use x and y for \mathbb{R}^2 . Define the unit circle

$$\mathcal{U} = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$$

and the quarter circle

$$\mathcal{Q} = \{(x, y) \in \mathbb{R}^2 \mid x, y \geq 0 \text{ and } x^2 + y^2 = 1\}.$$

Also, we define a projection map π to be

$$(2.1) \quad \pi : \mathbb{R}^3 - \{x_3 = 0\} \longrightarrow \mathbb{R}^2, \quad (x_1, x_2, x_3) \mapsto (x, y) := (x_1/x_3, x_2/x_3).$$

We will say that \mathbf{v} represents P when $\pi(\mathbf{v}) = P$.

Let

$$Q(\mathbf{x}) = x_1^2 + x_2^2 - x_3^2,$$

which we will refer as the Pythagorean quadratic form on \mathbb{R}^3 . However, many of our results here can be applied with some minimal modifications to other quadratic forms $Q(\mathbf{x})$ satisfying certain technical conditions, the conditions (Q-I)—(Q-III) in [4], to be precise. With such other cases in mind, we will try to present our arguments in the form that can be generalized easily in the future. Recall that a *quadratic space* $(\mathbb{R}^3, Q(\mathbf{x}))$ is a vector space \mathbb{R}^3 equipped with a quadratic form $Q(\mathbf{x})$ on \mathbb{R}^3 . It is well-known that the set of all quadratic forms $Q(\mathbf{x})$ is in one-to-one correspondence with the set of all symmetric bilinear forms on \mathbb{R}^3 , as well as that of all 3×3 symmetric matrices with real coefficients. In general, the relationship among an arbitrary $Q(\mathbf{x})$, the corresponding bilinear form, and M_Q is given by

$$(2.2) \quad \langle \mathbf{x}, \mathbf{y} \rangle = \frac{1}{2}(Q(\mathbf{x} + \mathbf{y}) - Q(\mathbf{x}) - Q(\mathbf{y})) = \mathbf{x}^T M_Q \mathbf{y}$$

for all \mathbf{x}, \mathbf{y} in \mathbb{R}^3 .

In our case of the Pythagorean form, the corresponding bilinear form is

$$\langle \mathbf{x}, \mathbf{y} \rangle = x_1y_1 + x_2y_2 - x_3y_3,$$

where $\mathbf{x} = (x_1, x_2, x_3)$ and $\mathbf{y} = (y_1, y_2, y_3)$, and

$$M_Q = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Define

$$C_Q = \{\mathbf{x} \in \mathbb{R}^3 \mid Q(\mathbf{x}) = 0\},$$

to be the set of all Q -null vectors in \mathbb{R}^3 . Also, if a linear map A on \mathbb{R}^3 to itself preserves $Q(\mathbf{x})$, that is, $Q(A\mathbf{x}) = Q(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^3$, then we say that A is *orthogonal with respect to $Q(\mathbf{x})$* . The group of all orthogonal maps is denoted by $O_Q(\mathbb{R})$. Clearly, every $A \in O_Q(\mathbb{R})$ maps C_Q to itself.

2.2. Reflections and their actions on \mathcal{U} and C_Q

Fix $\mathbf{z} \in \mathbb{R}^3$ with $Q(\mathbf{z}) \neq 0$. The *reflection* $s_{\mathbf{z}} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ of \mathbf{z} is defined to be

$$(2.3) \quad s_{\mathbf{z}}(\mathbf{x}) = \mathbf{x} - 2 \frac{\langle \mathbf{x}, \mathbf{z} \rangle}{Q(\mathbf{z})} \mathbf{z}.$$

It is easy to prove that $s_{\mathbf{z}} \in O_Q(\mathbb{R})$. Also, $s_{\mathbf{z}}^2 = \mathbf{1}_{\mathbb{R}^3}$, the identity map on \mathbb{R}^3 , so that $s_{\mathbf{z}}$ is its own inverse. In addition, $\det(s_{\mathbf{z}}) = -1$ because \mathbf{z} is an eigenvector of $s_{\mathbf{z}}$ with the eigenvalue -1 , while

$$\{\mathbf{x} \in \mathbb{R}^3 \mid \langle \mathbf{x}, \mathbf{z} \rangle = 0\}$$

is a two-dimensional eigenspace with the eigenvalue 1.

Define U_1, U_2, U_3 by

$$(2.4) \quad U_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad U_2 = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad U_3 = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Then it is easy to show that $U_1 = s_{(0,1,0)}$ and $U_3 = s_{(1,0,0)}$ (with the matrices being understood via the standard basis of \mathbb{R}^3). Also, $U_2 = U_1 U_3$ is the composite of the two reflections $s_{(0,1,0)}$ and $s_{(1,0,0)}$. Next, let

$$(2.5) \quad H = s_{(1,1,1)} = \begin{pmatrix} -1 & -2 & 2 \\ -2 & -1 & 2 \\ -2 & -2 & 3 \end{pmatrix}.$$

Then, the matrices M_1, M_2, M_3 in (1.1) are

$$(2.6) \quad M_1 = H U_1, \quad M_2 = H U_2, \quad \text{and} \quad M_3 = H U_3,$$

as first noted by Conrad in [6] and by Berggren [3] in a more indirect way. Equivalently,

$$(2.7) \quad M_1^{-1} = U_1 H, \quad M_2^{-1} = U_2 H \quad \text{and} \quad M_3^{-1} = U_3 H.$$

Central to us is to describe geometrically the actions of U_1, U_2, U_3, H on C_Q , as well as their induced actions on \mathcal{U} , whose meaning we now make precise below. Suppose that M is any one of U_1, U_2, U_3 , and H , or a finite product of

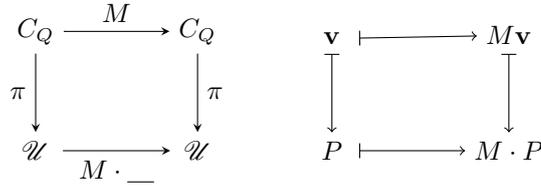


FIGURE 3. Action of M on C_Q and \mathcal{U}

them. If P is in \mathcal{U} , then we denote by $M \cdot P$ the point represented by $M\mathbf{v}$ when \mathbf{v} is any vector representing P . This is summarized in Figure 3.

Apply this for $M = U_j$. Then we obtain from (2.4) that

$$\begin{aligned} U_1 \cdot (x, y) &= (x, -y), \\ U_2 \cdot (x, y) &= (-x, -y), \\ U_3 \cdot (x, y) &= (-x, y), \end{aligned}$$

for any $(x, y) \in \mathcal{U}$. As for $M = H$, we compute $H\mathbf{v}$ using (2.5) for a few \mathbf{v} 's. The results are summarized in the table below.

\mathbf{v}	$H\mathbf{v}$
$(0, 1, 1)$	$(0, 1, 1)$
$(3, 4, 5)$	$(-1, 0, 1)$
$(4, 3, 5)$	$(0, -1, 1)$
$(1, 0, 1)$	$(1, 0, 1)$

Remembering that H is a reflection, thus of order 2, we see that H sends the subarc of \mathcal{Q} between $(1, 0)$ and $(\frac{4}{5}, \frac{3}{5})$ onto the quarter circle in the fourth quadrant, and vice versa. Recall from the definition of $d(x, y)$ in (1.2) and in Figure 2 that this arc is precisely the set of (x, y) whose Romik digit $d(x, y) = 1$. Likewise, H moves back and forth the subarc of \mathcal{Q} with $d(x, y) = 2$ onto the quarter circle in the third quadrant. Similarly, the subarc with $d(x, y) = 3$ is mapped by H onto the quarter circle in the second quadrant. This gives a geometric description of the actions of H and U_j 's on \mathcal{U} , as is pictured in Figure 4.

We now consider the actions of H and U_j 's on C_Q . As we understand their actions on \mathcal{U} it is enough to know the x_3 -coordinates of $H\mathbf{v}$ and $U_j\mathbf{v}$. First of all, U_j obviously preserves the x_3 -coordinates. As for H , we use the following proposition.

Proposition 2.1. *Let $\mathbf{v} = (a, b, c)$ (not necessarily in C_Q) with $c \neq 0$ and write $H\mathbf{v} = (a', b', c')$. Then*

$$(2.8) \quad 1 < \frac{a}{c} + \frac{b}{c} < 2,$$

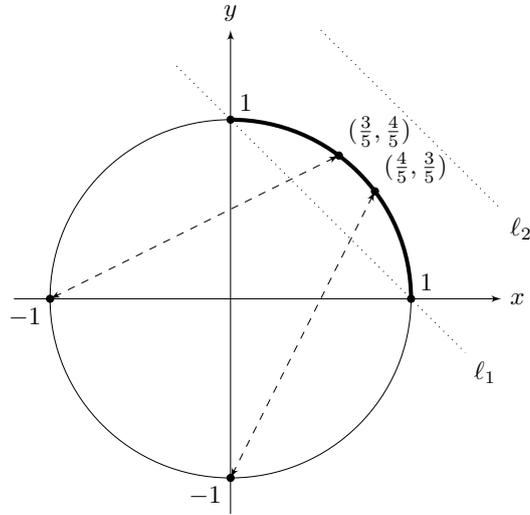


FIGURE 4. Action of H on \mathcal{U}

if and only if $|c'| < |c|$.

Proof. A straightforward calculation shows $c' = -2a - 2b + 3c$. The proof then follows immediately from the identity

$$\begin{aligned} c'^2 - c^2 &= (-2a - 2b + 3c)^2 - c^2 \\ &= 4c^2 \left(\frac{a}{c} + \frac{b}{c} - 1 \right) \left(\frac{a}{c} + \frac{b}{c} - 2 \right). \end{aligned} \quad \square$$

Note that the condition (2.8) is equivalent to saying that \mathbf{v} represents a point (x, y) in \mathbb{R}^2 between the two lines

$$\ell_1 : x + y = 1 \quad \text{and} \quad \ell_2 : x + y = 2,$$

as depicted in Figure 4.

Now, suppose $\mathbf{v} = (a, b, c) \in C_Q$ and write $H\mathbf{v} = (a', b', c')$. If \mathbf{v} represents a point P in (the interior of) \mathcal{Q} , then \mathbf{v} satisfies the condition (2.8), therefore, $|c'| < |c|$. In other words, H brings \mathbf{v} closer to the origin if $P \in \mathcal{Q}$. Likewise, while H moves \mathbf{v} away from the origin if $P \in \mathcal{U} - \mathcal{Q}$.

As a final note, we claim here that, if $c > 0$ for any $\mathbf{v} = (a, b, c) \in C_Q$, then $c' > 0$ as well. To prove this, define

$$C_Q^+ = \{(a, b, c) \in C_Q \mid c > 0\}$$

and

$$C_Q^- = \{(a, b, c) \in C_Q \mid c < 0\}.$$

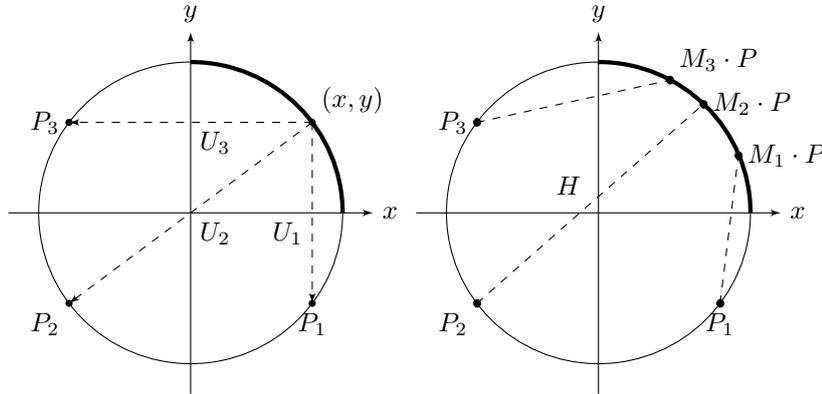


FIGURE 5. Actions of M_1, M_2, M_3 on $(x, y) \in \mathcal{Q}$. First, (x, y) moves to P_1, P_2, P_3 in the 4th, 3rd, 2nd quadrants under U_1, U_2, U_3 , respectively. These points are then brought back to \mathcal{Q} under H .

Note that C_Q^+ and C_Q^- are disjoint and connected, and that H is a continuous invertible map leaving stable the set

$$C_Q - \{0\} = C_Q^+ \cup C_Q^-.$$

So, H must leave stable each of C_Q^+ and C_Q^- separately, or H must swap them. But, the second is impossible because H fixes $(1, 0, 1) \in C_Q^+$. So, H must leave stable each of C_Q^\pm separately.

2.3. Proof of Berggren's theorem

Equipped with this geometric insight, we are now ready to prove Berggren's theorem. First, define a directed graph as follows:

- Vertex:** the vertices are all primitive Pythagorean triples $\mathbf{v} = (a, b, c)$, and
- Edge:** there is an edge from $\mathbf{v} = (a, b, c)$ (which we regard as a *parent*) to $\mathbf{v}' = (a', b', c')$ (a *child*) whenever $\mathbf{v}' = M_j \mathbf{v}$ for $j = 1, 2, \text{ or } 3$.

We show first that each vertex \mathbf{v} has outdegree 3, in other words, all

$$M_j \mathbf{v} = H U_j \mathbf{v}$$

for $j = 1, 2, 3$ are primitive Pythagorean triples. First, note that $U_j \mathbf{v} \in C_Q^+$ and H leaves C_Q^+ stable. So, $M_j \mathbf{v} \in C_Q^+$. We leave as an exercise for the reader to verify that, if $A \in O_Q(\mathbb{R})$ is *integral*, namely, if $A\mathbf{w} \in \mathbb{Z}^3$ for any $\mathbf{w} = (w_1, w_2, w_3) \in \mathbb{Z}^3$ and if $\gcd(w_1, w_2, w_3) = 1$, then the gcd of the three coordinates of $A\mathbf{w}$ is one, as well.

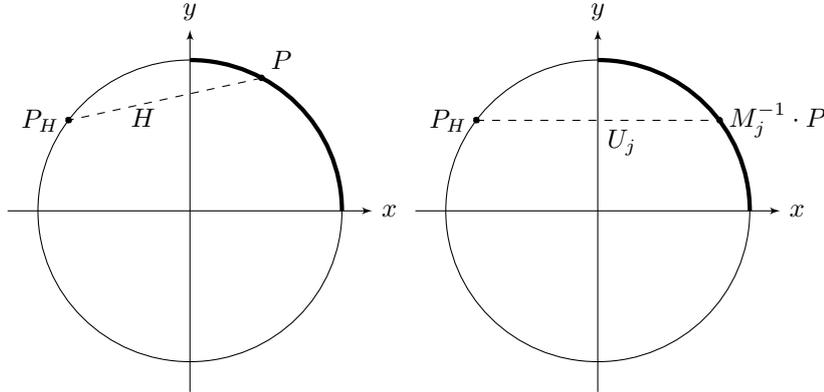


FIGURE 6. Action of M_j^{-1} on $P \in \mathcal{Q}$. First, H moves P outside of \mathcal{Q} to P_H . Then exactly one of U_1, U_2, U_3 , namely, U_j with $j = d(x, y)$ brings it back to \mathcal{Q} .

Now, it remains to show $HU_j\mathbf{v}$ represents a point in \mathcal{Q} . Clearly, the vector $U_j\mathbf{v}$ represents a point, say, P_j , which is not in \mathcal{Q} . Then H moves P_j back to one of the three subarcs of \mathcal{Q} (corresponding to the three Romik digits $d = 1, 2, 3$), as described in as Figure 5. This finishes proving the claim that \mathbf{v} has outdegree 3.

Next, suppose \mathbf{v} is a vertex not equal to $(3, 4, 5)$ and $(4, 3, 5)$. We argue that $\mathbf{v} = (a, b, c)$ has indegree 1, or, equivalently,

$$M_j^{-1}\mathbf{v} = U_jH\mathbf{v}$$

is a primitive Pythagorean triple for exactly one value of $j = 1, 2, 3$. In fact, if we let $j = d(x, y)$, the Romik digit of $P = (x, y)$ with $x = \frac{a}{c}, y = \frac{b}{c}$ (namely, P is the point represented by \mathbf{v}), then the point P_H represented by $H\mathbf{v}$ is on the second, third or fourth quadrant, depending on $j = 1, 2, 3$, respectively. Then, U_j moves P_H back to \mathcal{Q} , as in Figure 6. For the exceptional case $\mathbf{v} = (3, 4, 5)$ and $(4, 3, 5)$, we have $U_jH\mathbf{v} = (1, 0, 1)$ and $(0, 1, 1)$, which are not (technically) primitive Pythagorean triples. So, $(3, 4, 5)$ and $(4, 3, 5)$ are the only vertices of indegree 0.

Lastly, we claim that the actions of M_1, M_2, M_3 on a vertex $\mathbf{v} = (a, b, c)$ increase its x_3 -coordinate, while the actions of M_j^{-1} decrease it. For any j , let $(a', b', c') = M_j\mathbf{v} = U_jH\mathbf{v}$. Since \mathbf{v} represents a point in \mathcal{Q} we see that $H\mathbf{v}$ has greater x_3 -coordinate than \mathbf{v} , as was observed before. Also, U_j always preserves the x_3 -coordinates. This proves $c' > c$. Additionally, we note $\mathbf{v} = M_j^{-1}(a', b', c')$ (when j is the Romik digit of the point represented by (a', b', c')). This proves the claim.

To summarize, we proved that each primitive Pythagorean triple (a, b, c) (except for $(3, 4, 5)$ and $(4, 3, 5)$) satisfies the three properties

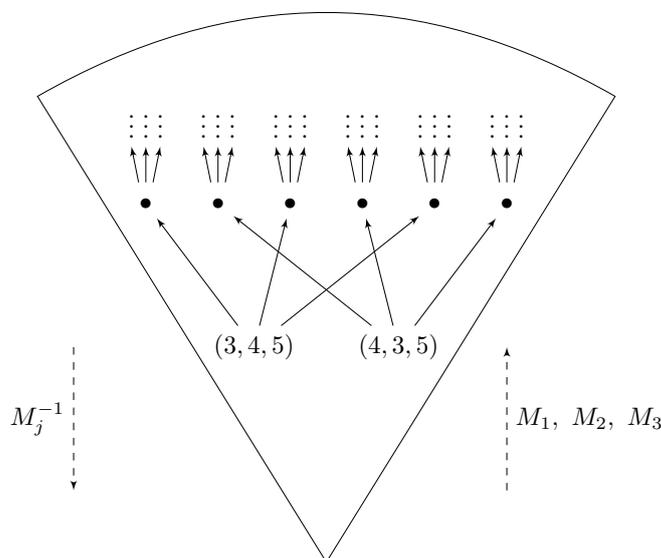


FIGURE 7. The vertices are drawn on C_Q^+ . Except for $(3, 4, 5)$ and $(4, 3, 5)$, each vertex is of outdegree 3 and of indegree 0. The actions of M_1, M_2, M_3 move the vertices upwards, increasing the x_3 -coordinates, while M_j^{-1} move them downwards, decreasing the x_3 -coordinates.

- (a, b, c) has outdegree 3,
- (a, b, c) has indegree 1,
- any vertex coming out of (a, b, c) has its x_3 -coordinate greater than c , while the vertex going into (a, b, c) has a smaller x_3 -coordinate than c .

We will say that a vertex (a, b, c) is a *funnel* vertex, if it satisfies the above three properties. The idea behind this terminology is this. Suppose $P \in \mathcal{Q}$. If we trace the vertices \mathbf{v} on C_Q^+ representing the points $P, T(P), T^2(P), \dots$, then we must move from a vertex to another along the *opposite* direction in the tree. So, at any vertex (except for $(3, 4, 5)$ and $(4, 3, 5)$), a T -orbit always moves *towards* the origin.

What we proved so far is summarized as:

Theorem 2.2. *All primitive Pythagorean triples (a, b, c) are funnel vertices, except for $(3, 4, 5)$ and $(4, 3, 5)$.*

Now, Berggren's theorem is an easy corollary to this, because any given primitive Pythagorean triple (a, b, c) , under the successive actions of M_j^{-1} , will funnel down until it reaches $(3, 4, 5)$ or $(4, 3, 5)$ and it must do so in a unique way.

3. Proof of Lagrange's Theorem

The goal of this section is to prove:

Theorem 3.1 (Lagrange's Theorem for the Romik system). *Let $P = (\alpha, \beta) \in \mathcal{Q}$. Its Romik digit expansion*

$$(\alpha, \beta) = [d_1, d_2, \dots,]_{\mathcal{Q}}$$

is eventually periodic if and only if P is defined over a real quadratic field, namely, there exists a squarefree positive integer D such that $\alpha, \beta \in \mathbb{Q}(\sqrt{D})$.

3.1. Q -cross product

Let $\mathbf{v}_1, \mathbf{v}_2$ be two vectors in \mathbb{R}^3 . We define the Q -cross product $\mathbf{v}_3 = \mathbf{v}_1 \times_Q \mathbf{v}_2$ by

$$(3.1) \quad \mathbf{v}_3 = M_Q^{-1}(\mathbf{v}_1 \times \mathbf{v}_2),$$

where $\mathbf{v}_1 \times \mathbf{v}_2$ is the usual cross product in \mathbb{R}^3 and M_Q is the symmetric matrix associated with the quadratic form $Q(\mathbf{x})$. For the case of Pythagorean form $Q(\mathbf{x}) = x_1^2 + x_2^2 - x_3^2$, which is all we need in this paper,

$$(3.2) \quad \mathbf{v}_1 \times_Q \mathbf{v}_2 = (b_1c_2 - b_2c_1, a_2c_1 - a_1c_2, a_2b_1 - a_1b_2),$$

if we write $\mathbf{v}_1 = (a_1, b_1, c_1)$ and $\mathbf{v}_2 = (a_2, b_2, c_2)$.

We deduce from (2.2) and (3.1) that $\mathbf{v}_3 = \mathbf{v}_1 \times_Q \mathbf{v}_2$ satisfies

- (A) $\langle \mathbf{v}_1, \mathbf{v}_3 \rangle = \langle \mathbf{v}_2, \mathbf{v}_3 \rangle = 0$, and
- (B) $Q(\mathbf{v}_3) = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3]$, where $[\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3] := (\mathbf{v}_1 \times \mathbf{v}_2) \cdot \mathbf{v}_3$ is by definition the triple product of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$.

For example,

$$\langle \mathbf{v}_1, \mathbf{v}_3 \rangle = \mathbf{v}_1^T M_Q M_Q^{-1}(\mathbf{v}_1 \times \mathbf{v}_2) = 0$$

and likewise $\langle \mathbf{v}_2, \mathbf{v}_3 \rangle = 0$. Also,

$$Q(\mathbf{v}_3) = \langle \mathbf{v}_3, \mathbf{v}_3 \rangle = \mathbf{v}_3^T M_Q M_Q^{-1}(\mathbf{v}_1 \times \mathbf{v}_2) = \mathbf{v}_3 \cdot (\mathbf{v}_1 \times \mathbf{v}_2) = (\mathbf{v}_1 \times \mathbf{v}_2) \cdot \mathbf{v}_3.$$

Conversely, if \mathbf{v}_1 and \mathbf{v}_2 are linearly independent, then the properties (A) and (B) characterizes $\mathbf{v}_1 \times_Q \mathbf{v}_2$ uniquely. First, the vectors satisfying (A) form a one-dimensional subspace of \mathbb{R}^3 , from which we choose a (nonzero) vector \mathbf{u} . Then, we take $\mathbf{v}_3 = k\mathbf{u}$ with $k = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{u}]/Q(\mathbf{u})$. It is easily verified that \mathbf{v}_3 satisfies (B).

Even though we do not need it in this paper, we present a more functorial construction of $\mathbf{v}_1 \times_Q \mathbf{v}_2$ for future reference. What we discuss in this paragraph is standard. See, for example, [7]. Let V be a n -dimensional vector space (over \mathbb{R}), equipped with a nondegenerate bilinear pairing $\langle \cdot, \cdot \rangle$. Then for each $p = 1, \dots, n$, the bilinear pairing $\langle \cdot, \cdot \rangle_p$ of $\bigwedge^p V$, which is defined by

$$\langle \mathbf{v}_1 \wedge \dots \wedge \mathbf{v}_p, \mathbf{w}_1 \wedge \dots \wedge \mathbf{w}_p \rangle_p = \det(\langle \mathbf{v}_i, \mathbf{w}_j \rangle)_{i,j=1,\dots,p},$$

is also nondegenerate. Fix a basis of $\bigwedge^n V$ (that is, an orientation or a volume form), say, $\tau \in \bigwedge^n V$. Then, for any $\lambda \in \bigwedge^p V$, there exists a unique element of $\bigwedge^{n-p} V$, which we denote by $\star\lambda$, satisfying

$$\lambda \wedge \mu = \langle \star\lambda, \mu \rangle_{n-p} \tau$$

for all $\mu \in \bigwedge^{n-p} V$. This way, we define *the Hodge star operator*

$$\bigwedge^p V \longrightarrow \bigwedge^{n-p} V, \quad \lambda \mapsto \star\lambda.$$

The relevance of this in our context is this. Let $V = \mathbb{R}^3$, equipped with the pairing $\langle \cdot, \cdot \rangle$ given by the Pythagorean form $Q(\mathbf{x})$ and choose the standard orientation $\tau = \mathbf{e}_1 \wedge \mathbf{e}_2 \wedge \mathbf{e}_3$. Then one can show as an exercise that

$$(3.3) \quad \mathbf{v}_1 \times_Q \mathbf{v}_2 = \star(\mathbf{v}_1 \wedge \mathbf{v}_2),$$

because $\star(\mathbf{v}_1 \wedge \mathbf{v}_2)$ satisfies the properties (A) and (B) above.

Proposition 3.2. *For any $A \in O_Q(\mathbb{R}^3)$ and $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^3$,*

$$A\mathbf{v}_1 \times_Q A\mathbf{v}_2 = (\det A)A(\mathbf{v}_1 \times_Q \mathbf{v}_2).$$

Proof. First of all, if \mathbf{v}_1 and \mathbf{v}_2 are linearly dependent, then both sides of the above equation are zero. So, we assume that \mathbf{v}_1 and \mathbf{v}_2 are linearly independent and we prove that

$$\mathbf{v} := (\det A)A(\mathbf{v}_1 \times_Q \mathbf{v}_2)$$

satisfies the properties (A) and (B) above with respect to $A\mathbf{v}_1$ and $A\mathbf{v}_2$. For (A),

$$\langle A\mathbf{v}_1, \mathbf{v} \rangle = (\det A)\langle A\mathbf{v}_1, A(\mathbf{v}_1 \times_Q \mathbf{v}_2) \rangle = (\det A)\langle \mathbf{v}_1, \mathbf{v}_1 \times_Q \mathbf{v}_2 \rangle = 0.$$

Likewise, $\langle A\mathbf{v}_2, \mathbf{v} \rangle = 0$. For (B), we note in general that

$$[A\mathbf{w}_1, A\mathbf{w}_2, A\mathbf{w}_3] = (\det A)[\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3]$$

for any $A \in O_Q(\mathbb{R})$ and $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3 \in \mathbb{R}^3$. So,

$$\begin{aligned} [A\mathbf{v}_1, A\mathbf{v}_2, \mathbf{v}] &= [A\mathbf{v}_1, A\mathbf{v}_2, (\det A)A(\mathbf{v}_1 \times_Q \mathbf{v}_2)] = (\det A)^2[\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_1 \times_Q \mathbf{v}_2] \\ &= (\det A)^2 Q(\mathbf{v}_1 \times_Q \mathbf{v}_2) = Q(\mathbf{v}), \end{aligned}$$

where the last equality is because A is orthogonal with respect to $Q(\mathbf{x})$. This proves (B). \square

Proposition 3.3. *Let $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^3$ and $\mathbf{v}_3 = \mathbf{v}_1 \times_Q \mathbf{v}_2$. Assume $\mathbf{v}_1 \in C_Q$, that is, $Q(\mathbf{v}_1) = 0$. Then*

$$Q(\mathbf{v}_3) = (\langle \mathbf{v}_1, \mathbf{v}_2 \rangle)^2.$$

Proof. It is possible to give a “coordinate-free” proof based on (3.3). Instead, we present an elementary approach beginning with (3.2). As the validity of the equation in this proposition remains unchanged under rescaling of \mathbf{v}_1 and \mathbf{v}_2 , without loss of generality, we may assume

$$\mathbf{v}_1 = (\alpha, \beta, 1), \text{ and } \mathbf{v}_2 = (\alpha', \beta', 1),$$

with $\alpha^2 + \beta^2 = 1$. (Later in the proof, we separately deal with the case when the x_3 -coordinate of \mathbf{v}_2 is zero.) Then (3.2) gives $\mathbf{v}_3 = (\beta - \beta', \alpha' - \alpha, \alpha'\beta - \alpha\beta')$, so

$$\begin{aligned} Q(\mathbf{v}_3) &= (\beta - \beta')^2 + (\alpha' - \alpha)^2 - (\alpha'\beta - \alpha\beta')^2 \\ &= 1 + (\alpha'^2 + \beta'^2) - 2(\alpha\alpha' + \beta\beta') - (\alpha'^2\beta^2 + \alpha^2\beta'^2 - 2\alpha\alpha'\beta\beta') \\ &= 1 + (\alpha'^2 + \beta'^2) - 2(\alpha\alpha' + \beta\beta') - (\alpha'^2(1 - \alpha^2) + (1 - \beta^2)\beta'^2 - 2\alpha\alpha'\beta\beta') \\ &= 1 - 2(\alpha\alpha' + \beta\beta') + (\alpha\alpha' + \beta\beta')^2 \\ &= (\langle \mathbf{v}_1, \mathbf{v}_2 \rangle)^2. \end{aligned}$$

For the case $\mathbf{v}_2 = (\alpha', \beta', 0)$, we have $\mathbf{v}_3 = (-\beta', \alpha', \alpha'\beta - \alpha\beta')$. And, the rest is similar to the above calculation. We omit the detail. \square

Assume further that $\mathbf{v}_1, \mathbf{v}_2 \in C_Q$ and that \mathbf{v}_1 and \mathbf{v}_2 are linearly independent. Then $\mathbf{w} := \mathbf{v}_1 \times_Q \mathbf{v}_2$ has a simple geometric description. The two subspaces

$$\langle \mathbf{v}_1 \rangle^\perp := \{ \mathbf{x} \in \mathbb{R}^3 \mid \langle \mathbf{v}_1, \mathbf{x} \rangle = 0 \}$$

and

$$\langle \mathbf{v}_2 \rangle^\perp := \{ \mathbf{x} \in \mathbb{R}^3 \mid \langle \mathbf{v}_2, \mathbf{x} \rangle = 0 \}$$

are tangent to C_Q and their intersection is the line spanned by \mathbf{w} . Also, we have $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle \neq 0$, otherwise the span of \mathbf{v}_1 and \mathbf{v}_2 would be contained in C_Q , contradicting the nondegeneracy of $Q(\mathbf{x})$. So, Proposition 3.3 shows that $W := Q(\mathbf{w}) > 0$. In particular, \mathbf{w} lies on the *one-sheeted hyperboloid*

$$(3.4) \quad \mathcal{H}(W) : x_1^2 + x_2^2 - x_3^2 = W.$$

Keeping the assumption that $\mathbf{v}_1, \mathbf{v}_2 \in C_Q$, we additionally suppose that \mathbf{v}_1 represents a point in \mathcal{Q} via the projection map π , which is defined in (2.1). In this case, the following proposition says that $\mathbf{v}_1 \times_Q \mathbf{v}_2$ represents a point in \mathbb{R}^2 “close to” \mathcal{Q} .

Proposition 3.4. *Suppose $\mathbf{v}_1, \mathbf{v}_2 \in C_Q$ are linearly independent. Assume that \mathbf{v}_1 represents a point in \mathcal{Q} . Write $\mathbf{w} := (w_1, w_2, w_3) = \mathbf{v}_1 \times_Q \mathbf{v}_2$ and let $W = Q(\mathbf{w})$.*

- (1) *If $w_3 > 0$, then $w_1 > -\sqrt{W}$ and $w_2 > -\sqrt{W}$.*
- (2) *If $w_3 < 0$, then $w_1 < \sqrt{W}$ and $w_2 < \sqrt{W}$.*

Proof. We will consider the sections of C_Q and of $\mathcal{H}(W)$ by the plane $x_3 = w_3$, which appear as the concentric circles depicted in Figure 8. First, assume $w_3 > 0$. Recall that \mathbf{w} lies in the intersection of $\langle \mathbf{v}_1 \rangle^\perp$ and $\langle \mathbf{v}_2 \rangle^\perp$. Note that $\langle \mathbf{v}_1 \rangle^\perp$ appears as a tangent line to the inner circle (the section of C_Q by $x_3 = w_3$) at \mathbf{v}_1 . Since $x, y \geq 0$ the intersection of $\langle \mathbf{v}_1 \rangle^\perp$ and $\langle \mathbf{v}_2 \rangle^\perp$ can lie only on the thick arc of the outer circle in Figure 8. It is easy to verify that the thick arc corresponds to the stated inequalities for w_1 and w_2 . The case for $w_3 < 0$ is similar. \square

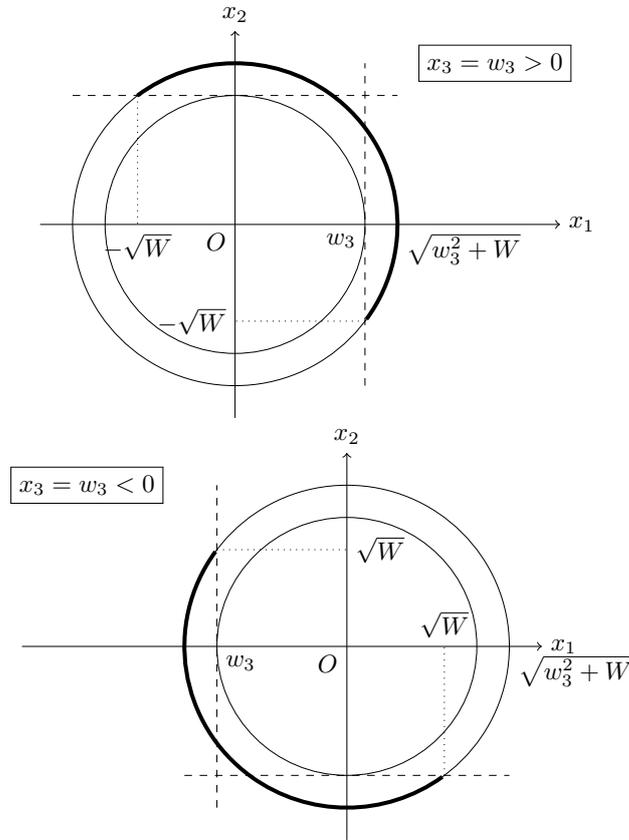


FIGURE 8. The sections of C_Q (the inner circle) and of $\mathcal{H}(W)$ (the outer one) by $x_3 = w_3$. The first picture is for $w_3 > 0$ and the second is for $w_3 < 0$.

The next proposition describes how the integrality of \mathbf{v} is reflected in the twisted product.

Proposition 3.5. *Suppose that $\mathbf{v} \in C_Q$ is defined over a real quadratic field $K = \mathbb{Q}(\sqrt{D})$, whose ring of integers will be denoted by \mathcal{O}_K . We denote by σ the nontrivial Galois automorphism of K over \mathbb{Q} . Assume that \mathbf{v} and \mathbf{v}^σ are linearly independent. If $\mathbf{v} \in (\mathcal{O}_K)^3$, then*

$$\mathbf{v} \times_Q \mathbf{v}^\sigma \in (\sqrt{D}\mathbb{Z}/2)^3.$$

Proof. Let $\mathbf{w} = \mathbf{v} \times_Q \mathbf{v}^\sigma$. It is easy to see from (3.2) that, whenever \mathbf{v} is in $(\mathcal{O}_K)^3$, \mathbf{w} is also in $(\mathcal{O}_K)^3$. At the same time, observe that

$$\mathbf{w}^\sigma = \mathbf{v}^\sigma \times_Q \mathbf{v} = -\mathbf{w}.$$

Hence, we have $\mathbf{w} \in (\sqrt{D}\mathbb{Q})^3$, where $\sqrt{D}\mathbb{Q} := \{\sqrt{D}r \mid r \in \mathbb{Q}\}$. This finishes the proof because $\sqrt{D}\mathbb{Q} \cap \mathcal{O}_K \subseteq \sqrt{D}\mathbb{Z}/2$. \square

3.2. Periodicity is sufficient for quadratic irrationality

We first prove in this subsection that, if $P = (\alpha, \beta) \in \mathcal{Q}$ has an eventually periodic Romik digit expansion, then P is defined over a real quadratic extension of \mathbb{Q} . Suppose that

$$(\alpha, \beta) = [f_1, f_2, \dots, f_l, \overline{d_1, \dots, d_k}]_{\mathcal{Q}} = [f_1, f_2, \dots, f_l, d_1, \dots, d_k, d_1, \dots, d_k, \dots]_{\mathcal{Q}}.$$

Define $P' = (\alpha', \beta') := T^l(P)$. Then clearly,

$$(\alpha', \beta') = [\overline{d_1, \dots, d_k}]_{\mathcal{Q}}$$

is purely periodic. Moreover, if we define $\xi := (\alpha, \beta, 1) \in C_{\mathcal{Q}}$, then ξ represents P , consequently, the vector

$$M_{f_1}^{-1} \dots M_{f_l}^{-1} \xi$$

represents P' . This argument shows that, if α', β' are in a real quadratic field K , then $\alpha, \beta \in K$ as well. Thus, it is enough to prove the statement for the purely periodic case. Therefore, we will assume from now on that

$$P = (\alpha, \beta) = [\overline{d_1, \dots, d_k}]_{\mathcal{Q}}$$

is purely periodic. Then the quadratic irrationality of P is an immediate consequence of Theorem 3.7 below. We give an easy lemma first.

Lemma 3.6. *Suppose that $A \in O_{\mathcal{Q}}(\mathbb{R}^3)$. Let \mathbf{v}_1 and \mathbf{v}_2 be eigenvectors of A , satisfying $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle \neq 0$, and let λ_1, λ_2 be their eigenvalues respectively. Then $\lambda_1 \lambda_2 = 1$. In particular, if \mathbf{v} is an eigenvector of A with $Q(\mathbf{v}) \neq 0$, then its eigenvalue is 1 or -1 .*

Proof. The first statement is an easy consequence of the identity

$$\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = \langle A\mathbf{v}_1, A\mathbf{v}_2 \rangle = \langle \lambda_1 \mathbf{v}_1, \lambda_2 \mathbf{v}_2 \rangle = \lambda_1 \lambda_2 \langle \mathbf{v}_1, \mathbf{v}_2 \rangle,$$

and the second statement follows from this by letting $\mathbf{v}_1 = \mathbf{v}_2$. \square

Theorem 3.7. *Let*

$$M = M_{d_1} \dots M_{d_k}$$

for some $d_1, \dots, d_k \in \{1, 2, 3\}$. Assume that not all $d_1 = \dots = d_k = 1$ and not all $d_1 = \dots = d_k = 3$. Or, equivalently, $M \neq M_1^k$ and $M \neq M_3^k$. Then there exists a real quadratic field extension K of \mathbb{Q} , satisfying the following properties:

- (1) *One of the eigenvalues for M is, say, $\lambda_1 > 1$ with $\lambda_1 \in K$. Another eigenvalue for M is $\lambda_2 := \lambda_1^\sigma$, with $0 < \lambda_2 < 1$. The last eigenvalue $\lambda_3 = \det(M)$, which is either 1 or -1 .*
- (2) *A (nonzero) eigenvector associated to λ_1 is Q -null, which represents a point (α, β) in \mathcal{Q} with $\alpha, \beta \in K$. We write $\mathbf{v}_1 := (\alpha, \beta, 1)$.*
- (3) *$\mathbf{v}_2 := (\alpha^\sigma, \beta^\sigma, 1)$ is an eigenvector associated to λ_2 .*

(4) $\mathbf{v}_3 := \mathbf{v}_1 \times_Q \mathbf{v}_2$ is an eigenvector associated to λ_3 .

Proof. For any finite digit sequence d_1, \dots, d_k , let us define the cylinder set

$$C(d_1, \dots, d_k) = \{(x, y) \in \mathcal{Q} \mid d(T^{j-1}(x, y)) = d_j \text{ for } j = 1, \dots, k\}.$$

Then $C(d_1, \dots, d_k)$ is just the closed sub-arc of \mathcal{Q} whose endpoints are represented by

$$M \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \quad \text{and} \quad M \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

In particular, $C(d_1, \dots, d_k)$ is contained in the interior of \mathcal{Q} unless all of d_j are 1, or all of d_j are 3.

Consider the continuous map from the unit circle \mathcal{U} to itself, defined by

$$(x, y) \mapsto \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \mapsto M \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \mapsto \left(\frac{v_1}{v_3}, \frac{v_2}{v_3} \right).$$

This map sends \mathcal{Q} to $C(d_1, \dots, d_k)$, therefore, must have at least one fixed point, say, (α, β) in $C(d_1, \dots, d_k)$. Then $\mathbf{v}_1 := (\alpha, \beta, 1)$ is an eigenvector of M . Call its eigenvalue λ_1 , then $\lambda_1 > 1$ because M must increase the x_3 -coordinates of the vectors representing any point in \mathcal{Q} (See Figure 7). Applying the same argument to the complement of $C(d_1, \dots, d_k)$ and M^{-1} , we see that there is another eigenvector $\mathbf{v}_2 := (\alpha', \beta', 1)$ with $(\alpha', \beta') \notin \mathcal{Q}$ associated to an eigenvalue, say, λ_2 .

So far, we found two linearly independent eigenvectors $\mathbf{v}_1, \mathbf{v}_2$ of M , both of which are Q -null. Note that $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle \neq 0$, otherwise the \mathbb{R} -linear span of \mathbf{v}_1 and \mathbf{v}_2 would be a subset of the Q -null set C_Q , contradicting the nondegeneracy of $Q(\mathbf{x})$. Then Lemma 3.6 shows that $\lambda_1 \lambda_2 = 1$. Moreover, Proposition 3.2 implies that $\mathbf{v}_3 := \mathbf{v}_1 \times_Q \mathbf{v}_2$ is an eigenvector associated with the third eigenvalue $\lambda_3 = \det(M) = \pm 1$. Observe that the characteristic polynomial of M is a monic polynomial with integer coefficients and its constant term is $\det(M) = \pm 1$. So, λ_1 and λ_2 must be quadratic irrationals and Galois conjugates one another. The remaining statements in the theorem follow from this easily. \square

Even though it is not needed for our proof of Lagrange's theorem, we obtain as a corollary to Theorem 3.7 an analogue of Galois' theorem for the Romik system. Recall that a classical theorem by Galois [13] relates the continued fraction expansion of a quadratic irrational to that of its Galois conjugate.

Corollary 3.8 (Galois' Theorem for the Romik system). *Suppose that*

$$P = (\alpha, \beta) = [\overline{d_1, \dots, d_k}]_{\mathcal{Q}},$$

and write $P^\sigma = (\alpha^\sigma, \beta^\sigma)$ for its Galois conjugate. Then

$$\begin{cases} \alpha^\sigma > 0 \text{ and } \beta^\sigma < 0 & \text{if } d_k = 1, \\ \alpha^\sigma < 0 \text{ and } \beta^\sigma < 0 & \text{if } d_k = 2, \\ \alpha^\sigma < 0 \text{ and } \beta^\sigma > 0 & \text{if } d_k = 3. \end{cases}$$

Moreover,

$$(|\alpha^\sigma|, |\beta^\sigma|) = \overline{[d_{k-1}, d_{k-2}, \dots, d_2, d_1, d_k]}_{\mathcal{Q}}.$$

Proof. Let $\mathbf{v}_1 = (\alpha, \beta, 1)$ as in Theorem 3.7 and write

$$M_{d_1} \cdots M_{d_k} \mathbf{v}_1 = \lambda_1 \mathbf{v}_1.$$

Conjugating this equation with σ and then rewriting this using (2.6) and (2.7),

$$(3.5) \quad HU_{d_1} HU_{d_2} \cdots HU_{d_k} \mathbf{v}_1^\sigma = \lambda_1^\sigma \mathbf{v}_1^\sigma.$$

Now, let $\mathbf{v}' := U_{d_k} \mathbf{v}_1^\sigma$ and multiply (3.5) by U_{d_k} (from the left) to obtain

$$(3.6) \quad M_{d_{k-1}} M_{d_{k-2}} \cdots M_{d_2} M_{d_1} M_{d_k} \mathbf{v}' = \lambda_1 \mathbf{v}'.$$

Here, we use the fact $\lambda_1^\sigma = 1/\lambda_1$ from Theorem 3.7. The equation (3.6) then shows that \mathbf{v}' is an eigenvector of

$$M' := M_{d_{k-1}} M_{d_{k-2}} \cdots M_{d_2} M_{d_1} M_{d_k},$$

associated with the eigenvalue $\lambda_1 > 1$. From Theorem 3.7 again, we conclude that the x_1 - and x_2 -coordinates of \mathbf{v}' are positive. Thus, $U_{d_k} \mathbf{v}_1^\sigma$ has positive x_1 - and x_2 -coordinates, so that $U_{d_k} \mathbf{v}_1^\sigma = (|x^\sigma|, |y^\sigma|, 1)$. This completes the proof. \square

3.3. Periodicity is necessary for quadratic irrationality

Let $P = (\alpha, \beta) \in \mathcal{Q}$ be defined over a (real) quadratic field $K = \mathbb{Q}(\sqrt{D})$. We prove that the Romik digit expansion of P is eventually periodic, completing the proof of Lagrange’s theorem for the Romik system (Theorem 3.1). Note that, if P is rational, its Romik digit expansion ends with infinite succession of 1’s or 3’s, thus is periodic. So, we assume in what follows that P is quadratic irrational.

We fix $\mathbf{v}_0 = (v_1, v_2, v_3) \in (\mathcal{O}_K)^3$ representing P . And define \mathbf{v}_n for each $n \geq 1$ to be

$$\mathbf{v}_n = M_{d_n}^{-1} \cdots M_{d_1}^{-1} \mathbf{v}_0,$$

so that \mathbf{v}_n represents $T^n(P)$ for all $n \geq 1$. Finally, we define

$$\mathbf{w}_n = \mathbf{v}_n \times_Q \mathbf{v}_n^\sigma$$

for all $n \geq 0$. Recall that $\det(M_1) = \det(M_3) = 1$ and $\det(M_2) = -1$. So, if we denote by ϵ_n the count of how many times the digit 2 appears in the sequence $\{d_1, \dots, d_n\}$, then $\det(M_{d_n}^{-1} \cdots M_{d_1}^{-1}) = (-1)^{\epsilon_n}$, and Proposition 3.2 shows

$$(3.7) \quad \mathbf{w}_n = (-1)^{\epsilon_n} M_{d_n}^{-1} \cdots M_{d_1}^{-1} \mathbf{w}_0$$

for each $n \geq 1$. Note that \mathbf{v}_0 and \mathbf{v}_0^σ are linearly independent because they represent two distinct points (α, β) and $(\alpha^\sigma, \beta^\sigma)$. Therefore, \mathbf{w}_0 is nonzero and all \mathbf{w}_n are thus nonzero as well.

Let $W = Q(\mathbf{w}_0) > 0$ and define the one-sheeted hyperboloid $\mathcal{H}(W)$ as in (3.4). Then (3.7) shows $\mathbf{w}_n \in \mathcal{H}(W)$ for all $n \geq 0$. In addition, thanks to Proposition 3.5, $\{\mathbf{w}_n\}_{n=0}^\infty$ is a discrete subset of $\mathcal{H}(W)$.

We claim further that $\{\mathbf{w}_n\}_{n=0}^\infty$ is a *bounded* subset of $\mathcal{H}(W)$, consequently a finite set. This will prove that the periodicity is a necessary condition for the quadratic irrationality, because then at least two \mathbf{v}_n 's, say, \mathbf{v}_{j_1} and \mathbf{v}_{j_2} with $j_1 < j_2$ represent the same point in \mathcal{Q} , therefore

$$(\alpha, \beta) = [d_1, \dots, d_{j_1-1}, \overline{d_{j_1}, d_{j_1+1}, \dots, d_{j_2}}]_{\mathcal{Q}}.$$

We now proceed to proving the claim that $\{\mathbf{w}_n\}_{n=0}^\infty$ is bounded.

It will be convenient to denote by $x_3(\mathbf{w})$ the x_3 -coordinate of \mathbf{w} . The boundedness of $\{\mathbf{w}_n\}_{n=0}^\infty$ is proven by

Step 1: There exists a (large) constant $R > 0$ which depends only on D and W such that, if

$$(3.8) \quad |x_3(\mathbf{w}_n)| > R,$$

then

$$(3.9) \quad |x_3(\mathbf{w}_{n+1})| < |x_3(\mathbf{w}_n)|.$$

In other words, if \mathbf{w}_n lies outside of a certain bounded set, then \mathbf{w}_{n+1} is closer to the origin (see (3.7)).

Step 2: If

$$(3.10) \quad |x_3(\mathbf{w}_n)| \leq R,$$

then

$$(3.11) \quad |x_3(\mathbf{w}_{n+1})| \leq 7\sqrt{2}R.$$

In other words, if \mathbf{w}_n lies *inside* of the same bounded set, then \mathbf{w}_{n+1} cannot escape too far away.

It is easy to deduce the boundedness from Steps 1 and 2 above since

$$|x_3(\mathbf{w}_{n+1})| \leq \max \left\{ |x_3(\mathbf{w}_{n+1})|, 7\sqrt{2}R \right\} \leq \max \left\{ |x_3(\mathbf{w}_n)|, 7\sqrt{2}R \right\}.$$

To prove Step 1, we begin by noting that any point (w_1, w_2, w_3) in the hyperboloid $\mathcal{H}(W)$ must lie close to C_Q if $|w_3|$ is large. Indeed, the distance between the sections (the two concentric circles in Figure 8) of C_Q and $\mathcal{H}(W)$ by the plane $x_3 = w_3$ is

$$\sqrt{w_3^2 + W} - |w_3| = \frac{W}{\sqrt{w_3^2 + W} + |w_3|},$$

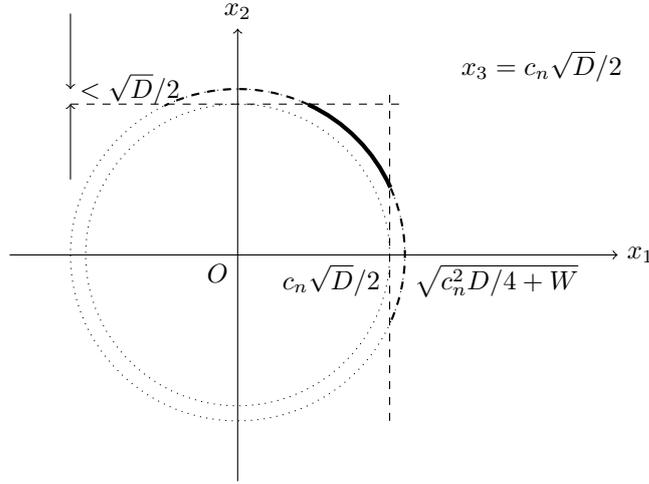


FIGURE 9. The sections of C_Q (the inner circle) and of $\mathcal{H}(W)$ (the outer one) by $x_3 = c_n\sqrt{D}/2$

which can be made arbitrarily small when $|w_3|$ is large. In particular, we choose $R > 0$ large enough so that

$$(3.12) \quad \sqrt{R^2 + W} - R < \sqrt{D}/2.$$

Note from (3.7) that

$$(3.13) \quad \mathbf{w}_{n+1} = \pm M_{d_{n+1}}^{-1} \mathbf{w}_n = \pm U_{d_{n+1}} H \mathbf{w}_n,$$

and assume that \mathbf{w}_n satisfies (3.8) above. Then (3.9) would follow from Proposition 2.1 if we show that \mathbf{w}_n meets the condition (2.8) (where \mathbf{w}_n will play the role of \mathbf{v} in Proposition 2.1).

We know from Proposition 3.5 that there exist $a_n, b_n, c_n \in \mathbb{Z}$ such that $\mathbf{w}_n = (\sqrt{D}/2)(a_n, b_n, c_n)$ for each $n \geq 0$. Let us treat the case $c_n > 0$ first, as the other case is similar. Consider the section of $\mathcal{H}(W)$ by $x_3 = c_n\sqrt{D}/2$ as in Figure 9. In order to verify (2.8) for \mathbf{w}_n , we invoke Proposition 3.4 which says in this case that \mathbf{w}_n must lie in either

- (i) $\{(w_1, w_2, w_3) \in \mathcal{H}(W) \mid w_3 = c_n\sqrt{D}/2, 0 < w_1, w_2 \leq c_n\sqrt{D}/2\}$ (the thick curve in Figure 9), or
- (ii) $\{(w_1, w_2, w_3) \in \mathcal{H}(W) \mid w_3 = c_n\sqrt{D}/2, w_1 > c_n\sqrt{D}/2 \text{ or } w_2 > c_n\sqrt{D}/2\}$ (the dashed curve in Figure 9).

If \mathbf{w}_n lies in (i), then (2.8) is easily verified. We show that \mathbf{w}_n cannot lie in (ii), which will then complete the proof of Step 1. From (3.8) and the condition

(3.12), we see that

$$\sqrt{\left(\frac{\sqrt{D}c_n}{2}\right)^2 + W} - \frac{\sqrt{D}c_n}{2} < \sqrt{R^2 + W} - R < \sqrt{D}/2.$$

So, if $a_n\sqrt{D}/2 > c_n\sqrt{D}/2$, then

$$c_n\sqrt{D}/2 < a_n\sqrt{D}/2 < (c_n + 1)\sqrt{D}/2,$$

which is impossible because a_n and c_n are integers. Likewise it is impossible to have $b_n\sqrt{D}/2 > c_n\sqrt{D}/2$. This proves that \mathbf{w}_n cannot lie in (ii).

We now move on to proving Step 2. Writing $\mathbf{w}_n = (w_1, w_2, w_3)$, the condition (3.10) implies

$$w_1^2 + w_2^2 = w_3^2 + R \leq R^2 + R,$$

so that

$$|w_1| \leq \sqrt{R^2 + R} \leq \sqrt{2}R,$$

and similarly, $|w_2| \leq \sqrt{2}R$. Now, we use (3.13), the triangle inequality and the simple calculation

$$H \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} * \\ * \\ -2w_1 - 2w_2 + 3w_3 \end{pmatrix}$$

together to deduce (3.11). This completes Step 2, thus the proof of the claim that $\{\mathbf{w}_n\}$ is bounded.

4. Some open problems

4.1. Berggren graphs over real quadratic fields

Fix a (real) quadratic extension K of \mathbb{Q} . Can we find Berggren trees defined over K , similar to Figure 1, namely, a collection of triples (x, y, z) with $x^2 + y^2 = z^2$ where x, y, z are positive (coprime) K -integers with the edges showing the actions of M_1, M_2, M_3 ?

When we consider K -integer triples (x, y, z) , it is convenient to remove the ambiguity stemming from the existence of *units* in K . As is well-known, there exists a *fundamental unit*, say, $\epsilon_K > 1$, such that the set of all the units of K are given by $\{\pm\epsilon_K^{\mathbb{Z}}\}$. (See, for example, Proposition 13.1.6 in [9].) Let us denote by $[x, y, z]$ the equivalent class defined by the relation

$$(x, y, z) \sim (x', y', z') \iff (x, y, z) = \epsilon_K^k (x', y', z')$$

for some $k \in \mathbb{Z}$. Note that any triple in the class $[x, y, z]$ represents a unique point $(x/z, y/z) \in \mathcal{Q}$. Write its Romik digit expansion

$$(x/z, y/z) = [d'_1, \dots, d'_t, \overline{d_1, \dots, d_k}]_{\mathcal{Q}}.$$

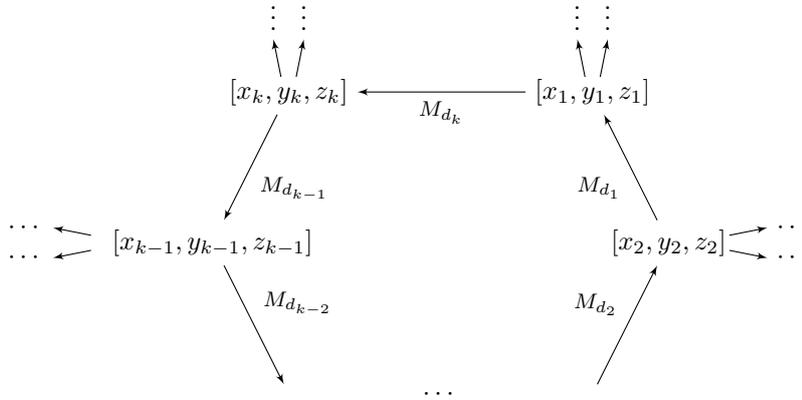


FIGURE 10. Purely periodic points form a circular root.

Then, our results in §3 show that the class $[x, y, z]$ belongs to a unique Berggren tree that grows out of a *circular root*, as pictured in Figure 10. Here the points $(x_1/z_1, y_1/z_1), \dots, (x_k/z_k, y_k/z_k)$ have the Romik digit expansions

$$\begin{aligned} (x_1/z_1, y_1/z_1) &= [\overline{d_1, \dots, d_k}]_{\mathcal{Q}}, \\ (x_2/z_2, y_2/z_2) &= [\overline{d_2, d_3, \dots, d_k, d_1}]_{\mathcal{Q}}, \\ &\dots \\ (x_k/z_k, y_k/z_k) &= [\overline{d_k, d_1, d_2, \dots, d_{k-1}}]_{\mathcal{Q}}. \end{aligned}$$

In other words, a tree growing out of the circular root will contain all the triples whose Romik digit expansions share the same period, up to circular permutation. Perhaps, a better nomenclature in this situation is a *Berggren graph*, as it is no longer a tree in the graph-theoretic sense. An example of a Berggren graph in $\mathbb{Q}(\sqrt{3})$ rooted at $[1, \sqrt{3}, 2]$ and $[\sqrt{3}, 1, 2]$ is provided in Figure 11.

The upshot of our discussion so far is the following. In order to answer the question posted at the beginning of this section, we must find all the circular roots defined over K , from which we then grow Berggren graphs to obtain the entire integer K -triples. Therefore, a natural question to ask is, *does there exist an algorithm for finding all the purely periodic points in \mathcal{Q} defined over K ?* Related, here is another question, perhaps a bit weaker. For a finite Romik digit sequence $\vec{d} = [d_1, \dots, d_k]$ with $d_j \in \{1, 2, 3\}$, define $(x(\vec{d}), y(\vec{d})) \in \mathcal{Q}$ to be the point whose Romik digit expansion is purely periodic with \vec{d} being its period, namely,

$$(x(\vec{d}), y(\vec{d})) = [\overline{d_1, \dots, d_k}]_{\mathcal{Q}}.$$

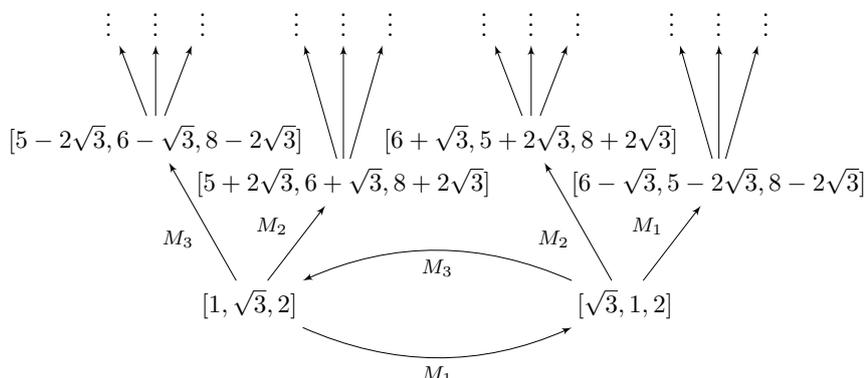


FIGURE 11. An example of Berggren graph in $\mathbb{Q}(\sqrt{3})$, rooted at $[1, \sqrt{3}, 2]$ and $[\sqrt{3}, 1, 2]$.

Then, does the number

$$N(k, K) := \#\{\vec{d} \in \{1, 2, 3\}^k \mid x(\vec{d}), y(\vec{d}) \in K\}$$

of such points *that are defined over* K tend to infinity as $k \rightarrow \infty$? If so, how fast? Does the growth rate depend on (the discriminant of) K ?

4.2. Romik systems defined by other quadratic forms

Another way to generalize the work in this paper is to consider Romik systems stemming from other quadratic forms considered in [4]. We believe that, for certain forms such as $Q(x, y, z) = x^2 + xy + y^2 - z^2$, much of the arguments presented here will carry over essentially in the same way. But, we don't expect that every quadratic form (with a Berggren tree) would always yield the same result. Particularly interesting to us is the form $Q(x, y, z, w) = x^2 + y^2 + z^2 - w^2$ for *Pythagorean quadruples*. This quadratic form and its Berggren tree have been recently studied by Chaubey et al. in [5] in a slightly different context. Will a Romik system arising from its Berggren trees consisting of Pythagorean quadruples have similar properties as described in this paper? Will there be any interesting arithmetical consequences?

References

- [1] R. C. Alperin, *The modular tree of Pythagoras*, Amer. Math. Monthly **112** (2005), no. 9, 807–816. <https://doi.org/10.2307/30037602>
- [2] F. J. M. Barning, *On Pythagorean and quasi-Pythagorean triangles and a generation process with the help of unimodular matrices*, Math. Centrum Amsterdam Afd. Zuivere Wisk. **1963** (1963), ZW-011, 37 pp.
- [3] B. Berggren, *Pytagoreiska triangular*, Tidskrift för elementär matematik, fysik och kemi **17** (1934), 129–139.

- [4] B. Cha, E. Nguyen, and B. Tauber, *Quadratic forms and their Berggren trees*, J. Number Theory **185** (2018), 218–256. <https://doi.org/10.1016/j.jnt.2017.09.003>
- [5] S. Chaubey, E. Fuchs, R. Hines, S. Robert, and E. Katherine, *The dynamics of super-Apollonian continued fractions*, Trans. Amer. Math. Soc. (2017); <https://arxiv.org/abs/1703.08616>.
- [6] K. Conrad, *Pythagorean descent*, <http://www.math.uconn.edu/~kconrad/blurbs/lin-multialg/descentPythag.pdf>.
- [7] R. W. R. Darling, *Differential Forms and Connections*, Cambridge University Press, Cambridge, 1994. <https://doi.org/10.1017/CB09780511805110>
- [8] L. Fishman, D. Kleinbock, M. Dmitry, S. Keith, and D. Simmons, *Intrinsic Diophantine approximation on manifolds: general theory*, Trans. Amer. Math. Soc. **370** (2018), no. 1, 577–599. <https://doi.org/10.1090/tran/6971>
- [9] K. Ireland and M. Rosen, *A Classical Introduction to Modern Number Theory*, second edition, Graduate Texts in Mathematics, **84**, Springer-Verlag, New York, 1990. <https://doi.org/10.1007/978-1-4757-2103-4>
- [10] D. Kleinbock and K. Merrill, *Rational approximation on spheres*, Israel J. Math. **209** (2015), no. 1, 293–322. <https://doi.org/10.1007/s11856-015-1219-z>
- [11] G. Panti, *A general Lagrange theorem*, Amer. Math. Monthly **116** (2009), no. 1, 70–74. <https://doi.org/10.4169/193009709X469823>
- [12] ———, *Billiards on pythagorean triples and their Minkowski functions*, arXiv:1902.00414 [math.NT].
- [13] O. Perron, *Die Lehre von den Ketterbrüchen*, Teubner, Leipzig, 1929.
- [14] D. Romik, *The dynamics of Pythagorean triples*, Trans. Amer. Math. Soc. **360** (2008), no. 11, 6045–6064. <https://doi.org/10.1090/S0002-9947-08-04467-X>

BYUNGCHUL CHA
 DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE
 MUHLENBERG COLLEGE
 2400 CHEW ST, ALLENTOWN, PA, 18104, USA
Email address: cha@muhlenberg.edu

DONG HAN KIM
 DEPARTMENT OF MATHEMATICS EDUCATION
 DONGGUK UNIVERSITY
 SEOUL 04620, KOREA
Email address: kim2010@dongguk.edu