Publl. Math. Debrecen
89/1-2 (2016), 203–221
DOI: 10.5486/PMD.2016.7470

On the reduction of binary quadratic forms

By AYBERK ZEYTIN (İstanbul)

Abstract. We give an interpretation of the reduction algorithm of Gauss in terms of çarks, which are certain types of infinite ribbon graphs (or infinite dessins). We then describe an alternative reduction which is slightly faster than Gauss'. We also solve the minimal value problem and describe an algorithmic solution to the representation problem of indefinite binary quadratic forms.

1. Introduction

Binary quadratic forms are homogeneous degree two polynomials in two variables with integer coefficients. They are in a sense the first non-trivial case of Diophantine equations. The history of the question of finding integer solutions of such equations dates back to ancient Greece and it was Gauss who has treated the case of binary quadratic forms in a complete and systematic manner [4].

The modular group, denoted by $\mathrm{PSL}_2(\mathbf{Z})$, acts on the set of binary quadratic forms by coordinate change. An orbit of this action is called the class of a form, and the stabilizer of a form is called the automorphism group of the form. Two forms are called equivalent if and only if they belong to the same orbit. The discriminant of a form and the set of values of a form are left-invariant by this action.

Gauss searched for a canonical representative of each class and defined reduced forms. Gauss has proved that if the form has negative discriminant, i.e. if

Mathematics Subject Classification: 11E16.
Key words and phrases: çarks, binary quadratic forms, reduction theory, representation problem of binary quadratic forms, minimum problem of binary quadratic forms.
This research is supported by Galatasaray University Research Fund Grant no. 13.504.001 and TÜBİTAK 3501 Career Grant no. 113R017.

the form is definite, then its class contains a unique reduced form. However, if the form has positive discriminant, i.e. if it is indefinite, then there always exist at least two reduced forms in its class. Gauss gave an algorithm whose input is an arbitrary form and whose output is a reduced form which is equivalent to the initial form[1]. Zagier also defined a reduction in [10, §2.13] which simplifies proofs, however, Gauss' algorithm is better suited for computational purposes.

In this work, we use a new tool called çark[2] discovered by the author and collaborators in [9] and solve classical questions and give alternative algorithms concerning binary quadratic forms. More precisely, the aim of the article is twofold. After recalling basic definitions and facts concerning binary quadratic forms, we give an interpretation of Gauss' reduction algorithm and then describe another reduction algorithm which is slightly faster than that of Gauss'. We then move on to the classical representation problem of binary quadratic forms which asks whether there exist integer solutions to an equation of the form $aX^2 + bXY + cY^2 = N$; where a, b, c are integers. If the form is definite, then there can only be at most finitely many solutions. However, in the indefinite case starting with one solution one may produce infinitely many solutions using elements of the automorphism group. Here we prove that an answer to the solubility of such an equation can be given after a finite number of steps and give an algorithm. Along the way we also touch upon some facts concerning binary quadratic forms which are merely observations in the language of çarks. We also solve the minimal problem which asks the smallest element of the set whose elements are absolute values of the numbers attained by an indefinite binary quadratic form.

Let us mention finally that our computations are done on PARI/gp [7]. We would also like to add that the reduction algorithm given in this article, together with further visualizations of çarks, is being developed and will be available on the homepage of the author under the name "Sunburst" [6].

2. Notation and terminology

This section is devoted to introducing the notation. We also state some results concerning binary quadratic forms.

2.1. Binary quadratic forms. A binary quadratic form is a homogeneous polynomial f of two integer variables, X and Y, of degree 2:

[1] Lagrange also defined a reduction theory which was slower than that of Gauss'.

[2] pronounced '*chark*'

$$f(X,Y) = aX^2 + bXY + cY^2 = (X\ Y)\begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix}(X\ Y)^t;$$

where a, b and c are integers. We write $f = (a,b,c)$ for short. The matrix $\begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix}$ is denoted by M_f. Forms f with negative discriminant, $\Delta(f) = b^2 - 4ac$, are called positive (resp. negative) definite if $a > 0$ (resp. $a < 0$) and forms with positive discriminant are called indefinite. A binary quadratic form, $f = (a,b,c)$, is said to be *primitive* if the greatest common divisor of a, b and c is 1. Throughout we will always consider primitive indefinite binary quadratic forms whose discriminant is not a perfect square[3].

The action of the modular group $\mathrm{PSL}_2(\mathbf{Z})$ on the set of all binary quadratic forms is defined as:

$$W \cdot M_f \mapsto W^t\, M_f\, W$$

The orbit of a binary quadratic form f under $\mathrm{PSL}_2(\mathbf{Z})$-action will be referred to as its class and denoted by $[f]$. For a binary quadratic form f, the group $\mathrm{Stab}(f)$ is called the automorphism group of f. The discriminant of f and the set of values of f, i.e. the set $\{f(X,Y)\,|\,X,Y \in \mathbf{Z}\}$, are invariant under this action.

2.2. The correspondence between çarks and binary quadratic forms. The modular group is isomorphic to the free product of a cyclic group of order 2 with a cyclic group of order 3: $\mathrm{PSL}_2(\mathbf{Z}) \cong \mathbf{Z}/2\mathbf{Z} * \mathbf{Z}/3\mathbf{Z}$. Throughout, we fix

$$S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},\ L = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}$$

to be the generators of the order 2 and 3 cyclic subgroups, respectively. We have

$$\mathrm{PSL}_2(\mathbf{Z}) \cong \left\langle S, L | S^2 = L^3 = I \right\rangle.$$

To the modular group we associate the bipartite Farey tree, $\mathcal{F}$, whose set of vertices, $V(\mathcal{F})$ is the disjoint union of $V_\circ$ and $V_\bullet$; where $V_\circ(\mathcal{F}) = \{\{W, WS\}\colon W \in \mathrm{PSL}_2(\mathbf{Z})\}$ and $V_\bullet(\mathcal{F}) = \{\{W, WL, WL^2\}\colon W \in \mathrm{PSL}_2(\mathbf{Z})\}$. We define the set of edges as $E(\mathcal{F}) = \{\{W\}\colon W \in \mathrm{PSL}_2(\mathbf{Z})\}$. There is an edge joining any two vertices v and v' of $\mathcal{F}$ if and only if their intersection is both non-empty and a proper subset of v and v'. In particular, there are no loops. The three edges W, WL and WL^2 emanating from the vertex $\{W, WL, WL^2\}$ are ordered as given, see Figure 1 and 2.

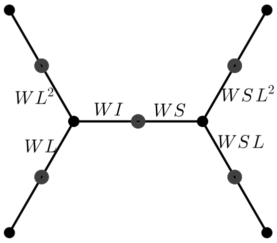

Figure 1. The bipartite Farey tree: local picture

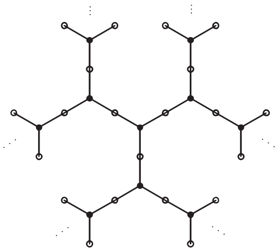

Figure 2. The bipartite Farey tree: global picture

The modular group acts on $\mathcal{F}$ from left by translation and therefore for every subgroup Γ of the modular group we may define the corresponding quotient graph $\Gamma\backslash\mathcal{F}$ called a *modular graph*[4], see [8] for further treatment of modular graphs. In the most general setting, given a set X with an action of a group G, for any subgroup H of G one can construct the quotient $H\backslash X$ as the set of orbits of H on X. For instance, the space X itself corresponds to the trivial subgroup. And, indeed, in the language of topological spaces this is reduced to the classical functor between subgroups of the fundamental group and topological covers of the space X. This applies to the case at hand, producing the quotient graphs. More precisely, the generator S acts like a rotation of degree 2 about the vertex $\{I, S\}$ (denoted by $\circ$) and L acts as a rotation of degree 3 about the vertex $\{I, L, L^2\}$ (denoted by $\bullet$). The classical base point considerations carry over to this setting as well. Taking into account the fact that any element of finite order in $\mathrm{PSL}_2(\mathbf{Z})$

[3] This is no loss in generality as the reduction procedure is the same in these cases.

[4] The resulting graph is called a 'dessin d'enfant' if it has finitely many edges. Our terminology unifies other essentially equivalent terms in the literature, such as Linienzug of Klein, lozenges, triangulations, etc.

is conjugate either to S or L or L^2 we conclude that there are only two modular graphs arising from such quotients, see Figure 3.

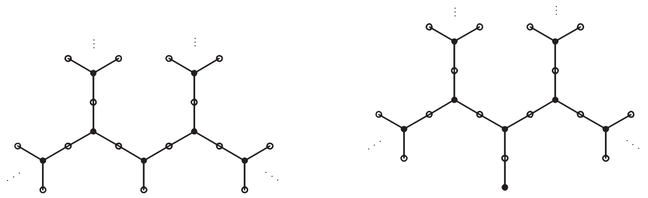

Figure 3. The modular graphs $\langle S\rangle\backslash\mathcal{F}$ and $\langle L\rangle\backslash\mathcal{F}$, respectively

One can immediately deduce from the action of L and S the fact that the LS acts as a translation on $\mathcal{F}$, and the quotient of $\mathcal{F}$ by the subgroup generated by LS is then an infinite planar graph with a unique cycle consisting of 2 edges, see Figure 4. For the algebraic definition of this construction the reader may consult [9].

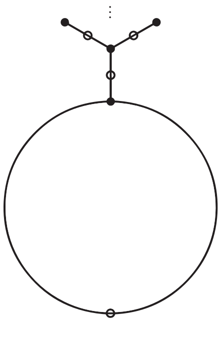

Figure 4. The modular graph $\langle LS\rangle\backslash\mathcal{F}$

Definition 2.1. A çark is the quotient of $\mathcal{F}$ by a non-trivial subgroup generated by one element.

In fact, the set of edges in a çark is the orbits of the subgroup $\langle W\rangle$ in the set of edges of $\mathcal{F}$, and the set of vertices of the çark can be identified with the orbits of $\langle W\rangle$ in the set of vertices of $\mathcal{F}$. The *shape* of the quotient graph is determined by the type of the element, and, for our purposes, we will only deal with those subgroups that are generated by a *hyperbolic* element. In this case, a *çark* is an infinite bipartite ribbon graph, denoted by Ç, which has a unique loop (called *spine*) and finitely many bipartite Farey tree components, called *Farey branches*, attached to all degree 3 vertices that are on the spine, see Figure 5. An *oriented çark* is a çark whose spine is oriented[5]. A çark is called *base-pointed* if an edge is specified to be the base edge. Note that the surface constructed from Ç is an annulus. Let us also note at this point that the spine of the çark is the quotient of the river which appeared in [3].

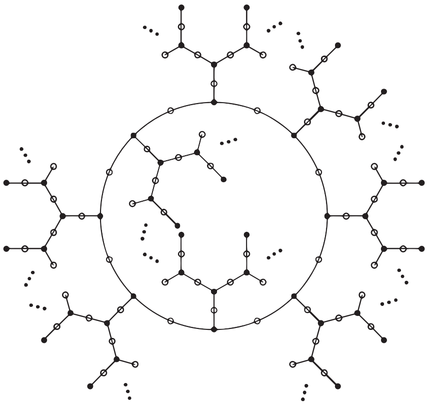

Figure 5. Çark corresponding to the class $[(-7, 8, 2)]$

Successive Farey branches pointing in the same boundary component are called a Farey bunch. We introduce the following notation: a Farey bunch of size n (i.e. a bunch containing n Farey branches) pointing in the inner boundary circle is denoted by a $+n$, and, similarly, a Farey bunch of size n pointing in the outer boundary circle is denoted by a $-n$. For instance, çark appearing in Figure 5 will be denoted by $[-4, +1, -2, +1]$. Remark that as we consider primitive forms, the çarks that we encounter will not be periodic. For instance, we will not be dealing with the çark $[-4, +1, -2, +1, -4, +1, -2, +1]$. Thus, we will always assume that çarks are not periodic. Table 1 gives a list of first few çarks of binary quadratic forms of Pell. For readers interested in computations, we refer to our software [6].

There is a one-to-one correspondence between çarks and classes of indefinite binary quadratic forms. In addition, we have the following base-pointed version:

Theorem 2.2 ([9, Corollary 3.2])**.** *There is a one-to-one correspondence between oriented, base-pointed çarks and indefinite binary quadratic forms.*

[5] In our drawings we will always assume that the spine of the çark is counterclockwise oriented.

binary quadratic form	corresponding çark	binary quadratic form	corresponding çark
$(1,0,-2)$	$[-2,2]$	$(1,0,-20)$	$[-2,8]$
$(1,0,-3)$	$[-1,2]$	$(1,0,-21)$	$[-1,1,-2,1,-1,8]$
$(1,0,-5)$	$[-4,4]$	$(1,0,-22)$	$[-1,2,-4,2,-1,8]$
$(1,0,-6)$	$[-2,4]$	$(1,0,-23)$	$[-1,3,-1,8]$
$(1,0,-7)$	$[-1,1,-1,4]$	$(1,0,-24)$	$[-1,8]$
$(1,0,-8)$	$[-1,4]$	$(1,0,-26)$	$[-10,10]$
$(1,0,-10)$	$[-6,6]$	$(1,0,-27)$	$[-5,10]$
$(1,0,-11)$	$[-3,6]$	$(1,0,-28)$	$[-3,2,-3,10]$
$(1,0,-12)$	$[-2,6]$	$(1,0,-29)$	$[-2,1,-1,2,-1,2,-1,1,-2,10]$
$(1,0,-13)$	$[-1,1,-1,1,-6,1,-1,1,-1,6]$	$(1,0,-30)$	$[-2,10]$
$(1,0,-14)$	$[-1,2,-1,6]$	$(1,0,-31)$	$[-1,1,-3,5,-3,1,-1,10]$
$(1,0,-15)$	$[-1,6]$	$(1,0,-32)$	$[-1,1,-1,10]$
$(1,0,-17)$	$[-8,8]$	$(1,0,-33)$	$[-1,2,-1,10]$
$(1,0,-18)$	$[-4,8]$	$(1,0,-34)$	$[-1,4,-1,10]$
$(1,0,-19)$	$[-2,1,-3,1,-2,8]$	$(1,0,-35)$	$[-1,10]$

Table 1. First few Pell çarks.

Indeed, given a base-pointed çark one makes a one-counterclockwise-oriented full turn around the spine. On visiting a vertex of degree two, one writes an S and on visiting a vertex of degree three one writes L if one proceeds left, and L^2 if one proceeds right afterwards. In the end, we obtain a word in S, L and L^2, and hence an element, say $W = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ in $\mathrm{PSL}_2(\mathbf{Z})$. The binary quadratic form, $f_{\text{Ç}}$ corresponding to Ç is the homogenization of the fixed point equation of W without common divisors:

$$f_{\text{Ç}}(X,Y) = \frac{c}{\delta}X^2 + \frac{d-a}{\delta}XY + \frac{-b}{\delta}Y^2;$$

where $\delta = \gcd(c, d-a, b)$. Remark that W is an automorphism of the corresponding binary quadratic form.

Example 2.3. For the base-pointed çark in Figure 6, we record the word $W = SL^2(SL^2)(SL)^2SL^2(SL)^3SL^2S = \begin{pmatrix} 33 & 26 \\ -14 & -11 \end{pmatrix}$. In this case, $f_{\text{Ç}} = (-7,-22,-13)$.

Remark 2.4. The correspondence stated in Theorem 2.2 carries over to the language of subgroups. More precisely, there are one-to-one correspondences between

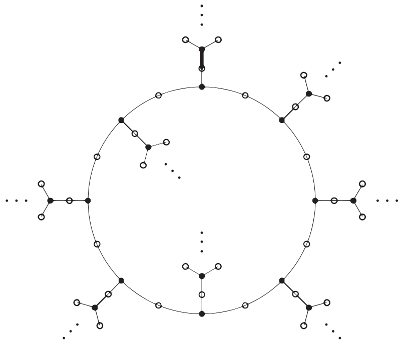

Figure 6. Çark with a base edge corresponding to the form $(-7,-22,-13)$ in $[(-7,8,2)]$

- the conjugacy classes of subgroups of $\mathrm{PSL}_2(\mathbf{Z})$ generated by a hyperbolic element and çarks,
- the subgroups of $\mathrm{PSL}_2(\mathbf{Z})$ generated by a hyperbolic element and base-pointed çarks,
- the set of pairs consisting of a subgroups of $\mathrm{PSL}_2(\mathbf{Z})$ generated by a hyperbolic element and a chosen generator and the set of oriented, base-pointed çarks.

Moreover, the conjugation action of $\mathrm{PSL}_2(\mathbf{Z})$ on itself amounts to translation of the base edge in the language of çarks.

2.3. Reduced forms and reduced edges. Every class of a binary quadratic form contains a finite set of distinguished forms that satisfy the following inequality:

$$\left|\sqrt{\Delta(f)} - 2\,|a|\right| < b < \sqrt{\Delta(f)}.$$

Binary quadratic forms satisfying this inequality are called *reduced* [4]. We define an edge of a çark to be reduced if it is neighbour to a + Farey bunch, see Figure 7. In [9], it is proven that the edges that represent reduced form in $[f]$ are exactly the reduced edges.

If f is indefinite, the set of reduced forms in the class of f has at least two elements. Similar to the Gauss' proof of finiteness of reduced elements, one can show that given an indefinite binary quadratic form $f = (a,b,c)$, forms $f' = (a',b',c')$ which are equivalent to f and satisfy $a'c' < 0$ are finite. We call such forms *spinal* or *semi-reduced*. In fact, the following is merely an observation:

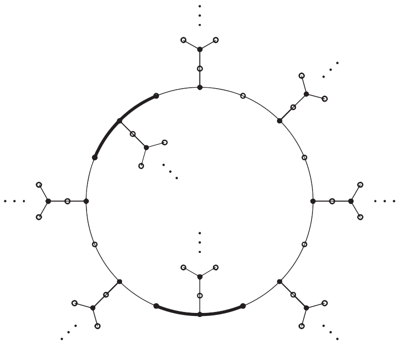

Figure 7. Bold edges correspond to reduced binary quadratic forms in $[(-7,8,2)]$

Proposition 2.5. *Given an indefinite binary quadratic form, there is a one-to-one correspondence between*

- *the set of spinal forms in the class* $[f]$ *and the edges of the spine,*
- *the set of forms,* $f' = (a,b,c)$ *in the class* $[f]$ *with* $a', c' > 0$ *and the edges on the Farey components of the çark which are directed in the inner boundary component (i.e. inner Farey component), and*
- *the set of forms,* $f' = (a,b,c)$ *in the class* $[f]$ *with* $a', c' < 0$ *and the edges on the Farey components of the çark which are directed in the outer boundary component (i.e. outer Farey component).*

And in particular, if f is an indefinite binary quadratic form, then $[f]$ contains at least 2 and at most finitely many reduced forms.

2.4. Reduced edges and faces of a çark. A *finite* path on a çark is defined as a sequence of edges of a çark, $\gamma = (e_1, e_2, \ldots, e_n)$, such that the intersection of two consecutive edges is a vertex and such that there are neither repetitions nor backtracks. The initial edge of γ is e_1 and the final edge is e_n. The initial vertex of γ is the vertex of e_1 which is not equal to the vertex defined by the orbit of the vertex $\{e_1, e_2\}$. Similarly, the final (or terminal) vertex of γ is the vertex of e_n which is not equal to the vertex defined by the orbit of the vertex $\{e_{n-1}, e_n\}$. The length of a finite path, $\ell(\gamma)$ is defined to be the number of edges it contains. An infinite and a bi-infinite path may also be defined accordingly [5].

A left (resp. right) turn path is a path on a çark in which every turn is a left (resp. right) turn. In fact, if γ is a finite left-turn path, then there is a word,

$M \in \mathrm{PSL}_2(\mathbf{Z})$ such that

$$\gamma = (M, MS, MSL, \cdots, M(SL)^{n-1}S)$$

or

$$\gamma = (M, MS, MSL, \cdots, M(SL)^{n-1}S, M(SL)^n)^6.$$

Similarly, if γ is a finite right-turn path, then there is a word, $M \in \mathrm{PSL}_2(\mathbf{Z})$ such that

$$\gamma = (M, MS, MSL^2, \cdots, M(SL^2)^{n-1}S)$$

or

$$\gamma = (M, MS, MSL^2, \cdots, M(SL^2)^{n-1}S, M(SL^2)^n).$$

In a similar fashion, one can define infinite and bi-infinite left and right-turn paths on a çark. Remark that every bi-infinite left-turn path is at the same time a bi-infinite right-turn path and vice versa. To avoid ambiguity, we will consider only bi-infinite left-turn paths.

Definition 2.6. A *face* of a çark, Ç, is defined to be a bi-infinite left-turn path on Ç. A face which includes an edge on the spine will be referred to as a *spinal* face and a face will be called *reduced* if it contains a reduced edge. Faces which are not spinal will be called *non-spinal.* A face which is not reduced is called *non-reduced.*

A few remarks are in order. If a face, φ, contains an edge on an inner (resp. outer) Farey component, then either all or all but finitely edges of φ lie on an inner (resp. outer) Farey component. Such faces are called *inner* (resp. *outer*) faces. We also conclude that a reduced face contains exactly two reduced forms. However, for any $k \in \mathbf{Z}_{\geq 2}$ there is a çark containing a face which has $2k$ many spinal edges. Spinal but not reduced faces contain exactly two spinal edges.

If φ is a non-reduced face, then there is a unique vertex, $\nu_\varphi = \{W, WS\}$, of φ with the property that the two infinite paths following the same path as φ and starting at ν_φ are left-turn and right-turn paths. This vertex is called the *vertex* of the face φ. The left-turn path starting at ν_φ is denoted by φ_l and the right-turn path is denoted by φ_r, see Figure 8.

Given a face φ of a çark Ç, we define the *level* of φ to be 0 if φ is a spinal face, and to be the minimum of the set

$$\{\ell(\gamma)\colon \gamma \text{ is any path from } \nu_\varphi \text{ to a vertex on the spine}\}$$

[6] By abuse of notation, we identify orbits with their representatives.

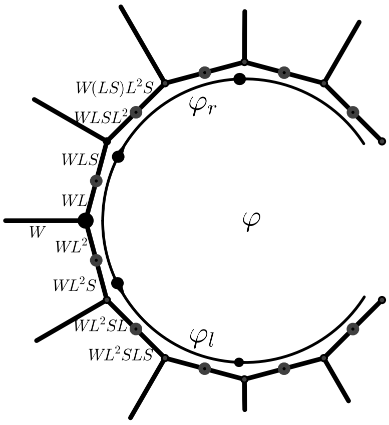

Figure 8. The face φ, its vertex $\{W, WL, WL^2\}$ and the paths φ_l and φ_r

if φ is an inner face and

$$\{-\ell(\gamma)\colon \gamma \text{ is any path from } \nu_\varphi \text{ to a vertex on the spine}\}$$

if φ is an outer face. Remark that the level of an edge is always an even integer. The set of faces, $F(\text{Ç})$ of a çark can be written as a disjoint union:

$$F(\text{Ç}) = \bigsqcup_{n \in 2\mathbf{Z}} F_n(\text{Ç});$$

where $F_n(\text{Ç})$ is defined to be the set of faces of Ç which are of level n.

3. Reduction theory of indefinite binary quadratic forms

3.1. Reduction algorithm of Gauss via çarks. Let us recall briefly the reduction algorithm of Gauss: let $f = (a, b, c)$ be any binary quadratic form. Let $t(f)$ be the function defined by:

$$t(f) = \left\{ \begin{array}{cl} \operatorname{sign}(c) \left\lfloor \frac{b}{2|c|} \right\rfloor & \text{if} \quad |c| \geq \sqrt{\Delta}, \\ \operatorname{sign}(c) \left\lfloor \frac{\sqrt{\Delta}+b}{2|c|} \right\rfloor & \text{if} \quad |c| < \sqrt{\Delta}. \end{array} \right\}$$

The following algorithm, see [1, §6.4] or [2, §5.6], takes f as an input and produces a reduced form f_r equivalent to f in finitely many steps:

0. Set $f_r = f$.
1. If f_r is reduced then return f_r
2. Else assign $U_{f_r} \cdot f_r$ to f_r and return to 1;

where $U_{f_r} = \begin{pmatrix} 0 & -1 \\ 1 & t(f_r) \end{pmatrix}$. Note that $U_{f_r} = S(LS)^{t(f_r)}$. If the form f is reduced, then the set $[f]$ contains a unique reduced form. If f is indefinite, the number of reduced forms is finite and always larger than 1.

Reduction algorithm of Gauss can be considered as an algorithm which produces a sequence of edges on the oriented çark with a base edge corresponding to f whose initial edge is the base edge and whose terminal edge is one of the reduced edges on the spine of Ç. We denote this sequence of edges by $\gamma_{f,Gauss}$. Remark that $\gamma_{f,Gauss}$ is not a path in our terminology as it has backtracks, see Remark 3.4.

Lemma 3.1. *Let Ç$_f$ denote the çark corresponding to the class of an indefinite binary quadratic form f and let f' be a spinal but non-reduced form in $[f]$. Then except for the terminal edge $\gamma_{f',Gauss}$ does not contain any reduced edge.*

Proof. Each step in the reduction algorithm can have only left turns or right turns, but not both. Hence the word recorded in a single step will contain powers of LS or powers of L^2S only. But, for an edge to be reduced either one has an LSL^2S sequence or L^2SLS sequence, which cannot be encoded in $\gamma_{f',Gauss}$. □

We conclude that once we obtain a spinal edge, then in at most one step the reduction algorithm terminates.

3.2. Reduction algorithm using values. Let $\mathcal{B} = \{e_1 = (1,0), e_2 = (0,1)\}$ be the standard *oriented*[7] basis of $\mathbf{Z}^2$ as a $\mathbf{Z}$ module and let $f = (a,b,c)$ be an indefinite binary quadratic form of discriminant $\Delta > 0$. Before describing the algorithm, let us note that it is simple to reduce an indefinite binary quadratic form of the form (a,b,a). First observe that such a form cannot be reduced, or even semi-reduced. Assume without loss of generality that $ab < 0$, if not, we can always replace (a,b,a) by $S \cdot (a,b,a) = (a,-b,a)$. Given such a form, both $L \cdot (a,b,a)$ and $L^2 \cdot (a,b,a)$ are semi-reduced. Indeed, $L \cdot (a,b,a) = (2a + b, -(2a+b), a)$ and $L^2 \cdot (a,b,a) = (a, -(2a+b), 2a+b)$. In both cases, the product $a(2a+b) = 2a^2 + ab$ is negative as $b^2 - 4a^2 > 0 \Leftrightarrow |b| > 2|a| \Leftrightarrow |ab| = -ab > 2a^2 = 2|a|^2$. A direct consequence of this observation is the following:

[7] We use oriented bases to obtain an element of $\mathrm{PSL}_2(\mathbf{Z})$. Similar constructions hold for $\mathrm{PGL}_2(\mathbf{Z})$, too.

Proposition 3.2. *Given an indefinite binary quadratic form $f = (a, b, a)$, the edge of the çark corresponding to f is always spinal, i.e. it is included in a spinal face.*

We can also deduce that the forms $LS \cdot (a, b, a)$ and $L^2S \cdot (a, b, a)$ are also reduced. More generally, one can show that the set of n's for which the form $(LS)^n \cdot (a, b, a)$ and $(L^2S)^n \cdot (a, b, a)$ is finite and depends on the class $[(a, b, a)]$. There are classes of forms which does not contain any binary quadratic form of the form (a, b, a), e.g. there is no such form of discriminant 40.

The next is an algorithm which terminates on arrival at a spinal edge, i.e. a semi-reduced form, of the spine.

0. If $ab > 0$, then replace f with $S \cdot f = (c, -b, a)$.
1. If $f(e_1)f(e_2) < 0$, return $f_{red} = (e_1|e_2) \cdot f$.
2. Else if $f(e_1) = f(e_2)$, then return $((e_1|e_2)\, L) \cdot f$.
3. Else if $|f(e_2)| > |f(e_1)|$, then let n be the first positive integer satisfying either $|f(e_1)| > |f(e_2 + n\, e_1)|$ or $f(e_1)f(e_2 + n\, e_1) < 0$ and replace e_2 by $e_2 + n\, e_1$, else if $|f(e_1)| > |f(e_2)|$, then let n be the first positive integer satisfying either $|f(e_2)| > |f(e_1 + n\, e_2)|$ or $f(e_2)f(e_1 + n\, e_2) < 0$ and replace e_1 by $e_1 + n\, e_2$ and go to step 1.

There always exists such an integer n as we have chosen a (together with c) and b to have opposite signs. The procedure terminates, because at each step values of the form at bases e_1 and e_2 decrease in absolute value. Let us also remark that in order to find the integer n, one does not have to compute many values of the form f. In fact, in case $|f(e_1)| > |f(e_2)|$, n is the first positive integer satisfying the inequality $|f(e_2)\, n^2 + b\, n + f(e_1)| < |f(e_2)|$; where b satisfies $b^2 - 4\, f(e_1)\, f(e_2) = \Delta$. Similarly, if $|f(e_2)| > |f(e_1)|$, then n is chosen as the smallest positive integer satisfying the inequality $|f(e_1)\, n^2 + b\, n + f(e_2)| < |f(e_1)|$.

Example 3.3. Consider the form $f = (-3367, 3956, -1162)$. $|f(1, 0)| = 3367 > 1162 = |f(0, 1)|$. In the first reduction step we have $n = 1$ and $e_1 = (1, 0)$ is replaced by $(1, 1)$. In the second reduction step we have $|f(1, 1)| = 573 < 1162 = |f(0, 1)|$. In this case, $|f(0, 1) + 1\,(1, 1))| < |f(0, 1)|$, i.e. $n = 1$ and e_2 is replaced by $(1, 2)$. In the next two steps, we have $n = 2$ and in the end of these two steps, $e_1 = (3, 5)$ and $e_2 = (7, 12)$. In the final reduction step, e_1 becomes $(10, 17)$, reduction matrix is $\begin{pmatrix} 10 & 7 \\ 17 & 12 \end{pmatrix}$, and the resulting form is $M \cdot f = (2, 8, -7)$, see Figure 9.

One of the outputs of this algorithm is the matrix $M = (e_1|e_2)$ if it ends with (1.) and $M\,L$ if it ends with (2.) which is, as a path on the çark (or a word in S, L and L^2) a semi-reduction path, i.e. a path on the çark from a given non-spinal form to a spinal form.

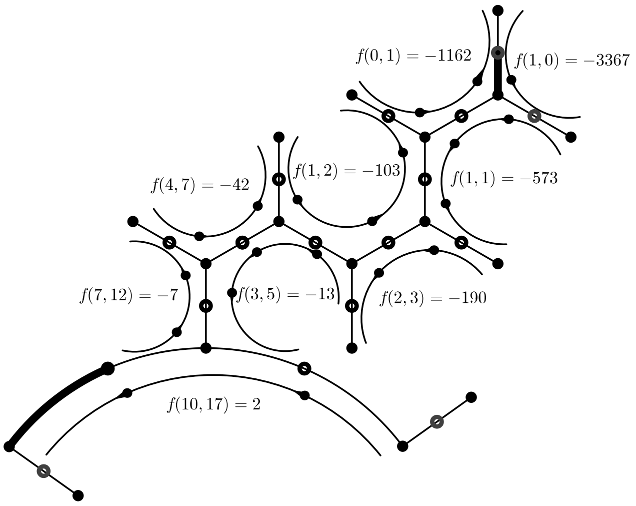

Figure 9. The form $f = (-3367, 3956, -1162) \in [(-7, 8, 2)]$ is equivalent to the spinal (in fact, reduced) form $f_o = (2, 8, -7)$ (bold edge on the spine)

Remark 3.4. The (semi-)reduction path produced by our algorithm in Example 3.3 is $(L^2, L^2S, \cdots, L^2SLS(L^2S^2)(LS^2)L^2S)$. This sequence of edges is a path because there are no cancellations in the final word. On the other hand, Gauss' algorithm produces $(L^2, L^2S, \cdots (L^2S)^2(L^2S)^3S(LS)^3SLS)$, which has backtracks.

4. Representation problem of binary quadratic forms

Given an integer N and a binary quadratic form $f = (a, b, c)$, the representation problem asks whether there are integers X and Y satisfying the equality $f(X, Y) = N$. For positive definite forms, the number of solution pairs, if any exists, is finite, see [1, §1.2.3]. In case of indefinite forms, the problem immediately reduces to solving the problem in the case of primitive forms. Note also that in this case if there is one solution, then there are infinitely many because the group of automorphisms of an indefinite binary quadratic form is infinite.

4.1. Labeled çarks. Let f be an indefinite binary quadratic form, let $Ç_f$ be the corresponding oriented, base-pointed çark. Starting once again with the oriented basis $\mathcal{B} = \{e_1 = (1,0), e_2 = (0,1)\}$, we may associate the value of f at these oriented bases as labels of the two faces neighbour faces to this edge. However, if $f = (a,b,c)$ is contained in a face, call φ, then so is $S \cdot f = (c,-b,a)$. To avoid this ambiguity, given an edge of $Ç_f$, from the set $\{f, S \cdot f\}$ we choose the one in which the product of its first component with second component is negative as a representative, i.e. we pick f if $ab < 0$, otherwise we pick $S \cdot f$, and call this the *labeling representative*, denoted by $f_{label} = (a_{label}, b_{label}, c_{label})$. Remark that if this pair is not on the spine then the labeling representative is the one closer to the spine among the two. We now assume that every edge not on the spine is oriented towards the spine of the çark. Given an edge, f', on $Ç_f$, we label the face on the left of f by $f(1,0) = a_{label}$ and the face on the right of f by $f(0,1) = c_{label}$. Note that via the method described, *all* faces of the çark, $Ç_f$, receive a unique label and further inner faces receive positive labels whereas outer faces receive negative labels.

Proposition 4.1. *There is a one-to-one correspondence between the set of values attained by a form f and the set of labels of the faces of the corresponding çark, $Ç_f$.*

Proof. Recall that an integer N is attained by a binary quadratic form if and only if there is form, f_N, in the class of the form in consideration with $f_N = (N,b,c)$. But each such N appears as a label in the labeled çark $Ç_{f_N}$, hence the result follows. □

4.2. Solving the minimum problem. Let us start this part with the following:

Theorem 4.2. *Given an indefinite binary quadratic form f let $Ç_f$ be the corresponding çark. Then the minimum of the absolute value of labels occurs at spinal faces of $Ç$.*

To prove, let us discuss a phenomenon that we call arithmetic progression on the labels of a çark: let $f = (a,b,c)$ be a binary quadratic form which corresponds to an edge of $Ç_f$ not on the spine. Without loss of generality, we suppose $f = f_{label}$. From f we may compute the labels of the remaining faces in the following manner: we label the two faces containing the edges $\{f, S \cdot f\}$ as described in the previous section and we associate $-b$ to the two edges $\{f, S \cdot f\}$. Now there

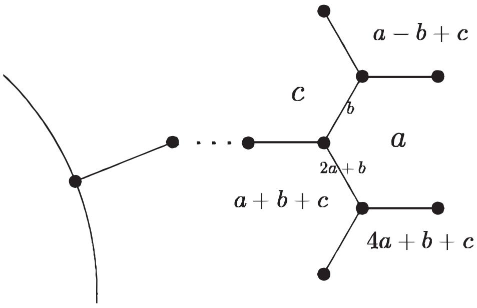

Figure 10. Arithmetic progression on the labels of a çark

are exactly two faces having distinct levels, say $|n_1| < |n_2|$ which have non-empty intersection with the two labeled faces. The label of the face having level n_2 is equal to $a-b+c$ and the label of the face of level n_1 is $a+b+c$, see Figure 10. In particular, taking into account the fact that the edges on the spine correspond to forms $g=(a',b',c')$ with $a'c'<0$, we conclude that as one moves away from the spine, the absolute value of the labels increase.

Remark also that if we label the pair of edges represented by $\{f, S\cdot f\}$ with b, then we also see that the labels of the edges also decrease as we move in the direction of the spine. As a result of the inequality satisfied by reduced forms, we conclude that labels of reduced edges are *local minimum*, that is for a reduced edge e_r, there is a positive integer n so that all edges whose distance to e_r is smaller than n has larger label than the label of e_r.

Let us state two immediate consequences of Theorem 4.2:

Corollary 4.3. *Fix an even integer n. Then the minimum label among the set $F_n(Ç)$ occurs at one of the faces which has non-empty intersection with a reduced face.*

In particular, setting $n=0$ we obtain the solution to the minimum problem, which asks the smallest element of the set $\{|f(X,Y)|\colon (X,Y)\in\mathbf{Z}^2\}$:

Corollary 4.4. *The minimum of a form f occurs at one of the reduced faces. More precisely, the smallest possible value of a form is the same as the smallest possible value of the set $\{|f'(0,1)| f' \in [f]$ and f' is reduced$\} \sqcup \{|f(1,0)| f' \in [f]$ and f' is reduced$\}$.*

In fact, more can be said about the minimal problem. For this let us give the following:

Definition 4.5. Given a primitive indefinite binary quadratic form f, a pair of *non-spinal* forms $\{f_o, S\cdot f_o\}$ of $Ç_f$ is called a *minimal edge* if the product of the labels of two faces which contains the pair $\{f_o, S\cdot f_o\}$ is smallest among all such products. The corresponding labeling representative is called a *minimal form.* If the two faces neighbour to a minimal edge are inner (resp. outer), then the corresponding pair/form is called an inner (resp. outer) minimal edge/form.

We note that inner and outer minimal edges may not be unique. For instance, any form whose çark has precisely two inner (resp. outer) faces possesses two minimal inner (resp. outer) minimal forms. The çark corresponding to the form $(2, 8, -7)$ has two outer minimal forms: one being the form $SL\cdot(3, 6, -7)$ and the other being $SL^2\cdot(-7, 6, 3)$. Both neighbouring faces are labeled -7 and -10.

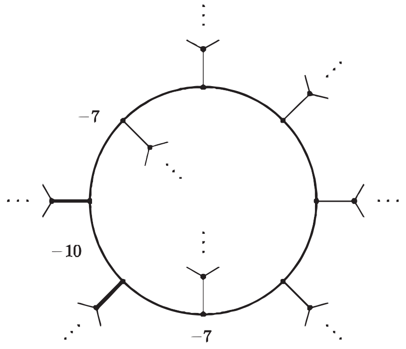

Figure 11. Outer minimal edges of the çark $(2, 8, -7)$

By definition of a minimal edge, at least one of the faces containing a minimal edge is reduced. More precisely, an inner minimal form must be contained in a reduced face whose label is smallest aming all reduced inner faces, and, similarly, an outer minimal form has to be contained in a face whose label is largest among all the outer reduced faces; as a result of Theorem 4.2 and the arithmetic progression on faces. Therefore, if f_o is an outer minimal edge, then either $LS\cdot f_o$ or $L^2S\cdot f_o$ (or both) is reduced. Analogously, if f_o is an inner minimal form, then either $L\cdot f_o$ or $L^2\cdot f_o$ (or both) is reduced. A minimal edge is contained in two faces. Among these, the face whose label's absolute value is smaller is called a *minimal face.* And from Corollary 4.3 we conclude that the smallest value has to intersect with a minimal face.

Corollary 4.6. *The minimum of a form is either equal to* $|((LS)\cdot f_o)(1,0)|$ *or* $\left|\left((L^2S)\cdot f_o\right)(1,0)\right|$*; where* f_o *is an outer minimal edge, or equal to* $((L)\cdot f_o)(0,1)$ *or* $\left((L^2)\cdot f_o\right)(1,0)$*; where* f_o *is an inner minimal edge.*

Remark 4.7. The Markoff number of an indefinite binary quadratic form is defined as

$$\mu(f) = \frac{\sqrt{\Delta}}{m(f)};$$

where $m(f)$ is the smallest element of the set of absolute value of values of the form $f = (a, b, c)$. As a result of the algorithm, given above solving the minimum problem, we also obtain a solution to determining the Markoff value of an indefinite binary quadratic form.

4.3. Solving the representation problem. A minimal face has to be reduced. Thus, of the two infinite paths, φ_l and φ_r of a minimal face φ, exactly one of them contains the minimal form $f_o = (a_o, b_o, c_o)$, say φ_l. Then by arithmetic progression, we see that the smallest of the set of absolute values of level n faces is the label on the non-reduced level n face containing the edge $(SL)^{|n|} \cdot f_o$. Analogously, if f_o is contained in φ_r, then such a face must contain the edge $(SL^2)^{|n|} \cdot f_o$. Let us summarize:

Theorem 4.8. *Given a binary quadratic form $f = (a, b, c)$ let f_o denote a corresponding (inner/outer) minimal form. Then the smallest value of the absolute value of labels of level n faces, denoted m_n, is:*

- $a_o - |\frac{n}{2}|b_o + (\frac{n}{2})^2 c_o$, *if f_o is contained in φ_l*
- $(\frac{n}{2})^2 a_o - |\frac{n}{2}|b_o + c_o$, *if f_o is contained in φ_r*

Now, given an equation

$$aX^2 + bXY + cY^2 = N \tag{1}$$

with $f = (a, b, c)$ being primitive indefinite binary quadratic form, in order to find a solution in $\mathbf{Z} \times \mathbf{Z}$, to the equation, we first find an inner minimal form if $N > 0$, and outer if $N < 0$. For this, one must find all reduced forms in the class $[f]$, or, equivalently, labels of all reduced faces. Then, by Theorem 4.1 and by Theorem 4.8 we conclude that it is enough to find n with $N \leq m_n$. Then there is an integral solution to Equation 1 if and only if there is a face labeled N in the labels of faces in the set

- $\bigsqcup_{0}^{n} F_n(\text{Ç}_f)$, if $N > 0$
- $\bigsqcup_{n}^{0} F_n(\text{Ç}_f)$, if $N < 0$

Example 4.9. Consider the equation $-3367X^2 + 3956XY - 1162Y^2 = -27$ and set $f = (-3367, 3956, -1162)$. Recall that the çark corresponding to f is given by $[-4, +1, -2, +1]$. In particular, there are four reduced forms in $[f]$: $(-7, 8, 2)$, $(3, 6, -7)$, $(-7, 6, 3)$ and $(2, 8, -7)$, in counterclockwise order. From this we immediately conclude that the minimal value attained by f is equal to 2. The forms $(-10, 20, -7)$ and $(-7, 20, -10)$ are the two minimal forms in $[f]$. So the minimal value of a level -4 face is -78. Now, labels of level -2 faces are -42, -58, -58, -42, -37, -37 and level 0 faces with negative label are -7, -13, -15, -13, -7, -10 in counterclockwise order. Therefore, given equation has no solution in integers. The equation $-3367X^2 + 3956XY - 1162Y^2 = -7$ on the other hand can be solved with solution $(7, 12)$, as also seen in Figure 9. Using the automorphism of f, namely $W = \begin{pmatrix} -3945 & 2324 \\ -6734 & 3967 \end{pmatrix}$, we obtain that any element of the set $\{W^n(7\ 12)^t \mid n \in \mathbf{Z}\}$ is a solution.

References

[1] J. Buchmann and U. Vollmer, Binary Quadratic Forms: An Algorithmic Approach, Algorithms and Computation in Mathematics, Vol. **20**, *Springer*, *Berlin*, 2007.

[2] H. Cohen, A Course in Computational Algebraic Number Theory, Graduate Texts in Mathematics, Vol. **138**, *Springer-Verlag*, *Berlin*, 1993.

[3] J. H. Conway, The Sensual (Quadratic) Form, Carus Mathematical Monographs, Vol. **26**, *Mathematical Association of America*, *Washington, DC*, 1997.

[4] C. F. Gauss, Disquisitiones Arithmeticae, *Yale University Press*, *New Haven, Conn.*, 1966.

[5] R. Halin, Über unendliche Wege in Graphen., *Math. Ann.* **157** (1964), 125–137.

[6] Team InfoMod, *Sunburst version 1.0*, 2015, math.gsu.edu.tr/azeytin/infomod/sunburst.

[7] The PARI Group, Bordeaux, *PARI/GP, version* 2.5.0, 2012, http://pari.math.u-bordeaux.fr/.

[8] A. M. Uludağ and A. Zeytin, A panorama of the fundamental group of the modular orbifold, In: Handbook of Teichmüller Theory, Volume VI., *European Mathematical Society (EMS)*, *Zürich*, 2016 (*to appear*).

[9] A. M. Uludağ, A. Zeytin and M. Durmuş, Binary quadratic forms as dessins, *Journal de Théorie des Nombres de Bordeaux* (*to appear*).

[10] D. B. Zagier, Zetafunktionen und quadratische Körper. Eine Einführung in die höhere Zahlentheorie, *Springer-Verlag*, *Berlin–New York*, 1981.

AYBERK ZEYTIN
DEPARTMENT OF MATHEMATICS
GALATASARAY UNIVERSITY
ÇIRAĞAN CAD. NO. 36
BEŞIKTAŞ, 34357, İSTANBUL
TURKEY

E-mail: azeytin@gsu.edu.tr
URL: http://math.gsu.edu.tr/azeytin/

(Received August 26, 2015; revised January 14, 2016)