Smallest Reduction Matrix of Binary Quadratic Forms
And Cryptographic Applications

Aurore Bernard\textsuperscript{1} and Nicolas Gama\textsuperscript{2}

\textsuperscript{1} XLIM, Limoges, France
\hspace{1em} aurore.bernard@xlim.fr
\textsuperscript{2} GREYC Ensicaen, Caen, France
\hspace{1em} nicolas.gama@greyc.ensicaen.fr

Abstract. We present a variant of the Lagrange-Gauss reduction of quadratic forms designed to minimize the norm of the reduction matrix within a quadratic complexity. The matrix computed by our algorithm on the input $f$ has norm $O\left(\frac{|f|^{1/2}}{\Delta_f^{1/4}}\right)$, which is the square root of the best previously known bounds using classical algorithms. This new bound allows us to fully prove the heuristic lattice based attack against NICE Cryptosystems, which consists in factoring a particular subclass of integers of the form $pq^2$. In the process, we set up a homogeneous variant of Boneh-Durfee-HowgraveGraham’s algorithm which finds small rational roots of a polynomial modulo unknown divisors. Such algorithm can also be used to speed-up factorization of $pq^r$ for large $r$.

1 Introduction

Binary quadratic forms appeared progressively in the 17-th century, when Descartes and Fermat first introduced the concept of coordinates as a tool to algebraically solve geometric problems. Those forms have wide applications in mathematics and physics, especially in geometry, numerical analysis or algebraic topology. A binary quadratic form is a homogeneous polynomial of degree two in two variables, which can be viewed as the Cartesian equation of a surface $f(x, y) = ax^2 + bxy + cy^2$ on a given basis of $\mathbb{R}^2$. Of course, this equation varies with the basis of expression, and it is natural to define an equivalence relation to regroup all these possible equations into classes. Over the real field, there are six classes corresponding to the Sylvester’s signatures. They can be distinguished by the sign of the discriminant $\Delta_f = b^2 - 4ac$, and the sign of $a + c$. Forms of strictly negative discriminant (imaginary forms) have a unique zero at the origin, which is also their unique local and global extremum. Forms of strictly positive discriminant (real forms) represent a saddle-shape.

Meanwhile, quadratic forms were also used over the integer ring by Fermat, Lagrange and Gauss to solve long standing problems from number theory. This time, binary quadratic forms are equations with integer coefficients of discrete
scatter-plots on a given lattice basis of $\mathbb{Z}^2$. One defines a similar equivalence relation by base change, except that transformation matrices are now unimodular, and that they preserve the value of the discriminant. Problems related to this equivalence are more complicated than on the real field: for instance, in both real and imaginary cases, we do not know any polynomial way to compute the number of equivalence classes of a given discriminant. Deciding the equivalence of two forms is easy in the imaginary case, where each class contains a unique reduced representative computable in polynomial time. However, the problem is hard in the real case, where there are, depending on the notion of reduction, either an exponential number of polynomially computable reduced representatives, or a few representatives computable in exponential time.

A reduction algorithm takes as input a quadratic form and outputs a reduced form and the reduction matrix, which is a unimodular base-change matrix used to obtain this form. The most famous polynomial time reduction algorithms are Lagrange algorithm [15] (1773) commonly known as "Gauss reduction" algorithm [11] (1801). In [14] (1980), Lagarias modified the Gauss reduction algorithm for make it more efficient. This algorithm is the one used in practice, and which we refer as the Gauss reduction algorithm, or Classical Gauss, if we need to differentiate it from new flavors which we propose.

The cryptanalysis of [6] shows experimental evidences that the small size of reduction matrices have important applications to the factorization of some large numbers used in public key cryptosystems, especially those of the NICE cryptosystems (see [12,13]). However the best currently known upper-bounds on the size of reduction matrices [14,1] are by an order too large, and keep all these results on the factorization heuristic. In this paper, we specially design an efficient variant of the Gauss reduction algorithm to minimize the size of transformation matrix, and we prove constructive upper-bounds which are tight both in the worst case and in the average case. These bounds, combined with an improvement of the methods of [6], allows us to prove all the above mentioned heuristics of on the factorization of integers from the NICE cryptosystems.

2 Preliminaries and Notation

In this section we recall some definitions and properties concerning binary quadratic forms. For a more detailed account of the theory see [5,4,9]. Then, we summarize some results on the norm of a matrix.

**Quadratic Forms.** A binary quadratic form $f$ is a homogeneous polynomial of degree two in two variables $f(x, y) = ax^2 + bxy + cy^2$ with $(a, b, c) \in \mathbb{Z}^3$ which we abbreviate as $f = (a, b, c)$. Throughout this paper the word form will be used in the sense of binary quadratic form. It is said primitive when $\gcd(a, b, c) = 1$. The discriminant of $f$ is $\Delta_f = b^2 - 4ac$. A discriminant $\Delta_f$ is called fundamental if all the forms of discriminant $\Delta_f$ are necessarily primitive: for example, it is the case of all odd and square-free integers. The set of all primitive forms of discriminant $\Delta_f$ is denoted $\mathcal{F}_{\Delta_f}$. We impose that the discriminant is not a perfect square then $a$ and $c$ are always non-zero. The form $f$ can be factored as
\( f(x, y) = a(x - y\zeta_1^r)(x - y\zeta_1^r) \) where \( \zeta_1^r \) and \( \zeta_1^r \) are the complex roots of the univariate polynomial \( f(x, 1) \) which we call the affine representation of \( f \). When \( \Delta_f > 0 \), each root of \( f \) live in \( \mathbb{R} \setminus \mathbb{Q} \) and the form is real. In this case, \( \zeta_1^{-1} \) will denote the smallest root and \( \zeta_1^+ \) the largest one. When \( \Delta_f < 0 \), the roots are in \( \mathbb{C} \setminus \mathbb{R} \) and the form is imaginary. We note \( \lambda(f) = \min \{|f(x, y)| : (x, y) \in \mathbb{Z}^2 \setminus (0, 0)\} \) the first minimum of \( f \).

**Composition Action.** We note \( M^t \) the transpose of a matrix \( M \). The polar representation of \( f \) is the symmetric matrix \( \left( \begin{array}{cc} a & \frac{b}{2} \\ \frac{b}{2} & c \end{array} \right) \) of determinant \(-\Delta_f/4\). Let \( M = \left( \begin{array}{cc} \alpha & \beta \\ \gamma & \delta \end{array} \right) \in M_2(\mathbb{Z}) \) be a \( 2 \times 2 \) matrix with integer entries which we often abbreviate as \((\alpha, \beta; \gamma, \delta)\). We note \( \text{Id} \) the identity matrix of \( M_2(\mathbb{Z}) \). The composition action of \( M \) on \( f \) is defined as the form \( g(x, y) = f(ax + \beta y, \gamma x + \delta y) \) and it is noted \( g = f.M \). The coefficients of \( g \) are \( g = (f(\alpha, \gamma), b(\alpha \delta + \gamma \beta) + 2(\alpha \alpha \beta + \gamma \gamma \delta), f(\beta, \delta)) \). We remark that for each root \( \zeta_g \) of \( g \), \( \left( \frac{\alpha \zeta_g + \beta}{\gamma \zeta_g + \delta} \right) \) is a root of \( f \). Finally, the polar representation of \( g \) is \( M^t f.M \) which implies that \( \Delta_g = \det(M)^2 \Delta_f \).

**Group action.** Let \( \text{GL}_2(\mathbb{Z}) \) be the general linear group of matrices in \( M_2(\mathbb{Z}) \) which are invertible and its subgroup \( \text{SL}_2(\mathbb{Z}) \) the special linear group of matrices which have a determinant equal to one. The action defined with either \( \text{GL}_2(\mathbb{Z}) \) or \( \text{SL}_2(\mathbb{Z}) \) on the set of primitive forms \( \mathcal{F}_\Delta \) of a given discriminant is a (right) group action. Two forms \( f \) and \( g \) are equivalent if they belong to the same \( \text{SL}_2(\mathbb{Z}) \)-orbit. In this case we note \( f \sim g \). We define \( \text{Aut}^+(f) \) the group of automorphisms of the form \( f \in \mathcal{F}_\Delta \) as \( \{M \in \text{SL}_2(\mathbb{Z}) : \text{trace}(M) > 0 \text{ and } f.M = f \} \). The set of all automorphisms of \( f \) is \( \pm \text{Aut}^+(f) \). The group \( \text{Aut}^+(f) \) is known to be cyclic, and we call its generator the fundamental automorphism of \( f \). The largest eigenvalue of the fundamental automorphism of \( f \) is the fundamental unit. It only depends on the discriminant \( \Delta_f \), and will be denoted \( \epsilon_{\Delta_f} \).

**Three special transformations.** We define the symmetry \( S = \left( \begin{array}{cc} 1 & 0 \\ 0 & -1 \end{array} \right) \), the exchange \( E = \left( \begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array} \right) \) and the translation by an integer \( T(h) = \left( \begin{array}{cc} 1 & h \\ 0 & 1 \end{array} \right) \). They are three (linear) transformations of \( \text{GL}_2(\mathbb{Z}) \). All matrices in \( \text{GL}_2(\mathbb{Z}) \) can be written as a product of powers of these three transformations and \( \text{SL}_2(\mathbb{Z}) \) is generated by the product \( E.S \) and \( T(1) \). The action of these transformations on \( f \) are \( f.S = (a, -b, c) \), \( f.E = (c, b, a) \) \( f.T(h) = (a, b + 2ah, f(h)) \). Note the important fact: the roots of \( f.S \) are the opposite of the roots of \( f \) and the roots of \( f.E \) are the inverse of the roots of \( f \), and that \( T(h) \) subtracts \( h \) to each roots of \( f \).

**Norms of matrices and forms.** Let \( M = (\alpha, \beta; \gamma, \delta) \) be a matrix in \( M_2(\mathbb{Z}) \). The Euclidean norm is \( \|M\|_2 = \sqrt{\alpha^2 + \beta^2 + \gamma^2 + \delta^2} \), and the maximum norm is \( \|M\| = \max(|\alpha|, |\beta|, |\gamma|, |\delta|) \). The norm \( \|M\| = \sup_{\|v\|_2 = 1} (\|M.v\|_2) \) is the induced Euclidean norm, which is also the square root of the largest eigenvalue of \( M^t.M \). All the norms are equivalent: \( \|M\| \leq \|M\|_2 \leq \|M\|_2 \leq 2\|M\| \).
Additionally, the induced norm is sub-multiplicative: if \( N \in \mathcal{M}_2(\mathbb{Z}) \) then \( \|MN\| \leq \|M\| \cdot \|N\| \) and \( \|\text{Id}\| = 1 \), and it is lower-bounded by the spectral radius \( \rho(M) \), which is the supremum among the absolute values of the eigenvalues of \( M \). By extension, we define the norms \( \|f\|, \|f\|_2 \) and \( \|f\| \) of a form as the corresponding norm of its polar representation.

### 3 A New Reduction Algorithm for Real Quadratic Forms

A form \( f = (a, b, c) \) is reduced if it satisfies two conditions simultaneously: a normalization condition, which defines the choice of the representative of \( b \mod 2a \), and a reduction condition, which often upper-bounds the size of \( |a| \) (or \( |c| \)). In the imaginary case, these conditions are very natural: a form is normal if and only if \( b \in [a] - |a|, |a| \) is minimal, and is reduced if additionally, \( |a| \) is the minimum \( \lambda(f) \). A single translation is needed to normalize any form. However, the reduction condition takes more steps to be achieved. The classical Gauss reduction reduces a form by successive swaps \( SE \) and normalizations \( T([-b/2a]) \) (see [1]) until \( f \) is reduced. The Gauss reduction algorithm operates in quadratic time (see [1,21,18]). For each form \( f \) of discriminant \( \Delta_f < -4 \), there exists a unique reduced form \( g \) in each equivalence class, and a unique reduction matrix \( M \in \text{SL}_2(\mathbb{Z}) \) such that \( f.M = g \). In this case \( \text{Aut}^+(f) = \{\text{Id}\} \).

In the real case (\( \Delta_f > 0 \)), the previous reduction conditions applied on \( f = (a, b, c) \) are too restrictive, since the smallest integers \( (\alpha, \beta) \neq (0, 0) \) such that \( |a| \geq \Delta_f \) and \( b \in [\sqrt{\Delta_f - 2|a|}, \sqrt{\Delta_f}] \) when \( |a| < \Delta_f \) and \( f \) is classically reduced if additionally, \( |\sqrt{\Delta_f - 2|a|}| < b < \sqrt{\Delta_f} \). It is known that only a finite subset of forms of discriminant \( \Delta_f \) are classically-reduced, and that they form a reduced cycle in each class. The Real-Gauss reduction algorithm, which uses the classical normalization, finds a reduced form equivalent to its input in quadratic time (see [1]).

In this paper, given a normalized form \( f \), we will bound the coefficients of the smallest reduction matrix \( M = (\alpha, \beta; \gamma, \delta) \) such that \( g = f.M = (a_g, b_g, c_g) \) is reduced. The case of imaginary forms is eased by the uniqueness of the reduction matrix. Lemma 5.6.1 in [1] give us that \( \|M\| \leq 2 \cdot \frac{\max\{|a|, |c|\}}{\sqrt{|\Delta_f|}} \). We improve this upper-bound with the following theorem:

**Theorem 1 (Imaginary Bound).** Let \( f = (a, b, c) \) be a normalized imaginary form of discriminant \( \Delta_f < 0 \), and \( M = (\alpha, \beta; \gamma, \delta) \) the reduction matrix such that \( g = f.M = (a_g, b_g, c_g) \), \( M \) satisfies these two upper-bounds:

1) \( \|M\| \leq \frac{2}{\sqrt{3}} \cdot \sqrt{\frac{|a|}{|a_g|}} \)

2) \( |\alpha\beta\gamma\delta|^{1/4} \leq |\gamma\delta|^{1/2} \leq \frac{2}{3^{1/4}} \cdot \left( \frac{|ac|}{|\Delta_f|^{1/4}} \right)^{1/4} \).
Proof. One has $|a_g| = |f(\alpha, \gamma)| = |a| \gamma^2 \left(\left(\frac{\alpha}{\gamma} + \frac{b}{2a}\right)^2 + \frac{|a_d|}{4a^2}\right)$, which can be lower-bounded by $\frac{|a_d|}{4a^2} \gamma^2$. It follows that $\gamma^2 \leq \frac{4|a_a|}{|a_d|}$, and similarly $\delta^2 \leq \frac{4|c_a|}{|a_d|}$. Therefore $|\gamma\delta| \leq \frac{4\sqrt{|a|}}{\sqrt[4]{|a_d|}}$. The first inequality comes from $3|a_g c_g| \leq |\Delta_f|$, because $g$ is reduced. Unless the transformation is trivial (Id or $SE$), the normalization condition induces the inequalities $|\alpha| \leq |\gamma|$ and $|\beta| \leq |\delta|$, which proves $|\alpha\beta\gamma\delta|^{1/4} \leq |\gamma\delta|^{1/2}$.

Thus, the norm of the reduction matrix is in fact basically in $O \left(\sqrt{\|f\|/\sqrt{\Delta_f}}\right)$.

In the real case however, this proof would not apply directly, because the term $\left(\left(\frac{\alpha}{\gamma} + \frac{b}{2a}\right)^2 - \frac{|a_d|}{4a^2}\right)$ can be exponentially close to 0. The problem is that in the real case, each reduced cycle contains a large (often exponential) number of equivalent reduced forms, and some of them are exponentially far from $f$. A constructive approach is needed to build a polynomial reduction matrix. The analysis of the Gauss reduction algorithm in [1,14] basically proves that the norm of the computed reduction matrix is bounded by $O(\|f\|)$. In this paper, we study a variant of this algorithm which finds a reduction matrix of norm $O \left(\sqrt{\|f\|/\sqrt{\Delta_f}}\right)$ and we verify that it is tight even in the average case.

We define new relaxed notions of reduction and normalization, and express them according to the roots of the forms, which is more intuitive than the classical conditions on the coefficients:

Definition 1. A real binary quadratic form $f$ is:

- primary normalized if $0 < \zeta_f^+ < 1$ and primary reduced if also $\zeta_f^- < -1$
- secondary normalized if $-1 < \zeta_f^- < 0$ and secondary reduced if also $1 < \zeta_f^+$.

Finally $f$ is largely reduced if it is either primary or secondary reduced.

Both primary and secondary notions are exchanged by the action of $S$, which negates the roots. Furthermore, primary and secondary reductions are exchanged by $E$, which inverts the roots. As usual, primary and secondary normalization can always be achieved by the action of some $T(h)$. Note that a classically normalized form, which has by definition at least one root in the interval $[-1, 1]$, is either primary or secondary normalized. Similarly, a classically reduced form $(a, b, c)$ is a largely-reduced form satisfying $b > 0$, which can again be ensured by the action of $S$. Our main contribution is to solve the following problems, which are equivalent.

Lemma 1. The two problems are equivalent:

1. Smallest $SL_2(\mathbb{Z})$ matrix Given a classically-normalized real form $f$, find $\mathcal{M} \in SL_2(\mathbb{Z})$ such that $f \mathcal{M}$ is classically-reduced and $\|\mathcal{M}\|$ is minimal.
2. Smallest $GL_2(\mathbb{Z})$ matrix Given a primary-normalized real form $f$, find $\mathcal{M} \in GL_2(\mathbb{Z})$ such that $f \mathcal{M}$ is largely-reduced and $\|\mathcal{M}\|$ is minimal.
Proof. From a solution \( \mathcal{M} \in \text{GL}_2(\mathbb{Z}) \) of Problem 2, one deduces a solution of Problem 1 by left-multiplication by \( \text{Id} \) or \( S \) to make the normalization of the input correspond, followed by a right-multiplication by \( \text{Id} \) or \( S \) to force the coefficient \( b \) of the reduced form to be positive, followed by a right-multiplication by \( \text{Id} \) or \( E \) so that the determinant is +1. The reduction of Problem 2 to Problem 1 is similar. Since \( \text{Id}, S \) and \( E \) are permutation matrices, they do not modify these norms \( \| \cdot \| \) or \( | \cdot | \). Remark that, reducing a problem to the other also preserves the absolute value of the product of the coefficients in each row of the reduction matrices.

Lemma 1 motivates the search of a reduction algorithm solving the less restrictive Problem 2, since we can use the above permutation matrices to return to classical notions in \( \text{SL}_2(\mathbb{Z}) \).

3.1 Algorithm and Analysis

Let \( f \) be a real form. We define the two integers \( h^+_f \) and \( h^-_f \) as \( h^+_f = \left\lfloor \zeta^+_f \right\rfloor \) and \( h^-_f = \left\lfloor \zeta^-_f \right\rfloor \). It is easy to show that \( h^+_f \) and \( h^-_f \) are respectively the unique integers such that \( f.T(h^+_f) \) is primary-normalized, and \( f.T(h^-_f) \) is secondary-normalized. Among the two integers \( h^-_f, h^+_f \) the one of smallest absolute value is noted \( h(f) \): that is \( h(f) = h^+_f \) if \( |h^+_f| < |h^-_f| \), and \( h(f) = h^-_f \) otherwise. In other words, \( h(f) \) is the shortest normalization of \( f \). As a comparison, there is only a single integer \( \nu_f \) in the classical case such that \( f.T(\nu_f) \) is classically-normalized, \( \nu_f \) being one of the integers \( h^-_f, h^+_f \) but not necessarily the one with the smallest absolute value. Our reduction algorithm, is a variant of the Gauss reduction which operates in \( \text{GL}_2(\mathbb{Z}) \). It alternates exchange \( E \) and the shortest normalization \( T(h(f)) \) at each loop, and terminates on a largely-reduced form. As we will see later, any kind of normalization by \( h^-_f \) or \( h^+_f \) would make a reduction algorithm terminate\(^1\), but the choice of the shortest normalization \( h(f) \) instead of the classical \( \nu_f \) (especially during the last steps) is the key element to minimize the reduction matrix. The main result of the section is the following theorem on the quality of the output of our algorithm, which is the real-case analogue of Theorem 1.

**Algorithm 1. RedGL2**

**Input:** \( f = (a, b, c) \) a primary-normalized form  
**Output:** \( f, \mathcal{M} \) a largely-reduced form and \( \mathcal{M} \in \text{GL}_2(\mathbb{Z}) \)

1: \( \mathcal{M} = \text{Id} \)  
2: while \( f \) not largely-reduced do  
3: \( f \leftarrow f.E \) and \( \mathcal{M} \leftarrow \mathcal{M}E \) \( \Rightarrow \text{Exchange step} \)  
4: \( f \leftarrow f.T(h(f)) \) and \( \mathcal{M} \leftarrow \mathcal{M}T(h(f)) \) \( \Rightarrow \text{Normalization step} \)  
5: end while  
6: return \( f \) and \( \mathcal{M} \)

\(^1\) The original Gauss algorithm of 1801 used actually the largest normalization at each step. The number of reduction steps is exponential on some entries. Lagarias introduced the classical normalization to obtain a quadratic complexity.
Theorem 2 (Real bound). Let \( f = (a, b, c) \) be a primary-normalized form of discriminant \( \Delta > 0 \). Given \( f \) as input, \( \text{RedGL2} \) terminates after at most 
\[
\left( \frac{\log |a|/\sqrt{\Delta}}{2 \log(\omega)} + 4 \right)
\]
iterations where \( \omega = \frac{1 + \sqrt{\Delta}}{2} \) is the gold number. Its output \( M = (\alpha, \beta \gamma, \delta) \) and \( f_r = f.M = (a_r, b_r, c_r) \) satisfies:

1) \( \| M \| \leq 4 \cdot \sqrt{|a|/\omega} \)

2) \(|a \beta \gamma \delta|^{1/2} \leq |\gamma \delta|^{1/2} \leq \sqrt{21} \cdot \sqrt{|a|/\sqrt{\Delta}}. \)

Before proving this theorem, we remark that the best known upper-bounds achieved by the classical Gauss algorithm under the same conditions (see theorem 4.4 of [1]) are \( \| M \| \leq |a| (1 + 1/\Delta) \) and \( |\gamma \delta|^{1/2} \leq (|a|/\sqrt{\Delta}) (1 + 1/\sqrt{\Delta}) \). They are basically the square of the upper-bounds of \( \text{RedGL2} \). Figure 1 and 2 illustrate respectively the families of forms \( F_n = (-n, b, 1) \) and \( G_n = (n, n, 1) \) with \( n \in \mathbb{N} \) and \( b = [2n^3 - 2/3] \), which are families of forms where the Gauss reduction algorithm outputs reduction matrices \( \sqrt{\Delta} \) times larger than our variant \( \text{RedGL2} \). Finally, note that a multiplicative triangular inequality on the norms of the polar representations of \( f = f_r . M^{-1} \) yields \( \sqrt{\| f \| \| f_r \|} \leq \sqrt{2} \| M \| \), which confirms the optimality of Theorem 2 in average. The analysis of Gauss reduction algorithm in [1] upper-bounds the number of iterations by \( \left( \frac{\log (|a|/\sqrt{\Delta})}{2 \log (2)} + 2 \right) \) reduction steps. Our upper-bound on the number of iterations of \( \text{RedGL2} \) is tight in the worst case, and is only by a multiplicative factor around 1.4 larger than the maximum number of iterations of the Gauss reduction algorithm. However the primary goal of \( \text{RedGL2} \) is the minimization of the reduction matrix.

3.2 Proof of Theorem 2

To prove Theorem 2, we first study the termination cases, characterized by the presence of integers between the roots of \( f.E \), and where the choice of the shortest normalization is of greatest importance. Eventually, we shall treat the general case and the complexity.

Termination cases. We first study the two cases where the algorithm terminates in a single step of reduction. The first one deals with normal form \( f \) containing exactly one integer between its roots. This is the only case where \( h_f^- = h_f^+ \), so all notions of normalizations (classical, primary, secondary, shortest) coincide.

Lemma 2. Let \( f = (a, b, c) \) be a real form satisfying \(-1 < \zeta_f^- < 0 < \zeta_f^+ < 1\), and \( h = h(f.E) \). The form \( f_r = f.ET(h) = (a_r, b_r, c_r) \) is largely-reduced, and its coefficients satisfy \( a_r = c, |c_r| \leq |a|, \) and \( h^2 |a_r| \leq |a| \).

Proof. The reduction matrix from \( f \) to \( f_r \) is \( ET(h) = (0, 1; 1, h) \). Consider the parabola \( p(x) = cx^2 + bx + a \) which is the affine representation of \( g = f.E \). Then we have \( h = h(g), \) and \( \zeta_g^- < h_g^- \leq -1 < 1 \leq h_g^+ < \zeta_g^+ \), \( c_r = p(h(g)) \) and \( p(0) = a \). By definition of \( h \) we have two cases: if \( -b/2c > 0 \) then we have \( h = h_g^- < 0 < -b/2c, \) else we have \( -b/2c < h = h_g^+ \). In both cases we
Let \( \nu_g \) be a real form satisfying 
\begin{equation}
\nu_g = \nu(h) = \left[ \zeta_g^+ \right],
\end{equation}
which can be \( O(\sqrt{\Delta}) \) smaller than the classical normalization \( \nu(g) = \left[ \zeta_g^- \right] \) in Gauss Algorithm. It is clear that \( \nu_g \) is the shortest normalization chosen by RedGL2.

This figure is the analogue for Lemma 3. In this case, the shortest normalization chosen by RedGL2 is \( h(g) = [\zeta_g^-] \), which can be \( O(\sqrt{\Delta}) \) smaller than the classical normalization in Gauss algorithm is \( \nu(g) = [\zeta_g^+] \). Inequality on the slopes of \( p \) before and after \( \zeta_g^- \) gives \( |\nu_g| \leq |a| \). Comparison of heights of the two rectangles on the same convex and decreasing branch of the parabola, gives \( |ch^2| \leq |a| \).

Theorem 2 holds in this termination case: the reduction matrix is \( M = (0, 1; 1, h) \). By Lemma 2, its norm satisfies \( \|M\| = h \leq \sqrt{|a|/|a_r|} \). Since \( f = f_r.M^{-1} \), its first coefficient is \( a = a_rh^2 - b_rh + c_r \), thus \( b_rh = -a + c_r + a_rh^2 \) and \( (b_rh)^2 - 4a_rh^2c_rh^2 = \Delta \cdot h^2 = a^2 + c_r^2 + a_r^2h^4 - 2ac_r - 2a_ch^2 - 2a_rch^2 \leq (|a| + |c_r| + |a_r|h^2)^2 \leq 9|a|^2 \), which proves the second point of Theorem 2.

The second case of single-step termination concerns normalized form \( f \) such that at least two integers lie between the roots of \( f.E \) (namely \( h_{f_r.E}^- < h_{f_r.E}^+ \)). We just write a proof for primary-normalized forms, but it can be easily extended to secondary-normalized forms.

**Lemma 3.** Let \( f = (a, b, c) \) be a real form satisfying \( 0 < \zeta_f^- < \zeta_f^+ < 1 \), and such that \( h_{f_{r.E}}^- < h_{f_{r.E}}^+ \). If \( h = h(f.E) \), then \( f_r = f.ET(h) = (a_r, b_r, c_r) \) is secondary-reduced, and its coefficients satisfy \( a_r = c, |c_r| \leq |a|, \) and \( h^2|a_r| \leq 4|a| \).

**Proof.** The proof of this lemma is also based on convexity inequalities. Let \( g = f.E \), of affine representation \( p(x) = cx^2 + bx + a \). Note that \( h = [\zeta_g^-] \geq 2 \). Again, one has \( p(0) = a, p(h) = c_r \). It follows from the definition that \( f_r \) is
secondary-reduced. The reduction matrix is $\mathcal{M} = (0, 1; 1, h)$, which proves $a_r = c$. Application of a convexity inequality (see Figure 2) on $p$ in the two intervals $[0; h - 1]$ and $[-\frac{b}{2a} - (h - 1); -\frac{b}{2a}]$ of same length yields $|a_r(h - 1)^2| \leq |a - p(h - 1)| \leq |a|$, therefore $|a_r|h^2 \leq 4|a_r|(h - 1)^2 \leq 4|a|$. Finally, another convexity inequality centered on $\zeta_g$ gives $\frac{p(0) - p(\zeta_g^-)}{a - \zeta_g} \leq \frac{p(h) - p(\zeta_g^+)}{h - \zeta_g}$, so $|a| = p(0) \geq \frac{\zeta_g^-}{h - \zeta_g} \cdot (-p(h)) \geq |c_r|$. □

Once again, Theorem 2 holds in this termination case, but this time, $\|\mathcal{M}\| = h \leq 2\sqrt{|a|/|a_r|}$ and $\Delta \cdot h^2 \leq (|a| + |c_r| + |a_r|h^2)^2 \leq (6|a|)^2$.

**General case.** We now prove the general case of Theorem 2. We call $f_i = (a_i, b_i, c_i)$ the successive values of $f$ at the beginning of the while loop of Algorithm 1, and $h_i = h(f_i, E)$. We suppose that the primary-normalized form $f_0$ does not have any integer between its roots (otherwise it would either already be reduced or as in Lemma 2). Thus $0 < \zeta_{f_0}^- < \zeta_{f_0}^+ < 1$. For each iteration $i$ in the loop, if there is at least one integer between the roots of $f_i, E$, we then set $m = i + 1$ and the algorithm reaches one of the two termination cases above. Otherwise the shortest normalization $h_i$ is the primary one $h_i = h_{f_i, E}^+ < h_{f_i, E}^-$. Thus $f_i$ is also primary-normalized and $0 < \zeta_{f_i}^- < \zeta_{f_i}^+ < 1$. Note that the distance between the roots strictly increases $|\zeta_{f_i}^- - \zeta_{f_i}^+| = |\zeta_{f_{i-1}}^- - \zeta_{f_{i-1}}^+| = |\zeta_{f_{i-1}}^- \times \zeta_{f_{i-1}}^+|^{-1}$, $|\zeta_{f_{i-1}}^- - \zeta_{f_{i-1}}^+| \geq |\zeta_{f_{i-1}}^- - \zeta_{f_{i-1}}^+|$. Such process can not hold forever, otherwise the integer sequence of the first coefficients $|a_i| = \sqrt{\Delta} |\zeta_{f_i}^- - \zeta_{f_i}^+|$ would be strictly decreasing. This proves the termination of the algorithm. The integer $m$ is the smallest index, such that $f_{m-1}, E$ contains at least one integer between its roots. The shortest normalization $h_{m-1} = h_{f_{m-1}, E}^+ \leq h_{f_{m-1}, E}^-$ is in this case secondary, and satisfies $h_{m-1} \geq 2$.

We eventually use the following lemma to conclude the proof of Theorem 2.

**Lemma 4.** Let $f = (a, b, c)$ and $g = (a_g, b_g, c_g)$ be two real forms and $\mathcal{M} = (\alpha, \beta, \gamma, \delta) \in GL_2(\mathbb{Z})$ such that $f \mathcal{M} = g$. If all the roots of $g$ are positive and $\gamma \geq 0$ and $\delta \geq 1$ then $|a_g|\delta^2 \leq |a|$.

**Proof.** If $\gamma = 0$, then $\mathcal{M}$ is triangular, so $|\alpha| = |\beta| = 1$ and $|a_g| = |a|$. We now suppose $\gamma > 0$. Let $\zeta_g$ be a root of $g$, then $\zeta_f = \frac{\alpha \zeta_g + \beta}{\gamma \zeta_g + \delta}$ is a root of $f$. We have $|\alpha/\gamma - \zeta_f| = 1/|\gamma^2 \zeta_g + \gamma \delta| < 1/\gamma \delta$ thanks to the positivity conditions. Since this bound holds for both roots of $f$, $|a_g| = \gamma^2 |a| \left| \alpha/\gamma - \zeta_f^+ \right| < |a|/\delta^2$. □

We continue the proof of Theorem 2 by applying this lemma to the main loop of RedGL2. Note that for each $i \in [1; m]$, the reduction matrix from $f_0$ to $f_i$ is

$$
\mathcal{M}_i = \begin{pmatrix} 0 & 1 \\ 1 & h_i \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & h_{i-1} \end{pmatrix} \ldots \begin{pmatrix} 0 & 1 \\ 1 & h_0 \end{pmatrix} = \begin{pmatrix} \alpha_i & \beta_i \\ \gamma_i & \delta_i \end{pmatrix}.
$$

(1)
Their coefficients are all positive, and satisfy these recurrence equalities for \( i \geq 2 \):
\[
\begin{align*}
\gamma_{i+1} &= \delta_i = h_{i-1} \delta_{i-1} + \delta_{i-2} \quad \text{and} \quad (\delta_0, \delta_1) = (1, h_1) \\
\alpha_{i+1} &= \beta_i = h_{i-1} \beta_{i-1} + \beta_{i-2} \quad \text{and} \quad (\beta_0, \beta_1) = (0, 1)
\end{align*}
\]

Since all the \((h_j)_{j=0..i}\) are greater than 1, it follows that \( \alpha_i \leq \min(\beta_i, \gamma_i) \leq \max(\beta_i, \gamma_i) \leq \delta_i \) and \( \|M_i\| = \delta_i \geq \omega^{i-2} \) by induction and comparison to the Fibonacci sequence \(^2\). Applying Lemma 4 on \( f_0 \) and \( f_{m-1} \) implies that \( \|M_{m-1}\|^2 \leq |a_0|/|a_{m-1}| \). At iteration \( m \), Lemma 4 can be applied to \( f_m, T(-1) \), which has positive roots and shares its first coefficient \( a_m \) with \( f_m \). The transformation matrix \( M_m T(-1) = M_{m-1} (0, 1; 1, h_{m-1} - 1) \) still satisfies the conditions of Lemma 4 because \( h_{m-1} \geq 2 \). We obtain \( \|M_m T(-1)\|^2 \leq |a_0|/|a_m| \), and finally \( \|M_m\|^2 \leq 4|a_0|/|a_m| \) after a backwards translation by \( T(1) \).

We already know that \( f_m \) is secondary-normalized and that the largest root of \( f_m \) is positive. There are two cases:

1. If the largest root of \( f_m \) is strictly greater than 1, then \( r = m, f_r \) is secondary-reduced, and the reduction matrix is \( M_m = (a_m, \beta_m; \gamma_m, \delta_m) \). One already has \( \|M_m\|^2 \leq 4|a_0|/|a_r| \). From \( f_0 = f_r M^{-1} \), we draw \( a_0 = a_r \delta_m^2 - b_r \delta_m \gamma_m + c_r \gamma_m^2 \), so \( \Delta \delta_m^2 \gamma_m^2 = (b_r \delta_m \gamma_m)^2 - 4a_r c_r \delta_m^2 \gamma_m^2 \leq (|a_0| + |a_r \delta_m^2| + |c_r \gamma_m^2|)^2 \). Since by construction \( \gamma_m^2 = \delta_m^2 \gamma_{m-1}^2 \) and by Lemma 3 applied on \( f_{m-1} \) and \( f_r, |c_r| \leq |a_{m-1}|, \) one finds \( \Delta \delta_m^2 \gamma_m^2 \leq (6 \cdot |a_0|)^2 \).

2. If the second root of \( f_m \) is strictly lower than 1, then by Lemma 2, \( f_{m+1} \) is reduced. The matrix of reduction is \( M = \begin{pmatrix} \alpha_r & \beta_r \\ \gamma_r & \delta_r \end{pmatrix} = M_m \cdot \begin{pmatrix} 0 & 1 \\ 1 & h_m \end{pmatrix} \), and \( r = m + 1 \). Thus \( \|M\|^2 \leq \|M_m\|^2 (1 + |h_m|)^2 \leq 4|a_0|/|a_m| \cdot 4h_m^2 \leq 16|a_0|/|a_r| \). One still has \( \Delta \delta_r^2 \gamma_r^2 \leq (|a_0| + |a_r \delta_r^2| + |c_r \gamma_r^2|)^2 \leq (21|a_0|)^2 \), because \( |c_r| \leq |a_m| \) by Lemma 2.

This concludes the proof of items 1) and 2) of Theorem 2. It remains the complexity issue, proved in the following paragraph.

**Complexity.** We now prove the number of iterations performed by RedGL2. Two steps before the end, at iteration \( r - 2 \) of RedGL2, we know that the form \( f_{r-2} = (a_{r-2}, b_{r-2}, c_{r-2}) \) satisfies \( \sqrt{\Delta} < |a_{r-2}| \), because the distance between the roots of \( f_{m+1} \) is smaller than 1. By Lemma 4 we have \( \omega^{r-4} \leq \|M_{r-2}\| \leq \sqrt{\frac{|a_0|/|a_r| - 2}{\log(|a_0|/\sqrt{\Delta})} \leq \sqrt{\frac{|a_0|}{\omega^2}} \). It follows that \( r - 4 \) is upper-bounded by \( \frac{\log(|a_0|/\sqrt{\Delta})}{2 \log(\omega)} \) steps where \( \omega = \frac{1+\sqrt{5}}{2} \).

The worst case complexity of algorithm RedGL2 is reached when all the normalizations occurring in the algorithm until the index \( r - 2 \) are by \( h = 1 \). For instance, we experimentally verify that it is the case on this family of inputs \( g.(T(-1)E)^n \) where \( g \) is reduced and \( n \) grows.
4 Proof of Heuristic Cryptanalysis of the NICE Cryptosystems

We propose an application of the results of the previous section to the cryptanalyses of the NICE cryptosystems. There are two variants, which are by chronological order NICE Imaginary [12] (with imaginary forms), and NICE Real [13] (with real forms). Their security relies on the intractability of factorization of the public discriminant $N$. They were designed for a similar level of security as RSA, but with faster decryption, since the decryption process has quadratic complexity. Both are now considered as broken. The first one succumbed by a proved arithmetic attack in [7]. However, the more general attack against both versions of NICE (in [6]) using lattice reduction remains only experimental and relies on two heuristic assumptions. In this paper, we provide an alternative point of view on the lattice attack, which allows to avoid the use of these heuristics and to prove the attack entirely.

Both variants of NICE (Real and Imaginary) have originally been described in terms of ideals of quadratic orders, and are based on a morphism between classes of primitive forms of fundamental discriminant $p$ and classes of primitive forms of non-fundamental discriminant $N = q^2p$. These notions are actually not needed here to understand the lattice attack, therefore we will here give a simple description solely in term of quadratic forms.

4.1 Lifting Quadratic Orders

We summarize some important properties on the relation between the sets $\mathfrak{F}_p$ and $\mathfrak{F}_N$ of primitives forms of discriminants respectively $p$ and $N = q^2p$, using the terminology we introduced in the last section. For the cryptographic interest we restrict ourselves to the case where $q$ is an odd prime. The following background theory can be found in [5,4,9].

Integer matrices of determinant $q$. We define an equivalence relation modulo $\text{SL}_2(\mathbb{Z})$ between two integer matrices $A$ and $B \in M_2(\mathbb{Z})$ by $A \equiv B \iff \exists M \in \text{SL}_2(\mathbb{Z}), \ AM = B$. The $2 \times 2$ integer matrices of determinant $q$ correspond to matrices of rank 1 mod $q$, they fall into $q + 1$ equivalence class, which are characterized by the (projective) direction from $\{0, 1, ..., q - 1, \infty\}$ of their image mod $q$. Each class contains a unique Hermite normal form: $Q_k = \begin{pmatrix} q & k \\ 0 & 1 \end{pmatrix}$, $k \in \{0, \ldots, q - 1\}$ or $Q_X = \begin{pmatrix} 1 & 0 \\ 0 & q \end{pmatrix}$.

Lift. As we can see in [5, section 7], for each form $f$ of discriminant $N = pq^2$ and each $M \in M_2(\mathbb{Z})$ of determinant $q$, there exists a (non-unique) form $g \in \mathfrak{F}_p$ such that $f = gM$. When $M = Q_X$, we define a particular function $\varphi$ (also called lift) which computes such $g \in \mathfrak{F}_p$ from $f = (a, b, c) \in \mathfrak{F}_N$ such that $\gcd(a, q) = 1$ as follows: $\varphi(f) = (a, \frac{b + 2ah}{q}, \frac{ah^2 + hh + c}{q^2})$ where $h \in [1 - q, \ldots, 0]$ and $h = -b/2a \mod q$. Note that all the divisions are exact since $f$ is primitive of discriminant $N = 0 \mod q^2$ and $q$ is an odd prime. It must be noted that the
lift preserves the first coefficient $a$ of the form. It is also clear that $\varphi$ preserves primary normalization, because its action on the roots of $f$ is a translation by $-h \in [0, q-1]$ followed by a division by $q$, which stabilizes the interval $[0, 1]$ of the largest root. Finally, equivalence of forms is stable by lift $\forall f, f' \in \mathfrak{F}_N$, $f \sim f' \implies \varphi(f) \sim \varphi(f')$.

The converse is in general false. Given a form $g \in \mathfrak{F}_N$ and $U$ its fundamental automorphism, there are exactly $q \sim (p/q)$ primitive forms (in $\mathfrak{F}_N$) among \{ $g.Q_0, \ldots, g.Q_{q-1}, g.Q_x$ \} where $(p/q)$ denotes the Legendre symbol. These forms split into $(q \sim (p/q))/s_q$ sets of $s_q$ equivalent forms (see [5] theorem 7.4), where $s_q$ is the order of $U$ modulo $q$. The fundamental unit $\epsilon_N$ is equal to the power $(\epsilon_p)^{s_q}$. These $(q \sim (p/q))/s_q$ different classes of equivalence are the only ones to be lifted to the class of $g$.

**Reduced cycle.** Let $g \in \mathfrak{F}_\Delta$ be a classically-reduced form of discriminant $\Delta > 0$, the right neighbour of $g$ is the classical normalization of $g.S(E)$. If we note $H(g)$ the largest normalization of $g$ (by the integer among $h^-_g, h^+_g$ of largest absolute value), then the right neighbour of $g$ is $g.S(E(H(g.S(E))))$. Successive iterations of the right neighbour enumerates all the reduced forms equivalent to $g$, and define the reduced cycle of the class of $g$. The cardinality of such reduced cycle is in $O(\log(\epsilon_\Delta))$ where $\epsilon_\Delta$ is the fundamental unit.

**Principal cycle, and $q$-belt.** The principal class of a discriminant $\Delta > 0$ is the class containing $(1, 1, *)$. The principal form is the classical-normalization of this form, and the principal cycle $1_\Delta$ is the reduced cycle of the principal class. Note that the principal class is the only class containing a form of first (or last) coefficient equals to 1.

We define the $q$-belt of a discriminant $N = pq^2$ as the set of all primary normalized forms $(q^2, kq, *)$ of the principal class. Necessarily, $k \in [-\sqrt{p}, 2q-\sqrt{p}]$. There are exactly $s_q - 1$ forms in the $q$-belt of $N$: let $g_0$ be the principal form $(1, *, *)$ of $\mathfrak{F}_N$ and $f = \varphi(g_0)$ is (necessarily) the principal form of $\mathfrak{F}_p$. Let $U$ be the fundamental automorphism of $f$, we set by induction $k_0 = \infty$ and $k_i$ the unique integer such that $UQ_{k_{i-1}} \equiv Q_{k_i}$ for $i \geq 1$. Note that $Q_{k_i} \equiv U^i Q_{k_0}$, and that the order of $U \mod q$ is precisely $s_q$, therefore the sequence $(k_i)$ is periodic and $k_{s_q} = k_0 = \infty$. Finally, the $q$-belt of $N$ is the set \{ $g_1 = f.Q_{k_1}, \ldots, g_k = f.Q_{k_{s_q-1}}$ \}. They are indeed primary-normalized and equivalent by construction. A transformation matrix from $g_i$ to $g_{i-1}$ is by construction $Q_{k_i}^{-1} UQ_{k_{i-1}} \in \text{SL}_2(\mathbb{Z})$, because $UQ_{k_{i-1}} \equiv Q_{k_i}$.

### 4.2 Cryptosystem Real NICE

We now describe the NICE Real encryption and decryption. The public key is a composite integer $N = pq^2$ and the secret key $(p, q)$ with $p$ and $q$ two distinct primes of the same size, satisfies two conditions:

- $p$ is a Schinzel prime [19] which is a positive squarefree integer of the form $p = A^2 x^2 + 2Bx + C$ with $A, B, C, x \in \mathbb{Z}$, $A \neq 0$ and $B^2 - 4AC$ dividing $4 \gcd(A^2, B)^2$. Such special primes implies a very low number of reduced
forms in each class, namely there are $O(\log(p))$ reduced forms in $\mathfrak{F}_p$ in each equivalence class ([8] and [22, theorem 5.8, p. 52]). It is therefore practical to enumerate every reduced form equivalent to a given one. With a generic discriminant, the number of reduced forms per cycle would be exponential, around $O(\log(p))$ (see [3]). To avoid any confusion, please note that even for a Schinzel prime, the number of classes in $\mathfrak{F}_p$ remains exponential.

$q$ is such that $s_q$ is linear in $q$. This imply that the number of reduced forms of discriminant $N = q^2p$ in each equivalence class is at least linear in $q$ and upper-bounded by $O(q \log(p))$, which is exponential.

The encryption of a message $m$ works as follows: $m$ is embedded into a (usually prime) integer $a \leq \sqrt{p}/2$ which satisfies some low-probability pattern, and such that $q^2p$ is a square modulo $a$. This integer is expanded into a quadratic form $f_s = (a, b', c')$ of discriminant $q^2p$ (which is not printed). The ciphertext is a random reduced form $f_c$ equivalent to $f_s$ (there are exponentially many). It can be generated from $f_s$ by successive multiplications by random unimodular matrices and reductions.

The decryption algorithm lifts the ciphertext in $\mathfrak{F}_p$ and enumerate all the reduced forms equivalent to $\varphi(f_c)$, looking for the pattern. Of course, the knowledge of $q$ is needed to compute $\varphi$. There are only $O(\log(p))$ of them. It will necessarily find it, because the (unknown) lift of $f_s \sim f_c$ is an equivalent form $\varphi(f_s) = (a, *, *)$, whose normalization $(a, *, *)$ is reduced due to the small size of $a$, and it satisfies the pattern by construction. Due to the small number of reduced forms, it is likely the only one of the small reduced cycle to satisfy the pattern, and the plaintext $m$ is eventually extracted from $a$.

### 4.3 Cryptanalysis

The cryptanalysis of NICE Real presented in [6] works as follows. The authors present an algorithm inspired of Coppersmith methods (see [10,17]), which solves in polynomial time the equation $au^2 + buv + xv^2 = 0 \mod q^2$ in the variables $(u, v, q)$ where $N = pq^2$ is known and $\max(|u|, |v|) = O(N^{1/9}).$ They call this algorithm Homogeneous-Coppersmith in [6]. Their cryptanalysis of NICE Real is: Pick a form $g$ of the principal cycle, and try to solve the equation $g(u, v) = 0 \mod q^2$ with Homogeneous-Coppersmith. Repeat this until it finds a solution $(u, v, q)$ and return the private key $q$.

The proof of the attack of [6] relies on this heuristic assumption:

**Assumption 1.** The cardinality of the set $A = \{g \in \mathbb{1}_N, \exists(u, v) \max(|u|, |v|) \leq O(N^{\frac{1}{\gamma}}) \text{ and } g(u, v) = 0 \mod q^2\}$ is linear in $s_q$.

---

3 The authors of [6] enumerates the forms sequentially, until it finds a solvable one. They need an assumption not only on the large number of such forms, but also on their regular repartition on the principal cycle. Randomizing the enumeration avoids to prove the assumption on regular repartition (Heuristic 2 in [6]), which is feasible using the distance introduced in Theorem 3, but is beyond the scope of this paper.
The authors of [6] experimentally verify this assumption. Namely, if \( \bar{g}_k \) denotes the reduction of the form \( g_k = (q^2, *, *) \) of the \( q \)-belt by Classical Gauss reduction. The bottom two coefficients of the reduction matrix satisfy \( \bar{g}_k(\delta, -\gamma) = q^2 \).

**Homogeneous-Coppersmith** experimentally recovers \( (\delta, -\gamma) \) for most of the \( \bar{g}_k \) and even a few of their direct left or right neighbours on the principal cycle. This indicates that the norm of the reduction matrix is in general upper-bounded by \( O(N^{1/9}) \).

However we also found rare cases of \( \bar{g}_k \) where the norm of reduction matrix was by an order greater than \( N^{1/9} \), and on which Homogeneous-Coppersmith algorithm cannot find any solution. We call these particular forms *unbalanced*, because they have in general an unusually small coefficient. The main three difficulties which prevented the authors of [6] to prove Assumption 1 were to justify that the proportion of unbalanced forms is negligible among the set of \{\( \bar{g}_k \)\}, that the reduction matrix using Classical Gauss reduction is bounded by \( O(N^{1/9}) \), and that Classical Gauss is injective on a large enough subset of the \( q \)-belt, which prevents \{\( \bar{g}_k \)\} from being too small.

Our first improvement in their analysis is to replace the Classical Gauss reduction algorithm with RedGL2. This allows to square-root the upper-bounds on the reduction matrix as of Theorem 2. Thus we define \( \bar{g}_k \) as the reduction by RedGL2 of the \( q \)-belt form \( g_k \) for each \( k \). We ensure that \( \bar{g}_k \) is classically reduced and that the reduction matrix has determinant +1 using Lemma 1. The first point of Theorem 2 implies that the norm of the reduction matrix is in \( O(N^{1/9}) \) as soon as the smallest coefficient of \( \bar{g}_k \) is greater than \( N^{1/9} \). We can either prove that this condition is satisfied by a large proportion of the \( g_k \), or we can also circumvent this limitation by using the second point of Theorem 2, which indicates that the size of the product \( |uv| \) is always upper-bounded by \( O(N^{1/6}) \).

We therefore improve the Homogeneous-Coppersmith algorithm so that it also finds unbalanced solutions: namely, we design a rational variant of Boneh-Durfee-HowgraveGraham algorithm [2] which in particular solves \( g(u, v) = au^2 + buv + cv^2 = 0 \mod q^2 \) on \( (u, v, q) \) as soon as the product \( |uv| \) is in \( O(N^{2/9}) \).

Our new polynomial attack on Nice Real is the following: Randomly select a form \( g \) on the principal cycle \( \mathbb{F}_N \), and try to solve \( g(u, v) = 0 \mod q^2 \) in \( (u, v, q) \) using Rational-BonehDurfeeHowgraveGraham. Repeat until it finds a solution, and return \( q \).

The proof of this attack works in two steps: first, we prove (in Theorem 3) that the above-defined \( \bar{g}_k \) represent a non-negligible proportion of the principal cycle, and second, we prove (in Section 4.4) that Rational-BonehDurfeeHowgraveGraham finds \( q \) from any of the \( \bar{g}_k \) in polynomial time.

**Definition 2 (distance).** we define a notion of distance between two equivalent forms \( f \sim g \) as \( \text{dist}(f, g) = \min\{\log(\|\mathcal{M}\|), \mathcal{M} \in SL_2(\mathbb{Z}) \text{ and } f.\mathcal{M} = g\} \).

Let \( f, g, h \) be three equivalent forms in \( SL_2(\mathbb{Z}) \) and \( f.\mathcal{M} = g \). The distance function satisfies the following properties:

1. \( \text{dist}(f, g) = \text{dist}(g, f) \geq 0 \)
2. \( \text{dist}(f, g) = 0 \iff f = g \) or \( f = g.SE \)
Proof. The first three points follow from basic properties of the induced norm, and the fact that only isometries have a unit norm. To prove the fourth statement, let $U$ be the fundamental automorphism of $f$, the eigenvalues of $U$ are $\epsilon_\Delta$ and $\epsilon_\Delta^{-1}$. Any non-trivial automorphism $V$ of $f$ satisfies $\|V\| \geq \epsilon_\Delta$, because $V$ is a non-zero power of $U$, and its spectral radius is a positive power of $\epsilon_\Delta$. The matrix $\mathcal{M}$ of the fourth point is necessarily the smallest transformation matrix from $f$ to $g$, otherwise any matrix $X \in \text{SL}_2(\mathbb{Z})$ such that $f.X = g$ and $\|X\| < \|\mathcal{M}\|$ would produce a non-trivial automorphism $\mathcal{M}X^{-1}$ of $f$ of too small norm $\|\mathcal{M}X^{-1}\| < \epsilon_\Delta$, which is impossible.

One of the greatest advantage of this distance is the fourth statement, which in general indicates that any polynomial transformation matrix is necessarily the smallest one. This allows to efficiently lower-bound a distance. As shown in the proof, it is essential that the group of automorphism is cyclic, the fourth statement would be false on $\text{GL}_2(\mathbb{Z})$. The authors of [6] used another distance between $(f,g)$, which could have been formalized as the smallest $k \in \mathbb{N}$ such that there exists $h_1, \ldots, h_k$ such that $\prod_{i=1}^k \text{SET}(h_i)$ transforms $f$ into $g$ or $g.\text{SE}$. Inside the reduced cycle, this corresponds to Shanks distance [20]. Unfortunately, it does not satisfy any equivalent of the fourth point: there is no way to efficiently verify that a given distance, as small as it could be, is correct. All the variants we found of this distance, which aims to approximate this statement, based either on the logarithms of the $h_i$ or some maximum norms, break the positive definiteness or the triangular inequality. This explains why we do not base our proof on Shanks distance and introduce our own instead.

Theorem 3. Given a NICE modulus $N = pq^2$, the set $\mathcal{A}' = \{ \hat{g}_k = \text{RedGL2}(g_k), k \in [1, \ldots, s_q - 1]\}$ of the reduced of the $q$-belt has at least $K.s_q$ elements for some constant $K > 0$.

Proof. We now call $U_p$ the fundamental automorphism of the principal form of $\mathfrak{f}_p$. We verify that $\|U_p^j\| \leq 2(\epsilon_p^j + \epsilon_p^{-j})$ and that for all $i,j$, $Q_{k_i}^{-1}U_p^jQ_{k_i+j}$ transforms $g_i$ into $g_{i+j}$. Its norm is bounded by $\frac{1}{q}\|Q_{k_i}\| \cdot \|U_p^j\| \cdot \|Q_{k_i+j}\| < 4q(\epsilon_p^j + \epsilon_p^{-j})$.

Due to point 4, for all $j \in [1, (s_q/2) - 2]$, the distance $\text{dist}(g_i, g_{i+j}) = \log(\|Q_{k_i}^{-1}U_p^jQ_{k_i+j}\|)$ is greater than $j \log(\epsilon_p) - \log(2q)$. By Theorem 2, the norm of the reduction matrix from a $g_i$ to $\hat{g}_i$ is upper-bounded by $2.21q^3/\sqrt{N} = 42q^2/\sqrt{\pi}$, and it follows that $\text{dist}(\hat{g}_i, \hat{g}_{i+j}) \geq j \log(\epsilon_p) - \log(3528q^3p)$. For this reason, if $j > \log(3528q^3p)/\log(\epsilon_p)$, then $\text{dist}(\hat{g}_i, \hat{g}_{i+j}) > 0$ and $\hat{g}_i \neq \hat{g}_{i+j}$. Using the NICE parameters, one has $\log(3528q^3p)/\log(\epsilon_p) < 3$, thus the forms $\hat{g}_1, \hat{g}_4, \hat{g}_7, \ldots, \hat{g}_{3n+1}$ are distinct (with $n \leq s_q/6$).
4.4 Rational Improvement of the Boneh-Durfee-HowgraveGraham’s Algorithm

In this section, we describe our Rational-BonehDurfeeHowgraveGraham algorithm as a variant of Boneh Durfee Howgrave-Graham algorithm [2] solving rational linear equations \( u/v - C = 0 \mod q \) in the variables \((u, v, q)\) when a multiple \( N = pq^r \) is known. The description of Rational-BonehDurfeeHowgraveGraham is summarized in Algorithm 2. Among others, it can be used to solve all the equations \( g_k(u, v) = au^2 + bw + cv^2 = 0 \mod q^2 \) of discriminant \( pq^2 \) of the previous section, because they are equivalent to \( u/v + b/2a = 0 \mod q \). Since the solution we are looking for satisfies \(|uv| = O(N^{1/6})\), the following Theorem 4 proves that Rational-BonehDurfeeHowgraveGraham finds all solutions \(|uv| = O(N^{2/9})\), and concludes the proof of our new attack on Nice Real.

More generally, given a polynomial \( P \), the technique due to Boneh Durfee Howgrave-Graham transforms the equation \( P(u/v) = 0 \mod q \), into a lattice \( L \) of dimension \( m \) and bounded determinant, and whose short vectors are orthogonal to the integer vector \( S = (u^m, u^{m-1}v, ..., uv^{m-1}, v^m) \). The solutions \( u \) and \( v \) can be extracted from any of those short lattice vectors. This lattice is described by a basis \( B \), whose rows contain the coefficients of \((m - 1)\)-degree polynomials having \( u/v \) as a root modulo a power of \( q \). When \( u \) and \( v \) have approximately the same size (like in Homogeneous-Coppersmith of [6]), the celebrated LLL reduction algorithm on \( B \) outputs directly the desired vector orthogonal to \( S \). Otherwise, when \( u \) and \( v \) are unbalanced, say for instance that \( u \) is 1000 times larger than \( v \), one first needs to re-balance the lattice by multiplying each \( i \)-th column by \( C^i \), where \( C \) is close to 1000, and only then reduce the basis. The original Boneh-Durfee-HowgraveGraham’s algorithm, which interests in integer solutions (arbitrary \( u \) and \( v = 1 \)), follows the above rule: the lattice basis which is actually LLL-reduced is the basis of Homogeneous-Coppersmith where each \( i \)-th column has been multiplied by \( X^i \), where \( X \) is a power of 2 just larger than the solution \( u \). More generally, if we don’t know the relative balance between \( u \) and \( v \) but only know that the size of \( uv \) is \( n \)-bits, then we can test the \( n \) possible powers of two sequentially within a linear-factor overhead. Besides, we remark that instead of multiplying the columns of the input Homogeneous-Coppersmith basis by \((1, 2, 4, ..., 2^m)\), we describe the exact same lattice by multiplying the columns of the LLL-reduced basis, and the second one is almost reduced (LLL terminates in a very few steps). Thus after the reduction of the first Homogeneous-Coppersmith basis, one obtains all the other possible balances of \( u \) and \( v \) for free.

**Theorem 4.** Given any integer \( N = pq^r \) (where \( p \) and \( q \) are unknown), and a bound \( \beta < \frac{1}{4}q^\log(q^\beta)\log(N) \), Algorithm 2 terminates in polynomial time, and finds a solution (if it exists) of the equation \( \frac{u}{v} = c \mod q \) where \((u, v)\) are unknown integers satisfying \(|uv| < \beta\).

**Proof.** Let \((U, V) \in \mathbb{R}^2\) such that \(|u| \leq U \) and \( v \leq V \). We use the same parameters \( m \in \mathbb{N}\setminus\{0\} \) and \( t = \left\lfloor \frac{(m+1)\log(q^\beta)}{\log(N)} \right\rfloor \).

We denote by \( \mathbb{R}_m[X, Y] \) the span of homogeneous polynomials of degree \( m \), and we define the isomorphism \( \varphi : \mathbb{R}_m[X, Y] \rightarrow \mathbb{R}^{m+1} \) which computes...
This allows to recover Castagnos for useful discussions and valuable comments on this paper.

Acknowledgements.

Further lead would be to extend these results on the reduction of the forms in systems \([12,13]\), which is unusual in the history of lattice based cryptology. A precision of our analysis, in the worst case and also in the average case, because we mostly manipulate only positive matrices, which are easy to bound. We saw that reduction algorithms are conceptually simpler to study in \(GL_2(\mathbb{Z})\), any integer linear combination \(R\) of \((\alpha_0, \ldots, \alpha_m)\) the first vector

\[
\begin{aligned}
&\text{Compute the family } P_k(X, Y) = N\left[\frac{x}{y}\right] \cdot (X - cY)^k \cdot Y^{m-k} \text{ for } k = [0..m] \\
&\text{Express (or update) the family } (P_k)_{k=0..m} \text{ on the monomial basis } \\
&\left(\frac{x^k Y^{m-k}}{U^k V^{m-k}}\right)_{k=0..m}, \text{ and form a matrix } B \in M_{m+1}(\mathbb{Z}) \\
&\text{LLL-reduce } B, \text{ and call } (\alpha_0, \ldots, \alpha_m) \text{ the first vector} \\
&\text{for each rational root } \frac{x}{y} \text{ of } R(X) = \sum_{k=0}^{m} \frac{\alpha_k}{\beta_k} X^k = 0 \text{ do} \\
&\text{if } |uv| \leq \beta \text{ and } \gcd(u-cv,N) \text{ is non-trivial return } (u,v) \\
&\text{end for} \\
&\text{end for}
\end{aligned}
\]

the coordinates of a polynomial on the basis \(\left(\frac{x^k Y^{m-k}}{U^k V^{m-k}}\right)_{k=0..m}\). For instance, \(\varphi(X^k Y^{m-k}) = U^k V^{m-k} e_k\) where \(e_k\) is the \(k\)-th canonical basis vector. Let \((P_k)_{k \in [0..m]}\) be the family \(P_k(X, Y) = N\left[\frac{x}{y}\right] \cdot (X - cY)^k \cdot Y^{m-k} \in \mathbb{R}_m[X, Y]\). By construction, any integer linear combination \(R \in \sum_{k=0}^{m} Z \cdot P_k\) satisfy \(R(0) = 0 \mod q^t\) and \(|R(u, v)| \leq \sqrt{m + 1} \cdot ||\varphi(R)||_2\) (using Cauchy-Schwartz inequality). We now suppose that \(\varphi(R)\) is a short vector of the lattice generated by the (triangular) basis \(B = (\varphi(P_k))_{k \in [1,m]}\). By that, we mean \(||\varphi(R)||_2 \leq (1.08)^{m+1} \det(B)^{1/(m+1)}\). Such a vector can be found by running the LLL algorithm on the lattice basis \(B\) (see [16]). The remainder of the proof is just a formal verification that when \(m\) grows, \(\det(B)\) is small enough to guaranty that \(|R(u, v)| < q^t\), and therefore that \(R(u, v) = 0 \text{ (in } \mathbb{Z})\). Since \(R\) is homogeneous, this allows to recover \(u\) and \(v\).

5 Conclusion

We saw that reduction algorithms are conceptually simpler to study in \(GL_2(\mathbb{Z})\), because we mostly manipulate only positive matrices, which are easy to bound. The precision of our analysis, in the worst case and also in the average case, allows us to fully prove a lattice-based total-break attack against Nice cryptosystems [12,13], which is unusual in the history of lattice based cryptology. A further lead would be to extend these results on the reduction of the forms in higher dimension.

Acknowledgements.

We would like to thank Fabien Laguillaumie and Guilhem Castagnos for useful discussions and valuable comments on this paper.
References