MATHEMATICS OF COMPUTATION
Volume 00, Number 0, Xxxx XXXX, Pages 000–000
S 0025-5718(XX)0000-0

EFFICIENT SOLUTION OF RATIONAL CONICS

J. E. CREMONA AND D. RUSIN

Abstract. We present efficient algorithms for solving Legendre equations over $\mathbb{Q}$ (equivalently, for finding rational points on rational conics) and parametrizing all solutions. Unlike existing algorithms, no integer factorization is required, provided that the prime factors of the discriminant are known.

1. Introduction

1.1. **Summary of results.** In this paper we give efficient methods of finding all rational points on a rational conic $\mathcal{C}$, given by non-singular homogeneous equation of degree 2:

$$\mathcal{C}: \quad f(X,Y,Z)=0. \tag{1}$$

One method for finding one rational point on $\mathcal{C}$, if one exists, is the original descent method of Legendre. We show how one may easily make a significant improvement to this (reducing the number of iterations from exponential in the size of the input to linear); and also, but with more work, make an even greater improvement. This last method involves no integer factorization other than that of the discriminant of the original equation (which is in any case necessary for deciding the solubility of (1)). It is the necessity of factorizing "spurious" integers arising during the course of the computation which is the bottleneck in simpler reduction methods; our "factorization-free" method avoids this entirely.

We also describe a factorization-free method of solution based on lattice reduction; this is not original, though apparently not well known.

We present examples and timings of our implementation of both methods; these indicate that the reduction method is faster in practice than the lattice-based method. Both are linear time, given a so-called solubility certificate (defined below), and probabilistic polynomial time given only the factorization of the discriminant.

As an example of the speed which is now attainable, the solution of an equation of the form $ax^2+by^2=cz^2$, where a, b and c are 200-digit primes, takes less than 2 seconds on a modest PC. Such a problem is not feasible to solve in reasonable time with Legendre's method (as in Maple, for example).

We also show how to parametrize all rational points on $\mathcal{C}$, given one point, in the most efficient way. This is necessary for several applications, such as to 2-descent on elliptic curves, and is also used for finding a small single solution to (1).

It would be useful and interesting to extend the algorithms presented here to number fields. We say little more about this here, but refer to the paper [11] by Pohst, and Simon's thesis [14].

The factorization-free algorithm presented here has been implemented in release 2.8 (July 2001) of the package Magma.

1991 *Mathematics Subject Classification.* Primary 11G30, 11D41.

1.2. **Background.** By the Hasse or local-global principle for curves of genus 0, the curve $\mathcal{C}$ has rational points if and only if it has points everywhere locally. Thus, testing (1) for solubility is easy, at least in theory, and in practice no harder than factorizing the discriminant of the given equation (see Section 1 below for details).

Our first main concern will be to find efficiently one solution when solutions exist. Here and throughout we will pass freely between the geometric language of "points on curves" and the Diophantine language of "solutions to equations". We always exclude the trivial solution $(x, y, z) = (0, 0, 0)$, as we are really interested in projective solutions $(x : y : z) \in \mathbb{P}^2(\mathbb{Q})$, each of which has a "primitive" representation with $x, y, z \in \mathbb{Z}$ and $\gcd(x, y, z) = 1$, unique up to sign.

Secondly, we will want to find a "small" solution. Holzer's Theorem (see below for a precise statement) asserts that a soluble equation always has solutions which are not too large in terms of the coefficients. Any given solution may be reduced, using a method of Mordell, until it satisfies Holzer's bounds. We present an alternative reduction method, faster than Mordell's, though the solution it gives may not be quite Holzer-reduced.

Finally, given one solution $P_0 = (x_0, y_0, z_0)$ to (1), one can write down a parametrization of all solutions of the form

$$X = Q_1(U, V), \quad Y = Q_2(U, V), \quad Z = Q_3(U, V) \tag{2}$$

where each $Q_i(U, V) \in \mathbb{Z}[U, V]$ is a quadratic form. Geometrically, the homogeneous coordinates $(U : V)$ parametrize the pencil of lines through P_0, each of which intersects the conic $\mathcal{C}$ in a unique second point. Our final task will be to find such a parametric solution which is as simple as possible. We will see that a parametrization exists such that the discriminants of the polynomials $Q_i(U, V)$ are prescribed in terms of the coefficients of the defining polynomial $f(X, Y, Z)$, independently of the particular basic solution P_0 found earlier. This last point is particularly significant in certain applications.

Our approach throughout will be algorithmic, and our results will be in the form of efficient algorithms to carry out the tasks we have just outlined. We will give examples to show that our method is more efficient, and leads to better (meaning smaller) solutions than those which can be found elsewhere (for example, by using the Maple computer algebra system). The mathematics here is entirely elementary, and mostly also quite well-known, but we are not aware of a systematic treatment of such equations in the literature which is both algorithmic and concerned with the size of the parametric solutions obtained.

A slightly different problem is to parametrize all "primitive" integer solutions (x, y, z) to (1) using integer quadratics $Q_i(U, V)$. Mordell showed that this is possible using a finite family of quadratic parametrizations of the form (2). Since we are interested in projective solutions we are not interested in the primitivity, and our task is therefore slightly simpler.

The application which led us to develop these methods is in higher descents on elliptic curves over $\mathbb{Q}$, starting with a descent via 2-isogeny. See [5] for details of this. Another application of which we are aware is the determination of explicit equations for hyperelliptic curves whose Jacobians are quotients of modular Jacobians: see the theses of Wilson [19] (Oxford, 1998) and Weber [17] (Essen, 1996) for examples of these. It is remarkable that an algorithmic solution to the problem of finding all rational points on a curve of genus 0 has not yet been perfected (as remarked by Mazur in [9]), given the current activity on a wide scale concerning constructive solutions to Diophantine equations of higher genus, so it is interesting to note that efficient solutions to this simpler problem are also required for the study of curves of higher genus.

We are grateful to Denis Simon for the reference [12].

2. Single Solutions

2.1. **Standard forms of equation.** By elementary algebra we may complete the square in the general quadratic form $f(X,Y,Z)$ to obtain the *diagonal form*

$$aX^2 + bY^2 + cZ^2 = 0, \tag{3}$$

often called Legendre's equation. Since the equation is assumed to be nonsingular, we have $abc \neq 0$. Furthermore, by simple scaling of the variables we may reduce to the case where the coefficients are integers which are (i) pairwise coprime, and (ii) square-free, so that abc is square-free. Achieving condition (i) only requires gcd computations, while (ii) requires factorization of the coefficients. We will assume throughout that this factorization is known. Such an equation (3) will be called *reduced*; it is unique, up to permutations of the variables and changing all the signs. Since real solubility requires that the coefficients do not all have the same sign, we also assume that $a > 0$, $b > 0$ and $c < 0$.

It will also sometimes be useful to put our equation into *norm form*

$$X^2 - aZ^2 = bY^2. \tag{4}$$

Solving this amounts to expressing b as a norm from $\mathbb{Q}(\sqrt{a})$, if possible. In this form we can require that a and b are both square-free integers, but not that they are coprime. Real solubility requires that a and b are not both negative. We will use this form for the first recursive solution of the equation below.

Lastly, for the applications to elliptic curves it is most convenient to use a form of the equation slightly more general than the diagonal form, which we call the *semi-diagonal form*:

$$aX^2 + bXZ + cZ^2 = dY^2. \tag{5}$$

Here we will require that all the coefficients are integers with d squarefree, $d(b^2-4ac)$ nonzero for nonsingularity, and $\gcd(a,b,c,d) = 1$. In our application we also have $ac \neq 0$ and so we will also assume this below.

2.2. **Local Solubility Criterion and Holzer's Theorem.** The necessary and sufficient criterion for solubility of (3) is simply that it should have solutions in $\mathbb{Q}_p$ for all primes p and also in $\mathbb{R}$. This result is usually referred to as Legendre's Theorem. For odd primes p not dividing abc there is always a local solution, so this only gives a finite number of conditions to check. Checking these conditions in practice does require us to factorize the coefficients. Suppose that (3) is reduced, so that abc is square-free. If p is odd and divides c (say), then solubility in $\mathbb{Q}_p$ follows from solubility modulo p (by Hensel's Lemma), and hence from the condition that the Legendre symbol $\left(\frac{-ab}{p}\right)$ is $+1$. Hence the local conditions for all odd finite primes are equivalent to the existence of solutions to the following quadratic congruences:

$$X_1^2 \equiv -bc \pmod{a}; \quad X_2^2 \equiv -ca \pmod{b}; \quad X_3^2 \equiv -ab \pmod{c}. \tag{6}$$

Moreover, the number of local conditions which fails must be even (by the product formula for the Hilbert symbol), so the solubility of these congruences together with the sign condition ensuring real solubility are already sufficient to ensure global solubility, and a 2-adic condition is not needed.

Definition 2.1. A triple $(k_1, k_2, k_3) \in \mathbb{Z}^3$ is called a *solubility certificate* for (3) if it gives a solution to the congruences (6).

We summarize the local solubility criterion as follows.

Lemma 2.1. *Let a, b and c be nonzero integers with abc squarefree, not all of the same sign. Then* (3) *has a solution if and only if a solubility certificate exists.*

If a, b and c are pairwise coprime (but not necessarily square-free), then the existence of a solubility certificate is sufficient, but no longer necessary, for the existence of solutions to (3).

A proof of the last statement is implicit in the algorithms below, which guarantee to deliver a solution from a solubility certificate provided only that a, b and c are pairwise coprime and not all of the same sign. That the existence of the certificate is not necessary when the coefficients are square-free may be seen from the equation $9X^2 - Y^2 - Z^2 = 0$, which has the solution $(1, 3, 0)$ but no certificate since the congruence $X_1^2 \equiv -1 \pmod 9$ has no solution.

Associated to the triple of coefficients (a, b, c) and the certificate (k_1, k_2, k_3) we will associate a 3-dimensional sublattice $\mathcal{L} = \mathcal{L}(a, b, c; k_1, k_2, k_3)$ of $\mathbb{Z}^3$, called the solution lattice for the certificate, as follows:

$$\begin{aligned} \mathcal{L}(a,b,c;k_1,k_2,k_3) = \{(x,y,z) \in \mathbb{Z}^3 \,|\, by &\equiv k_1 z \pmod a, \\ cz &\equiv k_2 x \pmod b, \\ ax &\equiv k_3 y \pmod c\}. \end{aligned} \tag{7}$$

The index of $\mathcal{L}(a, b, c; k_1, k_2, k_3)$ in $\mathbb{Z}^3$ is $|abc|$. One easily checks that for $(x, y, z) \in \mathcal{L}$, we have $ax^2 + by^2 + cz^2 \equiv 0 \pmod{abc}$. In the second and third algorithms we present below, we will construct a solution of (3) which lies in the solution lattice for any given solubility certificate.

The first algorithm we give below for solving conics itself constitutes a proof of Legendre's theorem, since it is guaranteed to find a solution unless either a quadratic congruence fails to be soluble, or the signs of the coefficients are wrong. Indeed, Legendre's own proof follows the same lines: see the account in Weil's historical book [18, p. 100]. Algorithmic solutions in the literature often follow essentially the same reduction procedure as Legendre (see [8], or [15] for a recent example). As we will see, this method has two disadvantages in practice: it takes many steps, each of which involves the factorization of an integer, and the resulting solution can be very large. Our first improvement already performs better in these respects; although it does not eliminate the factorization from each step, the number of steps is reduced, the numbers to be factorized are smaller, and the resulting solution is also smaller. Then the "factorization-free" version of the reduction method eliminates the need for any factorization, given a solubility certificate, giving even greater improvement and making possible the solution of equations whose coefficients have hundreds of digits in only a few seconds.

In the famous paper [1], in which higher descents were used to study the ranks of elliptic curves of the form $Y^2 = X^3 - DX$, the authors remark [1, p. 100] that the solution of various auxiliary conics is the most time-consuming part of the descent process. We also found this to be true (despite having 30 years of factorization technology to hand) before using the new methods described here.

Now assume that (3) is soluble. Holzer's theorem asserts that there exists an integral solution (x, y, z) with

$$|x| \le \sqrt{|bc|}; \qquad |y| \le \sqrt{|ac|}; \qquad |z| \le \sqrt{|ab|}, \tag{8}$$

or equivalently,

$$\max(|a|x^2, |b|y^2, |c|z^2) \le |abc|. \tag{9}$$

Such a solution we will call "Holzer-reduced". Holzer's Theorem is not trivial to prove: see [3] for a recent fairly short proof, improving earlier versions by Mordell and Cassels (see section 2.4 below for more on this). In Mordell's proof, one obtains a solution which does not necessarily satisfy Holzer's bounds, and then reduces the solution using the following lemma from [10, Theorem 5, p.47].

Lemma 2.2. *Let a, b and c be nonzero integers with abc squarefree, $a > 0$, $b > 0$ and $c < 0$, and let (x_0, y_0, z_0) be a solution of (3). If $|z_0| > \sqrt{ab}$ then there exists a solution (x_1, y_1, z_1) with $|z_1| < |z_0|$.*

We will give Mordell's construction below. After a finite number of steps, we arrive at a solution (x, y, z) with $|z| \leq \sqrt{ab}$, and then the inequalities on x and y follow immediately. We will also present a new method of reducing solutions which is faster than Mordell's, but only produces a solution satisfying

$$\max(|a|x^2, |b|y^2, |c|z^2) \leq \frac{4}{3}|abc|. \tag{10}$$

A similar result concerning small solutions to Legendre's equation over totally real number fields can be found in [11].

2.3. **Algorithm I: Legendre-type reduction.** The first algorithm for finding one solution works with the equation in the norm form (4), where the coefficients a and b are square-free nonzero integers, not necessarily coprime. By symmetry we may assume that $|a| \leq |b|$, interchanging a and b if necessary. The idea, which originates with Legendre, is to proceed by descent, reducing the problem of solving (4) to that of solving a similar equation with a smaller b coefficient. This step is repeated until $|b| < |a|$, after which a and b are interchanged. The base cases in which no further descent is necessary are trivially dealt with. The full procedure is as follows.

Algorithm I.

1. If $|a| > |b|$ then swap a and b, solve the resulting equation, then swap y and z in the solution obtained.
2. If $b = 1$ then set $(x, y, z) = (1, 1, 0)$ and stop.
3. If $a = 1$ then set $(x, y, z) = (1, 0, 1)$ and stop.
4. If $b = -1$ there is no solution (since a must be -1).
5. If $b = -a$ then set $(x, y, z) = (0, 1, 1)$ and stop.
6. If $b = a$ then let (x_1, y_1, z_1) be a solution of $X_1^2 + Z_1^2 = aY_1^2$, set $(x, y, z) = (ay_1, x_1, z_1)$ and stop.
7. Let w be a solution to $X^2 \equiv a \pmod{b}$ with $|w| \leq |b|/2$, and set $(x_0, z_0) = (w, 1)$, so that $x_0^2 - az_0^2 \equiv 0 \pmod{b}$.
8. Use lattice reduction to find a new nontrivial solution (x_0, z_0) to the congruence $X^2 - aZ^2 \equiv 0 \pmod{b}$, with $x_0^2 + |a|z_0^2$ as small as possible.
9. Set $t = (x_0^2 - az_0^2)/b$, and write $t = t_1 t_2^2$ with t_1 square-free.
10. Let (x_1, y_1, z_1) be a solution to $X^2 - aZ^2 = t_1 Y^2$; then
$$(x, y, z) = (x_0 x_1 + a z_0 z_1, t_1 t_2 y_1, z_0 x_1 + x_0 z_1)$$
is a solution to (4): stop.

By the end of step 6 we have reduced the problem to solving equations in which $|b| \geq 2$, $|b| > |a|$ and $a \neq 1$ (though $a = -1$ is possible). The reduction step proceeds by first solving the quadratic congruence

$$X^2 \equiv a \pmod{b}$$

to obtain a solution w with $|w| \leq |b|/2$. The usual algorithm for this step involves factorizing b, finding a square root of a modulo each prime divisor of b, and combining them with the Chinese Remainder Theorem. All these square roots must exist if the equation passes the local solubility criterion. We then have $w^2 - a = bt$, where the integer t satisfies

$$|t| < \frac{1}{4}|b| + 1 \leq \frac{1}{2}|b|;$$

(here we use $1 \leq |a| < |b|$). The standard algorithm found in the literature (as in [15], for example) omits step 8, using the fact that this value of t is strictly less than b to obtain a descent. This procedure works perfectly well in practice provided that the initial coefficients a and b are fairly small. The size of the larger coefficient is reduced by a factor of approximately 4 at each step; the main problem with large examples is the need to factorize all the coefficients b which arise, in order to solve the associated quadratic congruences.

Our improvement consists of inserting the extra step 8 above. We have one solution $(x_0, z_0) = (w, 1)$, to the congruence

$$X^2 - aZ^2 \equiv 0 \pmod{b}. \tag{11}$$

Using an elementary lattice reduction technique we find the solution (x_0, z_0) to this congruence which minimizes $x^2 + |a|z^2$, and set $t = (x_0^2 - az_0^2)/b$. This will be very much smaller than the earlier value of t. Explicitly, the minimal vector has length $O(b\sqrt{a})$, so we see that in Step 9, t will be $O(\sqrt{a})$. Thus while the unimproved method only reduces the size of ab (measured in bits, say) at a rate linear in the number of steps, in the improved method the size is reduced quadratically. One expects that the number of digits in ab should be roughly halved with each iteration. We give an example in the next section.

The rest of the procedure (steps 9 and 10) is identical, with or without the lattice reduction step 8. The formula in step 10 comes from the multiplicativity of the norm from $\mathbb{Q}(\sqrt{a})$ to $\mathbb{Q}$:

$$(x_0 + z_0\sqrt{a})(x_1 + z_1\sqrt{a}) = (x_0x_1 + az_0z_1) + (x_0z_1 + z_0x_1)\sqrt{a},$$

and hence

$$\begin{aligned} b(t_1t_2y_1)^2 = (bt_1t_2^2)(t_1y_1^2) &= (x_0^2 - az_0^2)(x_1^2 - az_1^2) \\ &= (x_0x_1 + az_0z_1)^2 - a(x_0z_1 + z_0x_1)^2. \end{aligned}$$

Note that in Step 9, it is not really necessary to factorize t, since t_1 need not be square-free in Step 10; but since solving the reduced equation in Step 10 will first involve factorizing t_1 to solve the congruence $X^2 \equiv a \pmod{t_1}$, there is no time lost in finding this square-free decomposition immediately.

The square-free decomposition is the main time-consuming step in the algorithm, together with the solution of the subsidiary quadratic congruences in Step 7. It involves factorization of the numbers t which arise during the course of the computation, but which need not be related in any direct way to the coefficients of the original equation. We have developed a way of avoiding this factorization altogether, which will be described below in Algorithm II. Starting with a solubility certificate, either the solubility certificate at each level will determine a solubility certificate at the next level (which immediately gives the solution to the quadratic congruence we need), or alternatively a square factor of one of the coefficients will be obtained, which also leads to a reduced problem. See Section 2.5 below.

A similar idea of using 2-dimensional lattice reduction to solve a modular version of our problem was described in the paper [12].

For completeness we give the details of the lattice reduction used in Step 8. Define a positive definite quadratic form on $\mathbb{Z}^2$ by

$$(u, v) \cdot (u', v') = (wu + bv)(wu' + bv') + |a|uu',$$

so that the (square) norm of (u, v) is

$$\|(u,v)\|^2 = (wu + bv)^2 + |a|u^2.$$

Let (u_0, v_0) be the nonzero vector in $\mathbb{Z}^2$ which minimizes this norm. One may find (u_0, v_0) by starting with the standard basis $(1,0)$, $(0,1)$ and applying Gaussian reduction. Then set $(x_0, z_0) = (u_0 w + bv_0, u_0)$: we have

$$x_0^2 - az_0^2 \equiv u_0^2(w^2 - a) \equiv 0 \pmod{b},$$

and $x_0^2 + |a|z_0^2 = \|(u_0, v_0)\|^2$ is minimal.

2.3.1. *Example.* To illustrate the dramatic improvement which the lattice reduction trick (Step 8 of the algorithm) provides in a non-trivial example, we take the equation (4) with $a = -113922743$ and $b = 310146482690273725409$, which occurs in [17]. The unimproved algorithm (omitting Step 8) proceeds with 18 reduction steps and the following sets of coefficients:

$$(a,b) = (-113922743, 310146482690273725409)$$
$$(a,b) = (-113922743, 63228882673342113344) \quad (a,b) = (-5941135, 690379)$$
$$(a,b) = (-113922743, 22155222796709666) \quad (a,b) = (690379, -5941135)$$
$$(a,b) = (-113922743, 13176519068967) \quad (a,b) = (690379, -436439)$$
$$(a,b) = (-113922743, 552039370818) \quad (a,b) = (-436439, 690379)$$
$$(a,b) = (-113922743, 10830811819) \quad (a,b) = (-436439, 52017)$$
$$(a,b) = (-113922743, 52527821) \quad (a,b) = (52017, -436439)$$
$$(a,b) = (52527821, -113922743) \quad (a,b) = (52017, -14)$$
$$(a,b) = (52527821, -5941135) \quad (a,b) = (-14, 52017)$$
$$(a,b) = (-5941135, 52527821) \quad (a,b) = (-14, 942)$$

At this stage, the congruence $X^2 + 14 \equiv 0 \pmod{942}$ yields the solution $x = 92$, and luckily $92^2 + 14 = 942t$, with $t = 9$. As this is a square, we obtain a solution to the equation at this level. Passing back up the stack we finally obtain the following solution to the original equation:

$$\begin{aligned}(x:y:z) = (&17096570497733995340458855914415817266660083175129\\ &: 971656516633305795680905979479465911216\\ &: 676684022080238402700088727243330689433972 29).\end{aligned}$$

By contrast, with the improved method we obtain the following much shorter sequence of coefficients:

$$(a,b) = (-113922743, 310146482690273725409)$$
$$(a,b) = (-113922743, 339)$$
$$(a,b) = (339, -113922743)$$

The last equation has solution $(31006 : 1 : 1781)$ and two back-substitutions lead to the solution $(320832774821087 : 21372 : -18438099853)$ of the original equation, considerably smaller than the previous solution found. Notice the dramatic reduction in the size of b at the first descent step compared with the first solution. The congruence $X^2 \equiv -113922743 \pmod{310146482690273725409}$ has solution $w = -88566846089432467791$, leading to $t = 25291553069336845336$; but then lattice reduction finds the solution $(x_0, z_0) = (824644660421, -93793135)$ to the congruence $X^2 \equiv -113922743Z^2 \pmod{310146482690273725409}$, which yields the much smaller value $t = 5424 = 4^2 \cdot 339$.

2.4. **Improving the solution.** The method of the preceding section will find one solution (x_0, y_0, z_0) to a diagonal equation (3), but this solution is not necessarily "Holzer-reduced". It is possible to reduce the size of a solution. We present two methods for this: the first, due to Mordell in [10, Theorem 5, p.47], is guaranteed to produce a Holzer-reduced solution after a finite number of steps, but is slow in practice since the number of steps appears to be linear in the size of z_0. The second method is based on the quadratic parametrizations which will be introduced in Section 3. This is much faster in practice. The solution it produces is not always Holzer-reduced, as it is only guaranteed to satisfy (10) rather than (9), though in practice it usually is.

As well as applying one or other of these reduction procedures to the solution produced at the end of the recursion, it is also possible to reduce all the intermediate solutions used in the back-substitution Step 10. This can be beneficial in practice for large problems, since otherwise the exponential growth in the size of intermediate solutions can cause serious degradation of the running time owing to the need to work with very large integers.

2.4.1. *Mordell's method for reducing solutions.* Mordell's method is used in [10] to prove that Holzer-reduced solutions always exist. It is not presented there as an algorithm, but is easily turned into one. We refer to [10, Theorem 5, p.47] for the proof that this method works (as stated in Lemma 2.2 above), giving here only a sketch.

Suppose that we have a primitive solution (x_0, y_0, z_0) to the equation (3), where abc is squarefree, a and b are positive and c is negative, with $|z_0| > \sqrt{ab}$. Since c is squarefree, we have $\gcd(x_0, y_0) = 1$.

If c is even, set $k = c/2$, solve $k = uy_0 - vx_0$ for u and v, and let w be the nearest integer to $-(aux_0 + bvy_0)/(cz_0)$. Then the equations

$$
\begin{aligned}
x &= \frac{1}{k}\left(x_0(au^2 + bv^2 + cw^2) - 2u(aux_0 + bvy_0 + cwz_0)\right) \\
y &= \frac{1}{k}\left(y_0(au^2 + bv^2 + cw^2) - 2v(aux_0 + bvy_0 + cwz_0)\right) \qquad (12) \\
z &= \frac{1}{k}\left(z_0(au^2 + bv^2 + cw^2) - 2w(aux_0 + bvy_0 + cwz_0)\right)
\end{aligned}
$$

define integers x, y, z which also satisfy (3), and which satisfy $0 < |z| < |z_0|$. (These follow easily from the identity $(aux_0 + bvy_0 + cwz_0)^2 + ab(uy_0 - vx_0)^2 = -kcz_0z$, together with the inequalities $|aux_0 + bvy_0 + cwz_0| \le \frac{1}{2}|cz_0|$ and $ab < z_0^2$.)

If c is odd, solve $c = uy_0 - vx_0$ for u and v. Now let w be the nearest integer to $-(aux_0 + bvy_0)/(cz_0)$ which has the same parity as $au + bv$. Then (12) with $k = 2c$ again defines an integral solution to (3), with $0 < |z| < |z_0|$.

If the new z is still too big, we apply this again; after a finite number of steps the Holzer bounds (8) will be satisfied.

When we apply this in practice, we may either just apply it once, at the end, or alternatively we may apply it to each solution in the recursive stack before back-substituting at Step 10.

2.4.2. *Reducing solutions via quadratic parametrization.* Starting from a primitive solution (x_0, y_0, z_0) to a diagonal equation (3) we apply the method of Section 3 below to obtain three parametrizing quadratic polynomials $Q_i(U, V)$ with respective discriminants $-4bc$, $-4ac$ and $-4ab$. As shown in the proof of Corollary 3.2 below, after applying Gaussian reduction to whichever one of the $\pm Q_i$ is a positive definite quadratic form, we obtain a parametrization whose leading coefficients give a solution (x_1, y_1, z_1) to (3) satisfying the "almost-Holzer" bounds (21). See Section 3 below for details.

When we apply this in practice, we have found that in most cases, the new solution obtained does in fact satisfy the Holzer bounds. Exceptional cases arise when the root of the reduced positive definite quadratic, which lies in the usual fundamental domain for $\mathrm{SL}(2,\mathbb{Z})$ in the upper half-plane, has imaginary part less than one.

For example, the equation $X^2+3Y^2=91Z^2$ has the solution $(19,1,2)$ which is not Holzer-reduced since $x^2=361>273=3\cdot 91$ and $z^2=4>3=1\cdot 3$. The parametrizing quadratics are $X=19U^2-16UV-11V^2$, $Y=U^2-20UV+9V^2$ and $Z=2(U^2-UV+V^2)$, with minimal discriminants $1092=-4\cdot 3\cdot(-91)$, $364=-4\cdot 1\cdot(-91)$ and $-12=-4\cdot 1\cdot 3$ respectively. The latter is positive-definite and reduced, with root $(-1+i\sqrt{3})/2$, and this method cannot reduce the solution further. However, applying Holzer's method once to $(19,1,2)$ gives the Holzer-reduced solution $(4,5,1)$. The corresponding parametrizing quadratics are $X=4U^2+30UV-12V^2$, $Y=5U^2-8UV-15V^2$ and $Z=U^2+3V^2$.

Note also that the leading coefficients of the reduced parametrizing quadratics need not necessarily be coprime, so the solution (x_1,y_1,z_1) may not be primitive; if not, we obviously obtain a further reduction by cancelling the common factor.

2.4.3. *Example.* Continuing our earlier example, the solution $(320832774821087 : 21372 : -18438099853)$ is not Holzer-reduced, but applying Mordell reduction once yields the Holzer-reduced solution $(30106379962113 : 7913 : 12747947692)$.

The much larger solution produced by the unimproved algorithm requires 27 steps of Mordell reduction to obtain the Holzer-reduced solution $(47464775475069 : 3131 : 2629196804)$.

Using the quadratic parametrization method to reduce the solutions, we obtain the new solutions $(7523107023591 : 7244 : 11931641701)$ (starting from the smaller original solution) and $(70647575606369 : 5679 : 6632499416)$ (starting from the larger). These solutions are both Holzer-reduced.

Note that, as illustrated by these examples, there is nothing at all canonical in the solutions obtained, even amongst those which satisfy the Holzer bounds. The solution obtained will depend on all the choices of modular square roots made along the way, each such choice being equally valid, and leading to a distinct solution. In fact, one remarkable feature of Holzer's theorem (apparent from its proof) is that it not only guarantees one reduced solution, but one in each class of modular solutions modulo abc, the number of which is around 2^k where k is the number of odd prime factors of abc.

2.5. **Algorithm II: factorization-free reduction method.** The preceding algorithm is adequate for solving equations where the coefficients are of "reasonable" size: reasonable in the sense that numbers of this size may be factorized quickly. But for larger problems, the time taken for intermediate factorizations make it impractical. For example, if we take the coefficients in (3) to be primes of around 100 digits (chosen so that (3) is soluble) then in the second step of the recursion one is likely to have to factorize a random integer with between 90 and 100 digits.

To avoid this, we have developed an alternative method which is quite similar in theory but avoids all factorization. The idea is that given a solubility certificate (k_1,k_2,k_3) for the diagonal equation (3) with coefficients (a,b,c), we can use it to construct a new solubility certificate for a reduced problem with smaller coefficients, without any (further) factorization, together with a linear transformation mapping solutions of the reduced problem to solutions of the original. While this idea is simple in principle, complications arise in practice, since at the general stage we cannot assume that the various triples of integers which arise as coefficients are square-free.

One starts with the data $(a, b, c; k_1, k_2, k_3)$ and recursively reduces this to a similar set of data which is smaller, in a suitably defined sense, until one reaches an easy case where the solution can be written down immediately or found directly. The coefficients a, b and c are assumed to be pairwise coprime, but not necessarily square-free. The essential idea is that whenever certain numbers which would be coprime in the square-free case are encountered and found not to be coprime, the common factor can be used to reduce the problem by giving a non-trivial square factor of one of a, b or c, which may then be divided out.

At each stage we take care to provide a solution which lies in the lattice defined by the current solubility certificate, so that at the end the solution obtained lies in the lattice defined by the original certificate.

The next lemma will deal with the base cases under this scheme.

Lemma 2.3. *If two of the coefficients of equation* (3) *are* ± 1*, then a solution in the solution lattice may be found from a solubility certificate in time* $O(\log |abc|)$*.*

Proof. By symmetry we may assume that $ab = \pm 1$. By a changing the sign of all three coefficients, and also of the solubility certificate (in order to keep the solution lattice unchanged) we may also assume that $a = 1$. The certificate consists of an integer $k = k_3$ satisfying $k^2 \equiv -b = \pm 1 \pmod{c}$.

If $a = 1$ and $b = -1$ then we have the trivial solution $(1, 1, 0)$, but we must find a solution satisfying $x \equiv ky \pmod{c}$ where k is a fixed square root of $+1$ modulo c. If $k \equiv \pm 1 \pmod{c}$ we simply use $(1, \pm 1, 0)$. Otherwise, let $c^+ = \gcd(k-1, c)$ and $c^- = \gcd(k+1, c)$ and set $z = c^+c^-/c$. One may check that in all cases $z = \pm 1$ or $z = \pm 2$; this is straightforward when c is square-free but needs a little care when $4 \mid c$. Now the required solution is $(x, y, z) = (\frac{1}{2}(c^- - zc^+), \frac{1}{2}(c^- + zc^+), z)$ which duly satisfies $x^2 - y^2 + cz^2 = 0$ and $x \equiv ky \pmod{c}$.

Now assume that $a = b = 1$, and so $c < 0$. Let $x + yi = \gcd(k + i, c)$ in the Euclidean ring $\mathbb{Z}[i]$ of Gaussian integers. Then it is easy to see (by considering the prime factorization of c) that $x^2 + y^2 = |c|$, so that $(x, y, 1)$ is a solution to (3), and satisfies $x \equiv ky \pmod{c}$ as required. The gcd may be computed in $O(\log |c|)$ steps by [13]. □

The general reduction step will start with a triple of coefficients (a, b, c), pairwise coprime but not necessarily square-free, defining an equation (3), together with a solubility certificate (k_1, k_2, k_3). We then construct a new equation

$$a'(X')^2 + b'(Y')^2 + c'(Z')^2 = 0, \tag{3$'$}$$

with smaller coefficients (a', b', c'), a new solubility certificate (k_1', k_2', k_3'), and a linear map T from the new solution lattice $\mathcal{L}'$ to $\mathcal{L}$, mapping solutions to $(3)'$ to solutions to (3).

During the main reduction step, it can happen that we find a non-trivial square factor u^2 of one of the coefficients. The following trivial lemma may then be used to reduce the problem; however, it is not possible to do so in such a way as to preserve the solution lattice. For this reason we do not in fact use this lemma in our implementation.

Lemma 2.4. *Let* (k_1, k_2, k_3) *be a solubility certificate for* (3)*, where the coefficients* (a, b, c) *are pairwise coprime. Suppose that* $u^2 \mid a$ *for some integer* $u > 1$*. Let* u' *be an inverse of* u *modulo* bc*. Then* $(k_1, u'k_2, u'k_3)$ *is a solubility certificate for the equation with coefficients* $(a/u^2, b, c)$*. Also, if* (x', y', z') *is a solution to the latter equation, then* (x', uy', uz') *is a solution to the original equation.*

Proof. Trivial. □

As an example to show that this lemma cannot be strengthened to respect the solution lattices in all cases, take the equation $p^2X^2 + Y^2 = Z^2$ with certificate $(k_1, 0, 0)$ satisfying $k_1^2 \equiv 1 \pmod{p^2}$, and take $u = p$. Then $u' = p$ also, and the reduced equation is $(X')^2 + (Y')^2 = (Z')^2$ with the same certificate $(k_1, 0, 0)$. The new solution lattice is the whole of $\mathbb{Z}^3$; given a solution (x', y', z') satisfying $(x')^2 + (y')^2 = (z')^2$ and the vacuous condition $y' \equiv k_1 z' \pmod 1$, the solution to the original equation is $(x, y, z) = (x', py', pz')$. This satisfies $y - k_1 z \equiv 0 \pmod p$, but not $0 \pmod{p^2}$.

The following lemma is crucial: it shows that we may find partial factorizations of any finite set of nonzero integers which approximate a full square-free decomposition, using only the operations of gcd and exact integer division.

Lemma 2.5. *Let a_i for $1 \le i \le n$ be nonzero integers. There exist integers b_i for $1 \le i \le n$, and pairwise coprime integers c_I indexed by the nonempty subsets $I \subseteq \{1, 2, \ldots, n\}$, such that for $1 \le i \le n$ we have*

$$a_i = b_i^2 \prod_{I,\ i \in I} c_I. \tag{13}$$

Moreover, these integers may be computed from the a_i using only the operations of gcd *and exact integer division, in $O(\sum \log(a_i))$ steps.*

Proof. We initialize by setting each $b_i = 1$, $c_{\{i\}} = a_i$, and the other $c_I = 1$. Then (13) is satisfied, but the coprimality conditions may not hold.

If $\gcd(c_I, c_J) = d > 1$ for two subsets I and J, we divide c_I and c_J by d, multiply c_{I+J} by d (where $I + J$ is the symmetric difference $(I \cup J) - (I \cap J)$), and multiply b_i by d for all $i \in I \cap J$. This preserves the relations (13) while decreasing the product of all the c_I by a factor of d. Hence, after a finite number of steps we achieve the conditions stated. □

Remark. Of course, without the last sentence of its statement, the lemma would be trivial, using the prime factorizations of the a_i, and we could even require that the c_I should be square-free. A useful trick in practice is to use a small amount of trial division at the start: to ensure that the c_I are not divisible by the square of any prime $p \le p_0$, say, we may (for each such p in turn) divide out the largest even power of p from $c_{\{i\}}$ and adjust b_i accordingly.

We will apply this lemma with $n = 3$ below.

Now we come to the main reduction step, which constructs a new reduced equation together with a solubility certificate, and an appropriate linear transformation.

Proposition 2.6. *Given data $(a, b, c; k_1, k_2, k_3)$ with (k_1, k_2, k_3) a solubility certificate for* (3) *and $|bc| > 1$, with (a, b, c) pairwise coprime and not all of the same sign. There is an algorithm, requiring no factorization, which* either *finds a non-trivial square factor of one of a, b, or c,* or *constructs a smaller set of data $(a', b', c'; k_1', k_2', k_3')$ such that (k_1', k_2', k_3') is a solubility certificate for the equation* $(3)'$*, together with a linear transformation from solutions of this equation to solutions of* (3).

Remark. When we use the algorithm described below, the situation where we fail to construct a reduced equation only arises when one of the coefficients is not square-free. At the top level we will insist that the coefficients are square-free, so this cannot happen there; this is reasonable, since the criterion for solubility given in Lemma 2.1 requires square-free coefficients. At lower levels, if we identify a non-trivial square factor of one of the coefficients, then rather than deal with this using Lemma 2.4, we instead pass the factor we have found back to the level above.

Proof. We will subdivide the proof into a number of steps.

Step 1: Preliminaries. Let $w \equiv c^{-1}k_1 \equiv -bk_1^{-1} \pmod{a}$. Consider the sublattice of $\mathbb{Z}^2$ defined by the congruence $y \equiv wx \pmod{a}$, with $\mathbb{Z}$-basis $(1,w)$, $(0,a)$, together with the weighted Euclidean norm $||(x,y)|| = |b|x^2 + |c|y^2$. Let (w_1,w_2) be a minimal nonzero vector in this lattice (obtained by Gaussian reduction). Then

$$k_1w_1 \equiv cw_2, \qquad k_1w_2 \equiv -bw_1 \pmod{a},$$

so that

$$bw_1^2 + cw_2^2 = at \tag{14}$$

with t a small integer. Explicitly, $0 < ||(w_1,w_2)|| \le \frac{4}{\pi}|a|\sqrt{|bc|}$, so we have $|t| \le \frac{4}{\pi}\sqrt{|bc|}$. Note that by minimality of the vector (w_1,w_2), we know that $\gcd(w_1,w_2)$ has no prime factors p which do not divide a, for then $p^2 \mid t$ and we could divide out a factor p^2 from (14). In particular, $\gcd(w_1,w_2)$ is coprime to bc.

To ease notation we use the abbreviations $(u,v) = \gcd(u,v)$, and write $u \perp v$ to mean $\gcd(u,v) = 1$.

Step 2: The reduced coefficients. Using Lemma 2.5, applied to the three integers bc, a, t, and using the fact that $a \perp bc$, we may write

$$bc = \alpha^2 b'c', \qquad a = \beta^2 n_1 n_3, \qquad t = \gamma^2 n_2 n_3 c'$$

where the integers n_1, n_2, n_3, b' and c' are pairwise coprime. If $|\alpha| > 1$ then either $u = (\alpha,b)$ or $u = (\alpha,c)$ gives a non-trivial square divisor u^2 of b or c respectively, and we may stop. So we may assume $\alpha = 1$. Similarly, we may assume $\beta = 1$, since otherwise we have a non-trivial square factor of a.

Hence the above equations simplify to

$$bc = b'c', \qquad a = n_1n_3, \qquad t = \gamma^2 a'c'$$

where $a' = n_2n_3$. Both triples (a',b',c') and (n_1,n_2,n_3) are pairwise coprime. Moreover, a', b' and c' cannot all have the same sign, since then $t > 0$ and $bc > 0$; but then (14) would imply that a, b, c all had the same sign.

Define $d_1 = (c,c')$ and $d_2 = (b,c')$. Then $c' = \pm d_1d_2$ with $d_1 \mid c$, $d_2 \mid b$, and $(d_1,d_2) = 1$. Adjust the sign of d_1 or d_2 if necessary so that $c' = d_1d_2$.

Remark. When we call this Proposition recursively with the reduced coefficients, it will return either a solution to the reduced equation or a non-trivial square factor of one of a', b' or c'. In the former case we transform the solution using Step 5 below, returning the result to the level above (or stopping if we are already at the top level). If we obtain a factor f^2 of a', then we divide a' by f^2 and multiply γ by f and repeat the process at this level. Similarly if we obtain a square factor of c'. Finally, if we obtain a non-trivial square factor f^2 of b', then at least one of $\gcd(f,b)$ and $\gcd(f,c)$ will be greater than 1, and can be passed back as a non-trivial square factor of b or c respectively. This final possibility cannot happen at the top level, where the coefficients are square-free. Clearly the number of these "back-tracking" steps will be finite.

Step 3: Refinements. In this step we show that various coprimality and divisibility conditions can be assumed between these variables, since otherwise non-trivial square factors of a, b or c are found. This will enable us to construct the new solubility certificate in Step 4.

- $d_i \mid w_i$ for $i = 1,2$:
  For $d_i \mid c' \mid t$, hence (14) implies that $d_i \mid w_i^2$. Let $e = (d_i,w_i)$, and write $d_i = eu$ and $w_i = ev$ with $u \perp v$. Then $d_i \mid w_i^2 \implies u \mid ev^2 \implies u \mid e \implies u^2 \mid d_i$. This gives a non-trivial square factor of either b or c, unless $u = 1$, so we may assume that $u = 1$ and deduce that $d_i \mid w_i$ for $i = 1,2$.

Now we may divide by $c' = d_1 d_2$ in (14) to obtain

$$d_1 \frac{b}{d_2}\left(\frac{w_1}{d_1}\right)^2 + d_2 \frac{c}{d_1}\left(\frac{w_2}{d_2}\right)^2 = aa'\gamma^2. \tag{15}$$

- $(b, \gamma) = (c, \gamma) = 1$:
  For let $d = (b, \gamma)$. Then $d \perp c$, so (14) implies that $d \mid w_2^2$. As before, this implies $d \mid w_2$ (else we obtain a non-trivial square factor of b). Then (14) implies $d^2 \mid bw_1^2$. But $(b, w_1, w_2) = 1$ by the remarks made in Step 1, so $d^2 \mid b$ and we have a square factor of b unless $d = 1$. Similarly, $(c, \gamma) = 1$, else we have a square factor of c.
- $(d_2, w_1) = (d_1, w_2) = 1$:
  For let $d = (d_2, w_1)$. Then $d \perp d_1$, so $d \mid (w_1/d_1)$. Now (15) implies $d \mid aa'\gamma^2$. But d is coprime to each of a, γ and a', since $d \mid d_2 \mid b \mid b'c'$, so $d = 1$. Similarly, $(d_1, w_2) = 1$.
- $(w_1, n_2) = (w_2, n_2) = 1$:
  For let $d = (w_1, n_2)$. Then $d \mid cw_2^2$ from (14), but $d \perp c$ since $n_2 \perp bc$, so $d \mid w_2^2$. Now a prime divisor p of d would divide both w_1 and w_2, but not divide $a = n_1 n_3$, contradicting the observation made below (14). Hence $d = 1$. The proof that $(w_2, n_2) = 1$ is similar.

Step 4: The new certificate. Next we define the new solubility certificate (k_1', k_2', k_3') for the equation with coefficients $a' = n_2 n_3$, $b' = (c/d_1)(b/d_2)$, $c' = d_2 d_1$ as follows.

$$k_1' = \begin{cases} -bw_1 w_2^{-1} & \pmod{n_2} \\ -k_1 & \pmod{n_3} \end{cases} \tag{16}$$

$$k_2' = \begin{cases} (a\gamma)^{-1} k_3 w_1 & \pmod{c/d_1} \\ (a\gamma)^{-1} k_2 w_2 & \pmod{b/d_2} \end{cases} \tag{17}$$

$$k_3' = \begin{cases} k_2 a' \gamma w_1^{-1} & \pmod{d_2} \\ -k_3 a' \gamma w_2^{-1} & \pmod{d_1} \end{cases} \tag{18}$$

This uniquely defines the k_i' modulo a', b', c' respectively by the Chinese Remainder Theorem, since $(n_2, n_3) = (c/d_1, b/d_2) = (d_1, d_2) = 1$. By Step 3, all the modular inverses in these formulae exist.

To check that we do indeed have a solubility certificate is now straightforward; each of the required quadratic congruences is proved in two steps using the factorization of the relevant modulus. Note that at present the signs of the k_i' are immaterial; but they will be important in Step 5.

- $(k_1')^2 + b'c' \equiv 0 \pmod{a'}$: First, modulo n_2 we have
  $$(k_1')^2 + b'c' \equiv (bw_1 w_2^{-1})^2 + bc \equiv bw_2^{-2}(bw_1^2 + cw_2^2) \equiv 0,$$
  since $n_2 | (bw_1^2 + cw_2^2)$; note that it also follows that $k_1' \equiv cw_2 w_1^{-1} \pmod{n_2}$. Modulo n_3 we have
  $$(k_1')^2 + b'c' \equiv k_1^2 + bc \equiv 0,$$
  since $n_3 | a$.
- $(k_2')^2 + a'c' \equiv 0 \pmod{b'}$: First, modulo c/d_1 we have
  $$\begin{aligned}(k_2')^2 + a'c' &\equiv (a\gamma)^{-2} k_3^2 w_1^2 + a'c' \\ &\equiv (a\gamma)^{-2}(k_3^2 w_1^2 + a^2 t) \\ &\equiv a^{-1}\gamma^{-2}(-bw_1^2 + at) \equiv 0,\end{aligned}$$

while modulo b/d_2 we have

$$\begin{aligned}(k_2')^2 + a'c' &\equiv (a\gamma)^{-2}k_2^2w_2^2 + a'c' \\ &\equiv a^{-1}\gamma^{-2}(-cw_2^2 + at) \equiv 0.\end{aligned}$$

- $(k_3')^2 + a'b' \equiv 0 \pmod{c'}$: Observe that the equation (14) implies that

$$(c/d_1)w_2^2 \equiv aa'd_2\gamma^2 \pmod{d_1}$$

on dividing by d_1 and then reducing modulo d_1. Now, modulo d_1 we have

$$\begin{aligned}d_2w_2^2((k_3')^2 + a'b') &\equiv d_2(k_3a'\gamma)^2 + a'b'd_2w_2^2 \equiv -d_2ab(a'\gamma)^2 + a'b'd_2w_2^2 \\ &\equiv -a'b(c/d_1)w_2^2 + a'b'd_2w_2^2 \equiv a'w_2^2(b'd_2 - bc/d_1) \\ &\equiv 0,\end{aligned}$$

since $bc/d_1 = b'c'/d_1 = b'd_2$. This implies that $(k_3')^2+a'b' \equiv 0 \pmod{d_1}$ since $(d_2w_2^2, d_1) = 1$. A similar calculation shows that $(k_3')^2 + a'b' \equiv 0 \pmod{d_2}$.

Step 5: The linear transformation. Recall that we defined in (7) a lattice $\mathcal{L} = \mathcal{L}(a, b, c; k_1, k_2, k_3)$ associated to equation (3) and its solubility certificate. Let $\mathcal{L}' = \mathcal{L}(a', b', c'; k_1', k_2', k_3')$ be the similarly defined lattice for the reduced data. We now define a linear transformation $T : \mathcal{L}' \to \mathcal{L}$, which maps solutions to the new equation into solutions to the original.

Given $(x', y', z') \in \mathcal{L}'$, set $T(x', y', z') = (x, y, z)$ where

$$\begin{aligned}x &= -\gamma n_3 x'; \\ y &= \frac{1}{n_2}\left(\frac{c}{d_1}\frac{w_2}{d_2}y' + w_1z'\right); \\ z &= \frac{1}{n_2}\left(\frac{b}{d_2}\frac{w_1}{d_1}y' - w_2z'\right).\end{aligned} \tag{19}$$

Assuming for the moment that $y, z \in \mathbb{Z}$ and that $T(\mathcal{L}') \subseteq \mathcal{L}$, direct calculation shows that

$$n_2^2\left(ax^2 + by^2 + cz^2\right) = aa'\gamma^2\left(a'(x')^2 + b'(y')^2 + c'(z')^2\right).$$

Hence T maps solutions of the equation $a'(x')^2 + b'(y')^2 + c'(z')^2 = 0$ to solutions of the equation $ax^2 + by^2 + cz^2 = 0$. Nontrivial solutions are mapped to nontrivial solutions, since T has nonzero determinant; specifically, another direct calculation shows that $|\det T| = (\gamma n_3)^3 n_1/n_2 \neq 0$.

Note that in general $T(\mathcal{L}') \neq \mathcal{L}$, since

$$\begin{aligned}[\mathcal{L} : T(\mathcal{L}')] &= [\mathbb{Z}^3 : \mathcal{L}'][\mathcal{L}' : T(\mathcal{L}')]/[\mathbb{Z}^3 : \mathcal{L}] \\ &= a'b'c'(\gamma n_3)^3 n_1/(n_2abc) \\ &= \gamma n_3^3.\end{aligned}$$

If t were square-free and coprime to a, we would have $\gamma = n_3 = 1$. In this situation, which is usually the case in practice, we do have $T(\mathcal{L}') = \mathcal{L}$.

It remains to show that (19) does define a well-defined map from $\mathcal{L}'$ to $\mathcal{L}$.

- $y, z \in \mathbb{Z}$: Since $n_2y \in \mathbb{Z}$ and $n_2 \perp c'$, it suffices to show that $c'(n_2y) \equiv 0 \pmod{n_2}$. But modulo n_2 we have

$$c'(n_2y) \equiv cw_2y' + c'w_1z' \equiv w_1(k_1'y' + c'z') \equiv 0,$$

since $k_1'y' + c'z' \equiv 0 \pmod{a'}$ (from the definition of $\mathcal{L}'$) and $n_2|a'$, and we have used $cw_2 \equiv k_1'w_1 \pmod{n_2}$. The verification that $z \in \mathbb{Z}$ is similar.

- $by \equiv k_1 z \pmod{a}$: using $n_2 \perp a$ and $c' \perp a$, we compute modulo a:

$$\begin{aligned} n_2 c'(k_1 z - by) &\equiv k_1(bw_1 y' - c'w_2 z') - b(cw_2 y' + c'w_1 z') \\ &\equiv by'(k_1 w_1 - cw_2) - c'z'(k_1 w_2 + bw_1) \equiv 0. \end{aligned}$$

- $cz \equiv k_2 x \pmod{b}$: Here it suffices to work modulo d_2 and b/d_2 separately, since they are coprime, and we may multiply by n_2 since $n_2 \perp b$.

  Modulo d_2, we have

$$\begin{aligned} n_2(k_2 x - cz) &\equiv -a'k_2\gamma x' - c\left(\frac{b}{d_2}\frac{w_1}{d_1}y' - w_2 z'\right) \\ &\equiv -a'k_2\gamma x' - b'w_1 y' \qquad (\text{since } d_2 | w_2) \\ &\equiv -a'k_2\gamma x' + w_1 k_3' x' \qquad (\text{since } k_3' x' \equiv -b'y' \pmod{c'} \text{ and } d_2|c') \\ &\equiv 0 \qquad (\text{since } k_3' w_1 \equiv k_2 a' \gamma). \end{aligned}$$

  Modulo b/d_2:

$$\begin{aligned} n_2(k_2 x - cz) &\equiv -a'k_2\gamma x' + cw_2 z' \\ &\equiv k_2' z' k_2 \gamma + cw_2 z' \qquad (\text{since } -a'x' \equiv k_2' z' \pmod{b'} \text{ and } (b/d_2)|b') \\ &\equiv a^{-1} k_2^2 w_2 z' + cw_2 z' \\ &\equiv a^{-1} w_2 z'(k_2^2 + ac) \equiv 0. \end{aligned}$$

- $ax \equiv k_3 y \pmod{c}$: We work modulo d_1 and c/d_1 separately.

  Modulo d_1, we have $ax \equiv k_3 y \iff k_3 x \equiv -by$, and

$$\begin{aligned} n_2(k_3 x + by) &\equiv -k_3 a'\gamma x' + b\left(\frac{c}{d_1}\frac{w_2}{d_2}y'\right) \\ &\equiv w_2(k_3' x' + b'y') \equiv 0. \end{aligned}$$

  Modulo c/d_1:

$$\begin{aligned} n_2(k_3 y - ax) &\equiv k_3(w_1 z') + a\gamma a' x' \\ &\equiv a\gamma(k_2' z' + a'x') \equiv 0. \end{aligned}$$

□

When implementing this method, we first factorize the coefficients of the given diagonal equation, removing square factors and common factors of the coefficients. Then we compute a solubility certificate, returning failure if none exists. A recursive procedure based on Lemma 2.3 and Proposition 2.6 is used to find a solution in the solution lattice defined by the certificate. Finally, the solution is reduced in size using the algorithms of Section 2.4.2 and (if necessary) Section 2.4.1.

This completes the description of the factorization-free reduction method.

2.6. **Algorithm III: lattice methods.** Both the preceding algorithms use 2-dimensional lattice reduction. One can also use the 3-dimensional lattice $\mathcal{L} = \mathcal{L}(a, b, c; k_1, k_2, k_3)$ (defined in (7)) directly as follows. As already observed, for $(x, y, z) \in \mathcal{L}$ we have $f(x, y, z) = ax^2 + by^2 + cz^2 \equiv 0 \pmod{abc}$. Moreover, Minkowski's Theorem implies that $\mathcal{L}$ contains a nonzero vector (x, y, z) satisfying Holzer's bounds (8). This implies that

$$-|abc| < f(x, y, z) < 2|abc|,$$

so that either $f(x, y, z) = 0$ or $f(x, y, z) = |abc|$. In the former case, we have a Holzer-reduced solution to (3), but in the latter case we do not have a solution. Various ways of fixing this problem have been proposed, by Mordell, Cassels and more recently by Cochrane and Mitchell in [3]. They impose extra 2-adic conditions to define a sublattice $\mathcal{L}'$ of index 2 in $\mathcal{L}$, such that points $(x,y,z) \in \mathcal{L}'$ satisfy $f(x,y,z) \equiv 0 \pmod{2abc}$, and apply a theorem of Gauss to assert the existence of a point $(x,y,z) \in \mathcal{L}'$ with $|f(x,y,z)| < 2|abc|$, giving a solution. The case $a = b = 1$ requires special treatment in the proof, as in Lemma 2.3.

In order to turn this into an algorithm for solving equations in practice, one needs methods of finding short vectors in 3-dimensional lattices, since the shortest vector in $\mathcal{L}'$ certainly gives a solution. In most cases, the first vector in an LLL-reduced basis of $\mathcal{L}$ gives a solution, and the following lemma[1] says that one does not have to look much further.

Lemma 2.7. *Let b_1, b_2, b_3 be an LLL-reduced basis of a 3-dimensional lattice $\mathcal{L}$. Then the shortest vector of $\mathcal{L}$ has the form $n_1b_1 + n_2b_2 + n_3b_3$ where each $n_i \in \{-1, 0, 1\}$.*

Remark. Since $\pm v$ have the same length, this leaves us with 13 nonzero vectors to check to find the shortest vector, given an LLL-reduced basis.

Proof. Let b_i^* for $i = 1, 2, 3$ be the orthogonalized basis vectors in $\mathbb{R}^3$, so that

$$\begin{aligned} b_1 &= b_1^*, \\ b_2 &= b_2^* + \mu_{21}b_1^*, \\ b_3 &= b_3^* + \mu_{31}b_1^* + \mu_{32}b_2^*, \end{aligned}$$

Since the b_i are LLL-reduced, we have $|\mu_{ij}| \leq 1/2$ for $1 \leq j < i \leq 3$, and

$$|b_i^*|^2 \geq \left(\frac{3}{4} - \mu_{i,i-1}^2\right)|b_{i-1}^*|^2 \geq \frac{1}{2}|b_{i-1}^*|^2$$

for $i = 2, 3$. Hence

$$|b_1^*|^2 \leq 2|b_2^*|^2 \leq 4|b_3^*|^2.$$

Let $x = \alpha_1b_1 + \alpha_2b_2 + \alpha_3b_3 \in \mathcal{L}$ with $\alpha_i \in \mathbb{Z}$. Then

$$|x|^2 = (\alpha_1 + \mu_{21}\alpha_2 + \mu_{31}\alpha_3)^2|b_1^*|^2 + (\alpha_2 + \mu_{32}\alpha_3)^2|b_2^*|^2 + \alpha_3^2|b_3^*|^2.$$

Suppose $|x| < |b_1|$. We will show that this forces $|\alpha_i| \leq 1$ for $i = 1, 2, 3$.

First of all, from $\alpha_3^2|b_3^*|^2 \leq |x|^2 < |b_1^*|^2 \leq 4|b_3^*|^2$ we have $\alpha_3^2 < 4$, so $|\alpha_3| \leq 1$.

Case 1: $\alpha_3 = 0$. Then $\alpha_2^2|b_2^*|^2 \leq |x|^2 < |b_1^*|^2 \leq 2|b_2^*|^2$ implies $\alpha_2^2 < 2$, so $|\alpha_2| \leq 1$. If $\alpha_2 = 0$, then $x = \alpha_1b_1$, so $\alpha_1 = 0$. Otherwise, $\alpha_2 = \pm1$, giving $|b_1^*|^2 > |x|^2 \geq (\alpha_1 + \mu_{21}\alpha_2)^2|b_1^*|^2$, which implies $(\alpha_1 \pm \mu_{21})^2 < 1$, so $|\alpha_1| \leq 1$ since $|\mu_{21}| < \frac{1}{2}$ and $\alpha_1 \in \mathbb{Z}$.

Case 2: $\alpha_3 = \pm1$. Now

$$\begin{aligned} |b_1^*|^2 > |x|^2 &\geq (\alpha_2 + \alpha_3\mu_{32})^2|b_2^*|^2 + |b_3^*|^2 \\ &\geq \frac{1}{2}(\alpha_2 \pm \mu_{32})^2|b_1^*|^2 + \frac{1}{4}|b_1^*|^2, \end{aligned}$$

so $(\alpha_2 \pm \mu_{32})^2 \leq \frac{3}{2}$, giving $|\alpha_2| \leq 1$. □

We now give a recipe for bases of the lattices $\mathcal{L}$ and $\mathcal{L}'$, in terms of the solubility certificate (k_1, k_2, k_3).

[1]shown to us by Robin Chapman

Basis for $\mathcal{L}$: Using the fact that a, b, c are pairwise coprime, solve the following for u, v, a' and b':

$$ub + vc = 1; \qquad aa' + bcb' = 1.$$

Now set

$$\alpha \equiv b'ck_1 \pmod{a}; \qquad \beta \equiv ua'bk_3 \pmod{bc}; \qquad \gamma \equiv va'ck_2 \pmod{bc}.$$

The following vectors give a basis for $\mathcal{L}$:

$$v_1 = (bc, 0, 0); \qquad v_2 = (a\beta, a, 0); \qquad v_3 = (\alpha\beta + \gamma, \alpha, 1);$$

for one easily checks that these vectors all satisfy the defining congruences for $\mathcal{L}$, and they evidently generate a lattice of index $|abc|$.

Basis for $\mathcal{L}'$: One easily checks that the map $\epsilon : \mathcal{L} \to \mathbb{Z}/2\mathbb{Z}$ given by $(x, y, z) \mapsto f(x, y, z)/(abc) \pmod 2$ is an additive homomorphism. It is surjective, since the images of $(bc, 0, 0)$, $(0, ac, 0)$ and $(0, 0, ab)$ (which all lie in $\mathcal{L}$) are bc, ac and ab (mod 2) respectively, and at least one of these is odd. Hence $\mathcal{L}' = \{v \in \mathcal{L} \mid f(v) \equiv 0 \pmod{2abc}\}$ is a sublattice of $\mathcal{L}$ of index 2. [Again, we are grateful to R.J.Chapman for this observation.]

Let v_i be a basis vector of $\mathcal{L}$ (from the above list) such that $\epsilon(v_i) = 1 \pmod 2$). Define

$$w_j = \begin{cases} 2v_i & \text{if } j = i; \\ v_j - v_i & \text{if } j \neq i \text{ and } \epsilon(v_j) = 1; \\ v_j & \text{if } j \neq i \text{ and } \epsilon(v_j) = 0. \end{cases}$$

Then w_1, w_2, w_3 is a basis for $\mathcal{L}'$.

Now use a standard integer LLL-algorithm, such as in [4], to find an LLL-reduced basis b_1, b_2, b_3 of $\mathcal{L}'$ with respect to the norm $||(x, y, z)||^2 = |a|x^2 + |b|y^2 + |c|z^2$. Then for at least one of the 13 nonzero vectors $v = n_1b_1 + n_2b_2 + n_3b_3$ (up to sign) we have $f(v) = 0$ by Lemma 2.7.

It would also be possible to use the algorithm of Vallée (see [16]) for finding the shortest vector in the 3-dimensional lattice $\mathcal{L}'$. We have not implemented this.

2.7. **Other methods.** Finally we mention that there are two methods for solving Legendre's equation due to Gauss: see [7, Arts. 294, 295]. These both involve the theory of reduction of ternary quadratic forms: specifically, in both solutions one constructs an indefinite ternary form of determinant -1 and reduces it to the form $x^2 + 2yz$ using a suitable unimodular substitution. While Gauss does give an algorithm for this reduction in [7, Arts. 272, 274], it does not seem to be very efficient in practice. Without a fast method of carrying out such a reduction, Gauss's methods of solving Legendre's equation are much slower than the method we presented above.

3. Parametric Solutions

Now we have one solution (x_0, y_0, z_0) to our equation (1), and we wish to parametrize all solutions. Our starting point is a classical method (see [10]), which was also used in [6] and may also be found in the book [15].

3.1. **The Diagonal Equation.** First assume that our equation is in diagonal form (3) with abc square-free. Assuming that $z_0 \neq 0$ by symmetry, one sets $X = x_0W + U$, $Y = y_0W + V$, $z = z_0W$ and eliminates W to obtain the following parametric solution:

$$\begin{aligned} x &= Q_1(U,V) = ax_0U^2 + 2by_0UV - bx_0V^2, \\ y &= Q_2(U,V) = -ay_0U^2 + 2ax_0UV + by_0V^2, \\ z &= Q_3(U,V) = az_0U^2 + bz_0V^2. \end{aligned} \tag{20}$$

These quadratics have the following discriminants:

$$\operatorname{disc}(Q_1) = -4bcz_0^2; \qquad \operatorname{disc}(Q_2) = -4acz_0^2; \qquad \operatorname{disc}(Q_3) = -4abz_0^2.$$

Also, the 3×3 matrix of coefficients of the Q_i (which is used in the application in [6]) has determinant $-4abcz_0^3$. We claim that the powers of z_0 which appear here are entirely superfluous and may be removed. This is hardly surprising, since we made an arbitrary choice of the variable Z at the start. But it is significant, since in many of the applications, such as the one in [6] and our own in 2-descent on elliptic curves, it is crucial to keep these quantities as small as possible, and to avoid introducing spurious prime factors. Our result is as follows.

Proposition 3.1. *Let a, b, c be nonzero integers with abc squarefree, such that the equation $aX^2+bY^2+cZ^2 = 0$ has a (nontrivial) solution. Then the set of all rational solutions may be parametrized in the form (2) where each $Q_i(U,V) \in \mathbb{Z}[U,V]$ is quadratic, with discriminants*

$$\operatorname{disc}(Q_1) = -4bc, \qquad \operatorname{disc}(Q_2) = -4ac, \qquad \operatorname{disc}(Q_3) = -4ab,$$

and the determinant of the coefficient matrix of the Q_i is $4abc$. Moreover, these discriminants cannot be further reduced.

Proof. We start with the parametrization given by (20) in terms of a primitive solution (x_0, y_0, z_0) with $z_0 \neq 0$. It is sufficient to find an integer e such that $Q_i(U + eV/z_0, V/z_0)$ has integer coefficients for $i = 1, 2, 3$, since this change of variables clearly reduces the discriminants of each Q_i by a factor of z_0^2 as required, and the coefficient determinant by a factor z_0^3.

Since a is squarefree, $\gcd(y_0, z_0) = 1$, so we can find an integer e satisfying

$$ey_0 \equiv x_0 \pmod{z_0^2}.$$

From $ax_0^2 + by_0^2 + cz_0^2 = 0$ it easily follows that $ae^2 \equiv -b \pmod{z_0^2}$, and then

$$\begin{aligned} eax_0 + by_0 &\equiv 0 \pmod{z_0^2}, \\ e^2ax_0 + 2eby_0 - bx_0 &\equiv 0 \pmod{z_0^2}. \end{aligned}$$

Now

$$Q_1(U + eV/z_0, V/z_0) = ax_0U^2 + 2\frac{eax_0 + by_0}{z_0}UV + \frac{e^2ax_0 + 2eby_0 - bx_0}{z_0^2}V^2,$$

which has integral coefficients[2]. A similar check shows that the coefficients of $Q_i(U + eV/z_0, V/z_0)$ are also integral for $i = 2$ and $i = 3$.

For the last statement, observe that since abc is squarefree, the only square dividing all the discriminants $-4ab$, $-4ac$, $-4bc$ is 4. Now $-ab$, $-ac$, and $-bc$ cannot all be discriminants: none is a multiple of 4, and they cannot all be congruent to 1 (mod 4) since their product is $-(abc)^2$. □

[2] In fact, R. Buchholz has observed that it is sufficient for e to satisfy $ey_0 \equiv x_0 \pmod{z_0}$; this may produce smaller coefficients at this stage, but the reduction given in Corollary 3.2 below makes this redundant.

Corollary 3.2. *With the notation as in Proposition 3.1, there exist values* (u_0, v_0) *of the parameters* (U, V) *such that* $\gcd(u_0, v_0) = 1$ *and if we set* $x_1 = Q_1(u_0, v_0)$, $y_1 = Q_2(u_0, v_0)$, $z_1 = Q_2(u_0, v_0)$ *then* $(x_1 : y_1 : z_1)$ *is a solution of* (3) *satisfying the "almost-Holzer" bounds*

$$|x_1| \le \sqrt{4|bc|/3}; \qquad |y_1| \le \sqrt{4|ac|/3}; \qquad |z_1| \le \sqrt{4|ab|/3}. \tag{21}$$

Proof. Assume, without loss of generality, that $a > 0$, $b > 0$ and $c < 0$. Then $Q_3(U, V)$ is definite (and even positive definite if we take $z_0 > 0$, as we may). Applying standard Gaussian reduction to Q_3, we find a unimodular substitution of the parameters (U, V), say $U = \alpha U' + \beta V'$, $V = \gamma U' + \delta V'$ with $\alpha\delta - \beta\gamma = \pm 1$, so that the transformed quadratic $Q_3^*(U', V') = Q_3(U, V)$ has leading coefficient $z_1 = Q_3^*(1, 0) = Q_3(\alpha, \gamma)$ satisfying $z_1^2 \le |\operatorname{disc}(Q_3)/3| = 4ab/3$. Applying the same transformation to $Q_1(U, V)$ and $Q_2(U, V)$ we obtain new parametrizing quadratics Q_i^* satisfying $aQ_1^*(U, V)^2 + bQ_2^*(U, V)^2 + cQ_3^*(U, V)^2 = 0$ and the same discriminants as the Q_i. Substituting $(U, V) = (1, 0)$ we obtain a new solution $x_1 = Q_1^*(1, 0)$, $y_1 = Q_2^*(1, 0)$, $z_1 = Q_3^*(1, 0)$, with $z_1^2 \le 4ab/3$. Finally, $ax_1^2 \le ax_1^2 + by_1^2 = cz_1^2 \le 4|abc|/3$ so that $x_1^2 \le 4|bc|/3$, and similarly $y_1^2 \le 4|ac|/3$. This proves the result, with $(u_0, v_0) = (\alpha, \gamma)$. ☐

3.2. **Example.** We apply the method of the previous section to the equation

$$X^2 + 113922743\,Z^2 = 310146482690273725409\,Y^2$$

treated earlier, starting with the primitive and Holzer-reduced solution $(x, y, z) = (70647575606369, 5679, 6632499416)$. We obtain the parametrization

$$\begin{aligned}
X &= 70647575606369\,U^2 - 272768472153240\,UV - 236838674874023V^2,\\
Y &= 5679\,U^2 - 536\,UV + 20073V^2,\\
Z &= 6632499416\,U^2 + 24254293278UV - 24587834368V^2.
\end{aligned}$$

These have discriminants $4 \cdot 113922743 \cdot 310146482690273725409$, $-4 \cdot 113922743$ and $4 \cdot 310146482690273725409$, as expected. While the size of the coefficients in this parametrization may seem large (up to 15 digits), recall that the coefficients of the original equation have 9 and 21 digits. By comparison, the parametrization given in [17], obtained using Maple, involves coefficients all of which have between 25 and 35 digits; and more seriously, the discriminants of the quadratics given there are k^2 times the ones given above, where $k = 2^5 \cdot 3 \cdot 59 \cdot 67 \cdot 79 \cdot 149 \cdot 1993 \cdot 7187 \cdot 45757 \cdot 16215770450329$.

3.3. **The Semi-Diagonal Equation.** For convenience for our elliptic curve applications, we give an alternative form of the Proposition 3.1 suited to the semi-diagonal form.

Proposition 3.3. *Let* a, b, c, d *be integers with* $acd(b^2 - 4ac) \ne 0$ *and* d *squarefree, such that the equation* $aX^2 + bXZ + cZ^2 = dY^2$ *has a (nontrivial) solution. Then the set of all rational solutions may be parametrized in the form* (2) *where each* $Q_i(U, V) \in \mathbb{Z}[U, V]$ *is quadratic, with discriminants*

$$\operatorname{disc}(Q_1) = 4cd, \qquad \operatorname{disc}(Q_2) = b^2 - 4ac, \qquad \operatorname{disc}(Q_3) = 4ad.$$

Proof. Rather than change variables and apply Proposition 3.1 it is simpler to start from scratch with a primitive solution (x_0, y_0, z_0) to (5). Set $\Delta = b^2 - 4ac$.

We first suppose that $y_0 \neq 0$, which will certainly be the case unless Δ is a square. One parametrization is given by

$$\begin{aligned} X &= Q_1(U,V) = x_0U^2 + 2(bx_0 + 2cz_0)UV + x_0\Delta V^2, \\ Y &= Q_2(U,V) = y_0U^2 - y_0\Delta V^2, \\ Z &= Q_3(U,V) = z_0U^2 - 2(bz_0 + 2ax_0)UV + z_0\Delta V^2, \end{aligned} \tag{22}$$

with $\mathrm{disc}(Q_1) = 16cdy_0^2$, $\mathrm{disc}(Q_2) = 4y_0^2\Delta$, and $\mathrm{disc}(Q_3) = 16ady_0^2$. These must now be divided by $(2y_0)^2$.

The argument is slightly complicated by the fact that we cannot assume that either $\gcd(x_0, y_0) = 1$ or $\gcd(z_0, y_0) = 1$. But $\gcd(x_0, z_0) = 1$ since d is squarefree, so without loss of generality we may assume that x_0 is odd, and we may factorize $y_0 = y_1y_2$ with $\gcd(2y_1, y_2) = \gcd(2y_1, x_0) = \gcd(y_2, z_0) = 1$. Hence by the Chinese Remainder Theorem we may find e satisfying

$$ex_0 \equiv -(2cz_0 + bx_0) \pmod{4y_1^2}, \tag{23}$$

$$ez_0 \equiv (2ax_0 + bz_0) \pmod{y_2^2}. \tag{24}$$

Explicitly, if $sx_0 + tz_0 = 1$ then we may set $e = t(2ax_0 + bz_0) - s(2cz_0 + bx_0) \pmod{4y_0^2}$. In particular, we may compute e in practice without having to determine the factorization $y_0 = y_1y_2$.

Using the fact that $ax_0^2 + bx_0z_0 + cz_0^2 \equiv 0 \pmod{y_0^2}$, simple calculations show that (23) also holds modulo y_2^2 and that (24) also holds modulo $4y_1^2$; hence both hold modulo $4y_0^2$. Also, squaring (23) and using $ax_0^2 + bx_0z_0 + cz_0^2 \equiv 0 \pmod{y_0^2}$ we find that $e^2 \equiv \Delta \pmod{4y_1^2}$, and (24) similarly implies that the same congruence holds modulo y_2^2, so we have $e^2 \equiv \Delta \pmod{4y_0^2}$. Now a trivial calculation shows that the quadratics $Q_i(U + eV/(2y_0), V/(2y_0))$ have integer coefficients and the properties stated.

The case where $y_0 = 0$ may easily be handled: this can only happen when Δ is a square, say $\Delta = \delta^2$. Note that $\delta \equiv b \pmod 2$. We start with the parametrization

$$\begin{aligned} x &= Q_1(U,V) = \frac{1}{2}(ad(\delta - b)U^2 + (\delta + b)V^2), \\ y &= Q_2(U,V) = a\delta UV, \\ z &= Q_3(U,V) = a^2dU^2 - aV^2, \end{aligned}$$

with discriminants $4a^2cd$, $a^2\Delta$ and $4a^3d$ respectively. Write $a = a_1a_2$ where $a_1 = \gcd(a, (\delta + b)/2)$. Then a_2 divides $(\delta - b)/2$, and a simple calculation shows that the quadratics $(1/a_1)Q_i(U/a_2, V)$ have the desired properties. □

4. Timings

We have implemented the algorithms described in Section 2 using C++ together with the LiDIA library (version 1.4) for multiprecision integer arithmetic and factorization routines. Modular square roots were computed using the implementation in LiDIA of Shanks's RESSOL algorithm in order to find certificates. The integer LLL algorithm was implemented by us following [4].

As sample problems we considered only diagonal equations $aX^2 + bY^2 + cZ^2 = 0$ where a, b and $-c$ are primes chosen so that the equation is soluble. The advantage of using prime coefficients is that the factorization-free solution methods then have to do no factorization at all (other than verifying that the coefficients are prime, which was done using a standard pseudo-primality test).

We precomputed sets of test data as follows, for

$$k \in \{5, 10, 15, 20, 25, 50, 75, 100, 125, 150, 175, 200, 500, 1000\}.$$

For each k, let a be the smallest prime above 10^k, and b the next prime after a; then for each of the next primes p after b such that the equation $aX^2 + bY^2 = pZ^2$ is soluble, we store the triple (a, b, c) with $c = -p$. The corresponding data set for each k will be denoted S_k. We precomputed datasets containing 100 triples for $k \le 200$, five triples for $k = 500$ and just one for $k = 1000$. This last one uses coefficients $a = 10^{1000} + 453$, $b = 10^{1000} + 1357$, and $c = -(10^{1000} + 2713)$. We remark that computing these data sets took longer than solving the corresponding equations using our algorithms.

Each of the algorithms was then used to compute (reduced) solutions to each of the equations in each data set. For $k > 20$ it was not practical to use Lagrange reduction (Algorithm I), either with (LAG+R) or without (LAG) the lattice reduction improvement described above. This was partly because of the excessively long time this would have taken, but also because a bug in LiDIA's MPQS factorization routine meant that integers of this size could often not be factored reliably, so that the timings obtained on repeated runs were very inconsistent. Since the coefficients used are prime, no factorization at all was needed for either the factorization-free reduction method (FFR) or the LLL-based methods.

The results are as follows, given in seconds, based on a DEC alpha EV6. Recall that each entry for $k \le 200$ gives the time taken to solve 100 different problems of size around 10^k for data set S_k, the datasets for $k = 500$ and $k = 1000$ having size 5 and 1 respectively.

k	LAG	LAG+R	FFR	LLL
5	35.243s	4.612s	0.407s	0.362s
10	169.764s	11.869s	0.765s	0.737s
15		18.554s	1.139s	1.185s
20		31.978s	2.537s	2.629s
25			2.763s	2.982s
50			7.168s	8.839s
75			13.073s	17.819s
100			21.920s	34.871s
125			30.611s	52.856s
150			40.603s	74.219s
175			57.991s	109.221s
200			73.597s	147.364s
500			32.372s	69.576s
1000			34.031s	79.320s

We may draw the conclusion that methods which do not require factorization at intermediate stages are much faster than those which do. Of the factorization-free methods, LLL and the factorization-free reduction methods are of comparable speed for small and medium-sized problems, but for larger problems the reduction method starts to gain, being twice as fast for the problems with 200 digit coefficients.

The above comparative timings between the FFR and LLL methods are somewhat misleading, however. By using equations with prime coefficients we have avoided all factorization in computing the solutions, but both the FFR and LLL methods start by computing the solubility certificate (k_1, k_2, k_3) for each equation, and this computation takes a substantial proportion of the total time. To investigate further, we isolated the time for this step, which involves the computation of three square roots modulo primes of size 10^k for each equation, and found that it takes almost all the time of the FFR method, and about half the time of the LLL method. In the following table, the first column of timings gives the times for just computing the certificates; the next two columns give the time to find the solutions, given the precomputed certificates using both FFR and LLL methods.

k	Certificate	FFR	LLL
5	0.237	0.181	0.175
10	0.470	0.257	0.329
15	0.735	0.327	0.522
20	1.981	0.404	0.766
25	2.093	0.484	1.025
50	5.801	0.914	3.283
75	10.809	1.411	7.290
100	18.343	2.055	16.372
125	26.485	2.746	25.966
150	35.092	3.475	38.445
175	50.607	4.633	58.593
200	63.534	5.883	83.435
500	30.003	1.888	39.575
1000	30.248	1.983	48.766

It is now apparent that our FFR method is very much faster than LLL in finding the solution from the certificate, by a factor of about 15 in the largest example. We give some more details of this last computation: 10 levels of recursion were needed; at each depth except one, the value of the variable γ is 1, the exceptional value being 5. This agrees with our expectation that $\gamma = 1$ in most cases. With the FFR method, the solution (x, y, z) produced initially had content $\gcd(x, y, z) = 6$, with no cancelling of common factors during the recursion (in order to stay on the appropriate lattice). This small content shows the efficiency of the formulae used to map the solutions back from lower levels. After cancelling this common factor, the non-reduced solution has integers x, y, z each of 1004 digits, with "Holzer measure" $\max\{x^2/|bc|, y^2/|ac|, z^2/|ab|\} = 5.7 \cdot 10^7$. After reduction (using the quadratic parametrization) we obtain the Holzer-reduced solution with integers of 1000 digits each and Holzer measure 0.54. The LLL method produces a solution which has Holzer measure 0.47, and again 1000 digits for each of x, y and z.

In our implementation of Lemma 2.5 we use the technique mentioned in the remark after that Lemma above, to ensure that the factors c_I are not divisible by p^2 for primes $p < 20$. For all the examples, this was sufficient to avoid ever having to backtrack, since (without this adjustment) the only square factors of the coefficients which were discovered at lower levels were products of the primes ≤ 11. A small time saving was achieved in this way.

Analyzing these two algorithms further may throw some light on this marked difference in their running times for large problems. In both cases we start by constructing a 3-dimensional lattice $\mathcal{L}$ in which the solution will lie. With the LLL method, we repeatedly find new bases for this same lattice, while with the FFR method we construct a new lattice at each step, and the only lattice reduction we do is on 2-dimensional projections of these. These successive lattices have decreasing index in $\mathbb{Z}^3$, and their bases have smaller and smaller integer coordinates, so we would expect the computations to be faster as we go deeper into the recursion.

References

1. B. J. Birch and H. P. F. Swinnerton-Dyer, *Notes on elliptic curves II*, J. Reine Angew. Math. **218** (1965), 79–108.
2. J. W. S. Cassels, *Lectures on elliptic curves*, London Mathematical Society Student Texts, no. 24, Cambridge University Press, 1991.
3. T. Cochrane and P. Mitchell, *Small solutions of the Legendre equation*, Journal of Number Theory **70** (1998), 62–66.
4. H. Cohen, *A course in computational algebraic number theory (third corrected printing)*, Graduate Texts in Mathematics, no. 138, Springer-Verlag, 1996.

5. J. E. Cremona, *Higher descents on elliptic curves*, preprint: see `http://www.maths.nott.ac.uk/personal/jec/papers/d2.ps`.
6. I. Gaál, A. Pethő, and M. Pohst, *Simultaneous representation of integers by a pair of ternary quadratic forms–with an application to index form equations in quartic number fields*, Journal of Number Theory **57** (1996), 90–104.
7. C. F. Gauss, *Disquisitiones arithmeticae*, Springer-Verlag, 1986.
8. K. Ireland and M. Rosen, *A classical introduction to modern number theory*, Graduate Texts in Mathematics, no. 84, Springer-Verlag, 1982.
9. B. Mazur, *On the passage from local to global in number theory*, Bull. Amer. Math. Soc. (N.S.) **29** (1993), no. 1, 14–50.
10. L. J. Mordell, *Diophantine equations*, Pure and Applied Mathematics, no. 30, Academic Press, 1969.
11. M. Pohst, *On Legendre's equation over number fields*, Publ. Math. (Debrecen) **56** (2000), 535–546.
12. J. M. Pollard and C. P. Schnorr, *An efficient solution of the congruence* $X^2 + kY^2 \equiv m \pmod{n}$, IEEE Transactions on Information Theory **33** (1987), no. 5, 702–709.
13. H. Rolletschek, *On the number of divisions of the euclidean algorithm applied to gaussian integers*, J. Symb, Comput. **2** (1986), 261–291.
14. D. Simon, *Équations dans les corps de nombres et discriminants minimaux*, Ph.D. thesis, Université Bordeaux I, 1998.
15. N. P. Smart, *The algorithmic resolution of Diophantine equations: a computational cookbook*, London Mathematical Society Lecture Notes Series, no. 117, Cambridge University Press, 1998.
16. B. Vallée, *Algorithmique dans les réseaux de petite dimension: un point de vue affine sur la récherche des minima*, Séminaire de Théorie des nombres de Bordeaux (1985–1986), no. 13.
17. H.-J. Weber, *Algorithmische Konstruktion hyperelliptischer Kurven mit kryptographischer Relevanz und einem Endomorphismenring echt größer als* $\mathbf{Z}$, Ph.D. thesis, Institut für Experimentelle Mathematik, University of Essen, 1997.
18. A. Weil, *Number theory: an approach through history from Hammurapi to Legendre*, Birkhäuser, 1984.
19. J. Wilson, *Curves of genus 2 with real multiplication by a square root of 5*, Ph.D. thesis, University of Oxford, 1998.

School of Mathematical Sciences, University of Nottingham, University Park, Nottingham NG7 2RD, UK.

E-mail address: `John.Cremona@nottingham.ac.uk`

Department of Mathematical Sciences, Northern Illinois University, DeKalb, IL 60115, USA.

E-mail address: `rusin@math.niu.edu`