Pell's Equation and Fundamental Units

Kaisa Taipale
University of Minnesota
Summer 2000

Pell's Equation and Fundamental Units

Pell's equation was first introduced to me in the number theory class at Caltech that I never completed. It was presented in the context of solving Diophantine equations - finding integral solutions to equations. It looks easy: find the smallest positive integer solutions (x, y) to the equation $x^2 - Dy^2 = 1$, with D a non-square positive integer. But althought there are several algorithms for finding solutions, a pattern in the solutions has never been detected. Moreover, values of D which one might expect to produce slightly different solutions can at times vary dramatically. A story might illustrate:

One day I was sitting happily at my computer, answering email, when I got a note that said merely, "The smallest non-trivial integer solution of $x^2 - 94y^2 = 1$ is $x = 2143295$ and $y = 221064$." This intrigued me, for some reason, and I asked what some other solutions were. I was urged to pick a number for D, and I picked 151. Solving these equations using a somewhat-better-than-brute-force method took one minute for $D = 94$ and three hours for $D = 151$, finally resulting in $x = 1728148040$ and $y = 140634693$. Before the solution was found I started to wonder why it was taking so long, and why I'd want to wait so long to hear those numbers, and whether or not there was any way to estimate the size of solutions so that I'd know whether or not I wanted to embark on any such searches in the future.

In response, I was handed a paper by Yoshihiko Yamamoto, published in 1971. Titled, "Real Quadratic Number Fields with Large Fundamental Units," it seemed to have nothing to do with Pell's equation. Upon the asking of some questions, like, "What's a fundamental unit?" and some research, the connection began to get clearer. Although at first glance the fundamental unit has little to do with the equation I saw in my number theory class it turns out that they are intimately connected.

The fundamental unit ϵ of the ring of algebraic integers in a real quadratic number field is a generator of the group of units (mod ± 1). For the subring $\mathbf{Z}[\sqrt{D}]$ (consisting of elements $x + y\sqrt{D}$ with $x, y \in \mathbf{Z}$) of the ring of integers in $\mathbf{Q}(\sqrt{D})$, ϵ is equal to $x + y\sqrt{D}$ where (x, y) is the smallest integer solution for Pell's equation. So finding bounds for the size of the fundamental unit is extremely useful, as when one knows ϵ and D one can find bounds for x and y.

Yamamoto's paper does not address the general subject of finding bounds for ϵ. In his introduction, he remarks upon the fact that there are many works dealing with the estimation of the fundamental unit, but that they primarily deal with fundamental units with "small orders of absolute value in comparison to their discriminants D." His paper deals instead with constructing real quadratic number fields with comparatively large fundamental units. The essence of the work is that one can find a positive constant c_1 such that

$$\log \epsilon > c_1(\log(\sqrt{D})$$

for sufficiently large D.

After the introduction, the paper is divided into four parts. In the first, Yamamoto outlines the idea of reduced quadratic irrationals. In the second, he links these to reduced ideals, which he also defines. Only in the third section does he directly approach the topic of bounds for ϵ. The fourth consists of examples. In outlining Yamamoto's research I will roughly follow his format, adding explanation or other materials wherever necessary.

The first thing Yamamoto addresses is reduced quadratic irrationals. Begin with a quadratic equation, $ax^2 + bx + c = 0$, $a > 0$ and $\gcd(a, b, c) = 1$. Then let α be a root of this equation and let D be the discriminant, equal to $b^2 - 4ac$. Then α is a real number, the root of a quadratic equation, and irrational, because the D here is the same as the D of Pell's equation. Thus the name quadratic irrational. There exists, too, a conjugate to α, denoted α'. We call α *reduced* if $\alpha > 1$ and $0 > \alpha' > -1$.

Every quadratic irrational is *equivalent* to a reduced quadratic irrational, where equivalence between two quadratic irrationals α and β is defined by the relation

$$\alpha = \frac{a\beta + b}{c\beta + d}.$$

Here a, b, c, d are integers and $ad - bc = \pm 1$. The equivalence of α and β means that they have the same discriminant D.

In addition to satisfying the conditions provided above, a quadratic irrational is deemed reduced if its continued fractional expansion is strictly periodic. The continued fractional expansion is found as follows:

$$\alpha = \alpha_1$$

$$\alpha_i = a_i + \frac{1}{\alpha_{i+1}}$$

Here a_i is the largest integer not larger than α_i. If $\alpha_{N+1} = \alpha$ the expansion has period N; if this is true for no N then α is not periodic. An example: for $D = 12$, $1 + \sqrt{3}$ is a reduced quadratic irrational. Not only do the inequalities hold — $1 + \sqrt{3} > 0$ and $0 > 1 - \sqrt{3} > -1$ — but the periodic expansion of $1 + \sqrt{3}$ is remarkably easy to find —

$$\alpha_1 = 1 + \sqrt{3} = 2 + \frac{1}{\frac{1+\sqrt{3}}{2}}$$

$$\alpha_2 = \frac{1 + \sqrt{3}}{2} = 1 + \frac{1}{1 + \sqrt{3}}$$

$$\alpha_3 = \alpha_1$$

The quadratic irrational numbers can be organized into sets based on the previous information. The most basic division is to group them based on whether or not they are reduced. We will use $A^* = A^*(D)$ to mean the group of all quadratic irrationals and $A = A(D)$ to mean the group of reduced ones. Beyond this, one can use the continued fractional expansion of α to divide A into cosets. The continued fractional expansion of α results in the sequence of numbers α_i(i=1, 2,..., N), and each of these sequences makes up a coset of A with respect to the equivalence relation. "Let $A = A_1 \cup A_2 \cup ... \cup A_h$ be the equivalence class decomposition of A, then the number h of the cosets is equal to the ideal class number of the field $F = Q(\sqrt{D})$ if D is the discriminant of F."

To me, this is a very interesting result. I had never seen even the phrase "class number" before reading this paper, and so I decided I should look it up. It turns out that the result just explained above is similar to a very classic way of finding the class number - John Stillwell explained class number in the introduction to Dirichlet's "Lectures on Number Theory" as "[t]he number of inequivalent forms with a given discriminant D." He was talking about binary quadratic forms ($ax^2 + 2bxy + cy^2 = n$, n an integer). *Equivalent* forms are forms for which a certain change of variables ($x' = \alpha x + \beta y$, $y' = \gamma x + \delta y$, $\alpha\delta - \beta\gamma = \pm 1$, $\alpha, \beta, \gamma, \delta$ all integers) does not change the values a form takes. An example of two *inequivalent* forms are those with $D = -5$, $x^2 + 5y^2$ and $2x^2 + 2xy + 3y^2$. These two expressions do not take the same values. There are many parallels between this way of finding class number and the specific method used above.

Most precisely, the class number is the cardinality of the *ideal class group*, which is the quotient group G/S where G is the group of all non-zero fractional ideals, and S is the subgroup of all non-zero principal fractional ideals. (These terms are defined later in this paper.) But another, perhaps more intuitive, way of thinking of class number is that it is very roughly the number of ways an algebraic integer can be factored in the ring of algebraic integers in the field $\mathbf{Q}(\sqrt{D})$. This brings us back to our real quadratic number field, and we can stop thinking about x, y, and changes of variable. It also makes more apparent the idea of class number as a measure of failure of unique factorization in a ring. This is intriguing idea, especially for those who are just venturing into the world of number theory and have never thought of the importance of unique factorization.

Even more interesting is a simple relationship that exists between class number and fundamental unit: $\frac{h\epsilon}{\sqrt{D}}$ has a constant value. Yamamoto does not use this result. It is startling enough to this author, though, that it may be an avenue of further inquiry. But back to work.

From α we can also find ϵ. It turns out that for $\alpha \in A$,

$$\alpha = \frac{a\alpha + b}{c\alpha + d}$$

Then $ad - bc = (-1)^N$, where N is the period of the continued fractional expansion of α, and $\epsilon = c\alpha + d$. There are several other ways to find ϵ from α, though. Proposition 1.2 in Yamamoto's paper says,

If D is equal to the discriminant of a real quadratic number field F and ϵ is the fundamental unit of F, then

$$\prod_{\alpha \epsilon A_i} \alpha = \epsilon$$

for any equivalence class $A_i (i = 1, 2, ..., h)$.

Corollary 1.3 says that in addition to that, $\prod_{\alpha \epsilon A} \alpha = \epsilon^h$. The corollary follows fairly easily from the proposition, and both can be proven by inductive reasoning.

The next section in Yamamoto's paper addresses the relationship between reduced quadratic irrationals and what will be called reduced ideals. Although the first section was mathematically accessible using only high-school algebra, the second section required me to introduce myself to abstract algebra, as well as a few more things.

The first thing Yamamoto brings up is ω. Let $\omega = \frac{D+\sqrt{D}}{2}$, and notice that 1 and ω make up a basis of the ring R of all algebraic integers in F, where $F = Q(\sqrt{D})$ is the real quadratic number field F with discriminant D. Then every integral ideal I has the form $[a, b + c\omega]$ with these conditions:

$$a, b, c \in \mathbf{Z}$$

$$a > 0, c > 0, ac = N(R)$$

$$a = b = 0 \bmod c$$

and

$$N(b + c\omega) = 0 \bmod ac$$

$$-a < b + c\omega' < 0$$

where ω' is the conjugate of ω. Then α=$\alpha(I) = \frac{b+c\omega}{a}$. We call α the quadratic irrational associated with the ideal I, and I is reduced if c=1 and $\alpha(I)$ is a reduced quadratic irrational.

A brief review of terms: An *ideal* I of a ring R is a subset of R for which, if $r \in R$, $rI = \{ra | a \in I\}$ and $Ir = \{ar | a \in I\}$. In addition, I is closed under multiplication by elements from R. In other words, if any element in a ring is multiplied by an element of one of the ring's ideals, their product is an element of the ideal. A quick example is the ideal generated by 2 in the ring of integers: multiply any integer by two and it's obvious that the product is a member of the set of even integers. Yamamoto refers to *integral ideals* when speaking of these subsets; this is to distinguish them from *fractional ideals*. Although never explicitly mentioned, fractional ideals are necessary in the definition of ideal class group, for instance, so here is a definition: Let $\mathbf{o}$ be the ring of algebraic integers in $\mathbf{Q}(\sqrt{D})$. Then a fractional ideal of $\mathbf{o}$ in $\mathbf{Q}(\sqrt{D})$ is a non-zero finitely-generated $\mathbf{o}$-module inside $\mathbf{Q}(\sqrt{D})$. (What does that last part mean? An $\mathbf{o}$-module M is finitely generated if there is a finite list $m_1, \ldots, m_n$ of elements in M so that $M = \mathbf{om_1} + \cdots + \mathbf{om_n}$. A module is to a ring as a vector space is to a field.) An integral ideal, then, is just a fractional ideal contained in $\mathbf{o}$. Yamamoto also writes of *principal* ideals — principal simply means that all the elements in the ideal are multiples of the generator.

The main proposition of section two of the paper is this (Proposition 2.1):

> *The map $I \to \alpha(I)$ gives a bijection of the set of all reduced ideals to the set $A = A(D)$ of all reduced quadratic irrationals with discriminant D. And it induces a bijection of the ideal class group of F to the set $A_1, A_2, ..., A_h$ of the equivalence classes of A.*

Of slightly less theoretical import, but extremely useful in the proof of the theorem the paper is aiming at, is Proposition 2.2:

> *An integral ideal I is reduced if (i) $N(I) < \frac{\sqrt{D}}{2}$ and (ii) the conjugate ideal I' is relatively prime to I.*

Once again, a definition of terms: If an ideal has the form $[a, b + c\omega]$, then its conjugate has the form $[a, b + c\omega']$. For two ideals to be relatively prime means that only products of their elements are in their intersection.

These two propositions are fairly significant. The technique of associating ideals with numbers is fairly common, and a good tool. Since numbers are often easier to deal with than ideals, a bijection of a set of ideals to a set of numbers can be a good way to turn an algebra problem into a problem with numbers rather than ideals. However, in this proof, we'll see that Theorem 3.1 deals with ideals, and that since the bijection exists the result holds for actual numbers.

Section three of Yamamoto's paper brings all the previous information together into a proof of the theorem given at the beginning. What follows, in my paper, is essentially a commentary on and explanation of the proof given in Yamamoto's note.

Theorem 3.1. Let $p_i (i = 1, 2, ..., n)$ be rational primes satisfying $p_1 < p_2 < \cdots < p_n$. Assume that there exist infinitely many real quadratic number fields F satisfying the following condition (*):

(*) Every p_i is decomposed in F into the product of two principal ideals m_i and m_i'.

Then there exists a positive constant c_0 depending only on n and $p_1, p_2, \cdots, p_n$ such that

$$\log \epsilon > c_0 (\log \sqrt{D})^{n+1}$$

holds for sufficiently large D, where D and ϵ are the discriminant and the fundamental unit for F.

Proof. *Consider the ideals I of the form*

$$I = \prod_{i=1}^{n} m_i^{e_i} m_i'^{f_i}$$

Then I is a principal integral ideal and reduced if (a) $N(I) = p_1^{e_1+f_1} \cdots p_n^{e_n+f_n} < \frac{\sqrt{D}}{2}$ and (b) $e_1 f_1 = \cdots = e_n f_n = 0$ (Proposition 2.2).

This is not hard to see: Proposition 2.2 says that an integral ideal I is reduced if $N(I) < \frac{\sqrt{D}}{2}$, and it is easy to see that (a) follows this exactly. Although slightly less obvious, (b) follows from part (ii) of Prop. 2.2. Condition (b) makes it necessary that either e_i or f_i is equal to 0, for any i. This ensures that I and I' are always relatively prime to each other.

Let $I_1, I_2, \cdots, I_t$ be the set of all reduced ideals obtained as above. Then the quadratic irrationals $\alpha_1, \alpha_2, \cdots, \alpha_t$ associated with them build a subset of the equivalence class A_1, say, corresponding to the principal ideal class.

The principal ideal class is the collection of fractional ideals which are principal. Most important in that definition is the fact that it is a subgroup of the group of all fractional ideals. This allows us to understand the next sentence:

So we get, from Proposition 1.2,

$$\epsilon = \prod_{\alpha \epsilon A_1} \alpha > \prod_{i=1}^{t} \alpha_i.$$

Proposition 1.2 says that ϵ is the product of all the α in A_1. This product is larger than the product of all $\alpha_i (i = 1, 2, \cdots, t)$ because the α_i make up a subset of A_1. Since all α are larger than one by defintion, multiplying more of them makes a bigger number.

On the other hand we have

$$\alpha_i = \frac{b_i + \omega}{N(I_i)} > (\frac{\sqrt{D}}{2})(p_1^{e_1+f_1} \cdots p_n^{e_n+f_n}),$$

where $I_i = [N(I_i), b_i + \omega]$ *is the canonical basis of* I_i.

This basis can be found from section two by noting that if I has basis $[a, b + c\omega]$, a condition on a is that $N(I) = ac$, and I is reduced if c=1. Thus the basis can be written as $[N(I), b + \omega]$.

Hence we get the following inequality

$$(3.1) \qquad \epsilon > \prod{}' \frac{\sqrt{D}/2}{p_i^{e_1+f_1} \cdots p_n^{e_n+f_n}} = \epsilon_0.$$

The product in (3.1) is taken over all integers e_i *and* f_i *satisfying*

$$(a') \qquad (e_1 + f_1)\log p_1 + \cdots + (e_n + f_n)\log p_n < \log(\frac{\sqrt{D}}{2}),$$

$$(b') \qquad e_i \geq 0, f_i \geq 0 \text{ and } e_i f_i = 0 (i = 1, 2, \cdots, n).$$

Conditions (a') and (b') follow easily from conditions (a) and (b) given earlier in the proof — for (a') use the laws for logarithms, for (b') notice that if $ab = 0$, a, b both real numbers, $a = 0$ or $b = 0$. The condition that e_i and f_i be positive is new but not surprising.

We have

$$(3.2) \qquad \prod{}' \frac{\sqrt{D}}{2} = (\frac{\sqrt{D}}{2})^t$$

The number t *equals to the cardinal of the set of* $2n$*-tuples* $(e_1, f_1, \cdots, e_n, f_n)$ *satisfying* (a') *and* (b').

Remember that t first came up as the number of ideals I obtained by multiplying factors of primes - check the beginning of the proof.

Then it holds

$$(3.3) \qquad t = \frac{2^n V}{P} + O((\log \frac{\sqrt{D}}{2}^{n-1}),$$

where V *is the volume of the* n*-simplex* Δ *in the* n*-dimensional euclidean space* $\mathbf{R}^n$*;*

$$\Delta = \{(x_1, \cdots, x_n) \in \mathbf{R}^n : x_1 \geq 0, \cdots, x_n \geq 0,\ x_1 + \cdots + x_n \leq \log(\frac{\sqrt{D}}{2})$$

and $\qquad P = (\log p_1)(\log p_2) \cdots (\log p_n).$

Notice that (a') is very similar to the second descriptor for the set Δ, and that all the x_i must be greater than or equal to 0. What Yamamoto is doing is saying that a parallel can be drawn between e_i and f_i and the x_i, so that standard results for n-simplices can be applied to this problem.

It would probably be a good idea, at this point, to explain what an n-simplex is. A 1-simplex is a line. A 2-simplex is a triangle. A 3-simplex is a tetrahedron. The pattern goes on like this, with n indicating the dimension of the shape with triangular 'sides.' The volume of the n-simplex approximates the number of lattice points inside - the number of e_i and f_i that satisfy the conditions given. Then the $O((\frac{\sqrt{D}}{2})^{n-1})$ term is an error term. For sufficiently large D, the n-simplex estimation is fairly accurate.

When all the vertices of an n-simplex have coordinates $(0, \cdots, 1, \cdots, 0)$, in $\mathbf{R}^n$, then the n-volume of the simplex is $\frac{1}{n!}$. However, in this case the coordinates of the vertices x_i have a condition imposed on them:

$x_1 + \cdots + x_n \leq (\log \frac{\sqrt{D}}{2})$. Yamamoto combines these facts next.

We have

$$V = \int_\Delta dx_1 \cdots dx_n = \frac{1}{n!}(\log \frac{\sqrt{D}}{2})^n.$$

Almost there...

For the product of all denominators in the right side of (3.1), we have

$$(3.4) \qquad \log \Pi'(p_1^{e_1+f_1} \cdots p_n^{e_n+f_n})$$

$$= \sum{}' [(e_1 + f_1) \log p_1 + \cdots + (e_n + f_n) \log p_n]$$

$$= \frac{2^n}{P} \int_\Delta (x_1 + \cdots + x_n) dx_1 \cdots dx_n + O((\log \frac{\sqrt{D}}{2})^2)$$

$$= \frac{2^n n}{(n+1)!P} (\log \sqrt{D})^{n+1} + O((\log \sqrt{D})^n).$$

This is a fairly easy-to-follow set of manipulations, using only logarithm laws, integration, and algebra.

From (3.3) and (3.4), we get

$$[3.45] \qquad \log \epsilon_0 = \log(\frac{\sqrt{D}}{2})^t - \log \Pi'(p_1^{e_1+f_1} \cdots p_n^{e_n+f_n})$$

$$= \frac{2^n}{(n+1)!P} (\log \sqrt{D})^{n+1} + O((\log \sqrt{D})^n).$$

Our theorem follows from this and (3.1).

Recall that (3.1) says $\epsilon > \prod{}' \frac{\sqrt{D}/2}{p_i^{e_1+f_1} \cdots p_n^{e_n+f_n}} = \epsilon_0$. Take the log of both sides, then substitute the expression in [3.45] for $\log \epsilon_0$. $\frac{2^n}{(n+1)!P}$ is c_0 in the theorem, and for sufficiently large D the error term in [3.45] is negligible. Thus Theorem 3.1 is proven.

Theorem 3.2 is the last theorem in Yamamoto's paper, and I will include it for the sake of completeness. Not much comment is needed; anything difficult to understand I leave, as an exercise, to the reader.

Theorem 3.2 *For the case $n = 2$, the assumption of Theorem 3.1 is satisfied by the following F's: $F = \mathbf{Q}(\sqrt{m_k})$*

$$m_k = (p^k q + p + 1)^2 - 4p \quad (k = 1, 2, \ldots),$$

where we set $p = p_1$ and $q = p_2$.

Proof *We easily see that $m_k = 1 (mod\ p)$, $_k = (p-1)^2 (mod\ q)$ and $m_k = 1 (mod\ 4)$ ($m_k = 1 (mod\ 8)$ if $p = 2$. Hence each one of p and q is decomposed into the product of two distinct prime ideals in F (if m_k is not a square). Set $p = mm'$ and $q = nn'$. From the definition of m_k, it holds*

$$(3.5) \qquad (p^k q + p + 1)^2 - m_k = 4p$$

$$(3.6) \qquad (p^k q + p + 1)^2 - m_k = -4p^k q.$$

From (3.5), m and m' are both principal (set $m = (\frac{p^k q+p+1+\sqrt{m_k}}{2}$), for example.) From (3.6), either $m^k n$ or $m'^k n$ is principal. Since m^k and m'^k are principal, both n and n' are also principal. So the condition (*) in Theorem 3.1 is satisfied. Finally, the infiniteness of the number of F's given above is as follows. Set $k = 2j$ (we consider the case where k is even), then

$$m_k = m_{2j} = (p^{2j}q + p + 1)^2 - 4p = q^2 p^{4j} + nq(p+1)p^{2j} + (p-1)^2.$$

Since the diophantine equation

$$Dy^2 = q^2 x^4 + 2q(p+1)x^2 + (p-1)^2$$

has only a finite number of rational integral solutions (x, y) for a fixed integer D (Siegel's theorem), $\mathbf{Q}(\sqrt{m_{2j}})$ represents infinitely many real quadratic number fields F for $j = 1, 2, \ldots$. This completes the proof.

So the theorems about bounds for fundamental units in real quadratic number fields have been proven. I can find, if I want, a lower bound for the fundamental unit ϵ in a real quadratic number field with a given large discriminant D; from this I can find some estimate for the solution to Pell's equation with the same D. Let us return, then, to the original question: How long is it going to take to find that solution to $x^2 + 151y^2 = 1$? How big is it going to be? And the answer is this: although we now have a method for finding a lower bound for ϵ, the method itself takes implementation. It is necessary to ascertain how many primes smaller than 151 decompose in $\mathbf{Q}(\sqrt{151})$ into the product of two principal prime ideals. This is not the easiest of tasks — I could write a computer program to do it for me, but if I am willing to do that then I might as well write a computer program to find the solutions to Pell's equation directly, and perhaps use a more efficient algorithm so that it doesn't take three hours. So here my exploration stopped. My summer has ended, and so has the theoretical part of the question I investigated. Now it's all application...