Math 4527 (Number Theory 2)

Lecture #7 of 38 $\sim$ February 3, 2021

Transcendence + Pell's Equation (Part 1)

- Transcendence
- Pell's Equation Background
- Pell's Equation and The Norm Map on $\mathbb{Z}[\sqrt{D}]$
- Pell's Equation and Rational Approximation

This material represents §6.2.6-6.3.1 from the course notes.

Transcendence, I

Last time, I explained how we can use our rational approximation results to prove that a given real number α is irrational:

Proposition (Irrationality and Approximation)

A real number α is irrational if and only if there exist infinitely many distinct rational numbers p/q such that $|\alpha - p/q| < 1/q^2$.

Example: The real number $\alpha = \sum_{k=0}^{\infty} 10^{-3^k}$ is irrational because its partial sums $p_n/q_n = \sum_{k=0}^{n} 10^{-3^k}$ all satisfy the inequality $|\alpha - p_n/q_n| < 1/q_n^2$.

Transcendence, II

As first observed by Liouville, we can extend this criterion to exclude algebraic numbers that are roots of higher-degree polynomials by increasing the exponent of q in the bound on the right-hand side. First, some preliminaries:

- We say that a number $\alpha \in \mathbb{C}$ is algebraic if α is the root of some nonzero polynomial $p(x)$ with rational coefficients.
- If we consider all of the possible polynomials in $\mathbb{Q}[x]$ of which α is a root, by the well-ordering principle we can see that there is some polynomial of minimal degree d of which α is a root.
- We refer to this degree d as the algebraic degree of α over $\mathbb{Q}$. There is a unique monic polynomial of this degree d of which α is a root; this polynomial is called the minimal polynomial of α over $\mathbb{Q}$.

Transcendence, III

Examples:

1. Quadratic irrationals have algebraic degree 2 over $\mathbb{Q}$, since they are roots of quadratic polynomials but not any polynomial of lower degree.
2. The number $\sqrt[4]{2}$ has minimal polynomial $x^4 - 2$ over $\mathbb{Q}$ (although this is not completely trivial to prove) and thus has algebraic degree 4.
3. The number $\sqrt{2} + \sqrt{3}$ is algebraic because it is the root of the polynomial $x^4 - 5x^2 + 1$, and in fact (though this is again harder to prove) that polynomial is its minimal polynomial.
4. The number π is not algebraic because it is not the root of any nonzero polynomial with rational coefficients. (This is even harder to prove, of course.)

Transcendence, IV

Some other observations:

- The minimal polynomial is always irreducible (otherwise, one factor would have α as a root and have smaller degree) and it cannot have any repeated roots (otherwise m and its derivative m' would have a factor $x - \alpha$ in common).
- We can also clear denominators to see that any algebraic number is the root of a polynomial with integer coefficients.
- If α is the root of some polynomial $c_d x^d + c_{d-1} x^{d-1} + \cdots + c_0$ where the c_i are integers, then $c_d \alpha^d + c_{d-1} \alpha^{d-1} + \cdots + c_0 = 0$.
- If we set $\beta = c_d \alpha$, by rescaling we can see that β is a root of the polynomial $x^d + c_{d-1} c_d x^{d-1} + \cdots + c_0 c_d^{d-1}$, which is monic and has integer coefficients.
- Thus, up to an integer factor, any algebraic number is the root of a monic polynomial with integer coefficients.

Transcendence, V

With these preliminaries finished, we can now give Liouville's result:

Theorem (Liouville's Approximation Theorem)

Suppose α is algebraic of degree $n > 1$ over $\mathbb{Q}$ and that its minimal polynomial $m(x)$ has integer coefficients. Then there exists a positive real number A such that $|\alpha - p/q| \geq A/q^n$ for any rational number p/q.

The idea of the proof is to use the mean value theorem to bound the difference between $m(\alpha)$ and $m(p/q)$ and the fact that we can express $m(p/q)$ as $1/q^n$ times an integer.

We can also reduce to the situation where the minimal polynomial is monic by rescaling α, as we noted on the last slide. So the given assumption is not really a restriction.

Transcendence, VI

Proof:

- Suppose α is algebraic of degree $n > 1$ over $\mathbb{Q}$ and that its minimal polynomial $m(x)$ has integer coefficients and factors as $m(x) = (x - \alpha)(x - \beta_1)(x - \beta_2) \cdots (x - \beta_{n-1})$ over $\mathbb{C}$.
- Note that the β_i are distinct from α because $m(x)$ cannot have repeated roots as we noted earlier.
- Now define M be the maximum value of $|m'(x)|$ on the interval $[\alpha - 1, \alpha + 1]$, and set $A = \min(1, 1/M, |\alpha - \beta_i|)$ over all of the roots β_i.
- We claim this value of A satisfies the given inequality.
- To show this, suppose otherwise, so that p/q is rational and has $|\alpha - p/q| < A/q^n$. Then because $A \leq 1$, we have $p/q \in (\alpha - 1, \alpha + 1)$.
- Also, because $A \leq |\alpha - \beta_i|$, we see that $p/q \neq \beta_i$ for any i, and there is no root of $m(x)$ between α and p/q.

Transcendence, VII

Proof (continued):

- Now write $m(x) = x^d + c_{d-1}x^{d-1} + \cdots + c_0$.
- Then $m(p/q) = (p/q)^d + c_{d-1}(p/q)^{d-1} + \cdots + c_0$
  $= (1/q^d) \cdot [p^d + c_{d-1}p^{d-1}q + \cdots + c_0 q^d]$.
- So $|m(p/q)| \geq 1/q^d \cdot \left|p^d + c_{d-1}p^{d-1}q + \cdots + c_0 q^d\right| \geq 1/q^d$ because the term inside the absolute values is an integer and it cannot be zero since $m(p/q) \neq 0$.
- Now, by the mean value theorem, there exists x_0 in the interval with endpoints p/q and α such that $m(\alpha) - m(p/q) = m'(x_0) \cdot (\alpha - p/q)$. Taking absolute values yields $|m(\alpha) - m(p/q)| = |m'(x_0)| \cdot |\alpha - p/q|$.
- By assumption we have $A \leq 1/M$ and $|m'(x_0)| \leq M$, and also $m(\alpha) = 0$ and $|m(p/q)| \geq 1/q^d$.
- So we get $|\alpha - p/q| = \dfrac{|m(p/q)|}{|m'(x_0)|} \geq \dfrac{A}{q^d}$ as desired.

Transcendence, VIII

Roughly speaking, this result says that if we have an algebraic number α of degree d, then we cannot find a rational approximation that is "too close" to α.

- If we flip the condition around, then if we have a real number α that we *can* approximate extremely well, then it cannot be algebraic.
- More precisely: if α is an irrational real number such that there exists a constant $c > 0$ and a sequence of rational numbers p_n/q_n such that $|\alpha - p_n/q_n| < c/q_n^n$, then α is transcendental.
- The point is that this sequence of rational numbers contradicts the assertion that α is algebraic of degree n for every n, by Liouville's theorem, and so α must be transcendental.

Transcendence, IX

We can use a similar sort of construction as we used earlier to construct transcendental numbers.

- We can construct such an α and corresponding rational approximations p_n/q_n by taking α to be an infinite series whose terms drop in size very quickly.
- Here, we want the tail after the nth partial sum p_n/q_n to be on the order of $1/q_n^n$ rather than $1/q_n^2$: this will guarantee that α will be transcendental.

Transcendence, X

Example: Show that $\alpha = \sum_{k=0}^{\infty} 10^{-k!}$ is transcendental.

- Let $p_n/q_n = \sum_{k=0}^{n} 10^{-k!}$ be the nth partial sum of the series. We observe that $q_n = 10^{k!}$ since each of the other terms has a denominator dividing $10^{-k!}$.
- Furthermore, it is easy to see (e.g., from the decimal expansion of α) that the size of the tail $\sum_{k=n+1}^{\infty} 10^{-k!}$ is at most $2 \cdot 10^{-(n+1)!}$.
- Then we have an easy bound $|\alpha - p_n/q_n| < 2 \cdot 10^{-(n+1)!} = 2(10^{-n!})^{n+1} = 2/q_n^{n+1} < 1/q_n^n$. Since all of the partial sums of this series are distinct, we obtain infinitely many such p_n/q_n, and therefore by our result above, α is transcendental.

Pell's Equation Intro, I

We now switch back into discussing Diophantine equations. We will spend the next few lectures studying the class of equations of the form $x^2 - Dy^2 = r$ where D is a positive nonsquare integer and r is an arbitrary integer.

- Such equations are often referred to under the general heading of Pell's equation, named after the English mathematician John Pell.
- However, this name is a misattribution by Euler, and it is quite possible that Pell never actually studied these equations.
- Equations of this type have been studied throughout history, with notable early contributions made by the Indian scholars Brahmaguptra, Bhaskara II, and Narayana.
- Certain instances of Pell's equation (most notably $D = 2$) were also studied by the ancient Greeks, including Diophantus.

Pell's Equation Intro, II

What we would like to be able to do is find a recipe for generating solutions to Pell's equation in the situations that they do exist, and to understand more about the structures of these solutions.

- The general approach we will follow is similar to the treatment developed by Lagrange in the mid-1700s.
- The main ideas in the more modern approach to Pell's equation is to exploit the language of more general rings to simplify much of the calculation.
- Perhaps unsurprisingly, the techniques we will develop will also eventually involve our results on continued fraction expansions and rational approximations.

Pell's Equation Intro: $D = 2$, I

Let's start by exploring the case $D = 2$ for various small r: thus, we are seeking integer solutions to the Diophantine equation $x^2 - 2y^2 = r$ for small values of r.

- We can do a search by plugging in small nonnegative values of x and y from 0 to 20 and looking for pairs where $x^2 - 2y^2$ is close to zero.
- On the next slide, I've collected them via the value of r.

Pell's Equation Intro: $D = 2$, II

Solutions to $x^2 - 2y^2 = r$ for small $|r|$ and $0 \leq x, y \leq 50$:

r	1	2	3
(x, y)	$(1, 0)$,$(3, 2)$,$(17, 12)$	$(2, 1)$,$(10, 7)$	none
r	-1	-2	-3
(x, y)	$(1, 1)$,$(7, 5)$,$(41, 29)$	$(0, 1)$,$(4, 3)$,$(24, 17)$	none
r	4	5	6
(x, y)	$(2, 0)$,$(6, 4)$,$(34, 24)$	none	none
r	-4	-5	-6
(x, y)	$(2, 2)$,$(14, 10)$	none	none
r	7	8	9
(x, y)	$(3, 1)$,$(5, 3)$,$(13, 9)$,$(27, 19)$	$(2, 4)$,$(20, 14)$	$(3, 0)$,$(9, 6)$
r	-7	-8	9
(x, y)	$(1, 2)$,$(5, 4)$,$(11, 8)$,$(31, 22)$	$(0, 2)$,$(8, 6)$,$(48, 34)$	$(3, 3)$,$(21, 15)$

Pell's Equation Intro: $D = 2$, III

A few basic observations:

- For some values of r (namely, $r = \pm 3$) there seem to be no solutions (though of course the table itself does not prove this, since the search is only for $x, y \leq 50$), while for other small values of r there are solutions.
- It also seems that there is a solution to $x^2 - 2y^2 = r$ if and only if there is also a solution to $x^2 - 2y^2 = -r$.
- The smaller values of $|r|$ have more solutions in the range we searched than the larger values did.
- The solutions for a fixed value of r seem to grow fairly quickly: no value of r has more than 3 solutions in the range we searched.

Pell's Equation Intro: $D = 2$, IV

By using some basic modular arithmetic we can show that there are no solutions to $x^2 - 2y^2 = r$ for some values of r.

- We cannot expect to get any contradictions modulo 2, and it is not hard to check that $x^2 - 2y^2$ can also take any value modulo 4.
- However, if we work modulo 8, we can see that $x^2 \in \{0, 1, 4\}$ mod 8 and $-2y^2 \in \{0, 6\}$ mod 8.
- Thus, $x^2 - 2y^2 \in \{0, 1, 2, 4, 6, 7\}$ mod 8, and so it cannot be congruent to 3 or 5 mod 8.
- This means that for $r \equiv 3, 5$ mod 8, there are no solutions to $x^2 - 2y^2 = r$. This explains the lack of solutions for $r = \pm 3, \pm 5$ in our table.

Pell's Equation Intro: $D = 2$, V

We can also get some more nonexistence results by looking at other small moduli.

- For example, if we look modulo 3, then because $x^2 - 2y^2 \equiv x^2 + y^2 \bmod 3$, this quantity can equal zero mod 3 only when $(x, y) \equiv (0, 0) \bmod 3$.
- But in such cases, $x^2 - 2y^2 = r$ is then actually divisible by 9, and so it cannot be 3 or 6 modulo 9.
- This means that for $r \equiv 3, 6 \bmod 9$, there are no solutions to $x^2 - 2y^2 = r$. This explains the lack of solutions for $r = \pm 3, \pm 6$ in our table.

Pell's Equation Intro: $D = 2$, VI

Solutions to $x^2 - 2y^2 = r$ for small $|r|$ and $0 \leq x, y \leq 50$:

r	1	2	3
(x, y)	$(1,0)$,$(3,2)$,$(17,12)$	$(2,1)$,$(10,7)$	none
r	-1	-2	-3
(x, y)	$(1,1)$,$(7,5)$,$(41,29)$	$(0,1)$,$(4,3)$,$(24,17)$	none
r	4	5	6
(x, y)	$(2,0)$,$(6,4)$,$(34,24)$	none	none
r	-4	-5	-6
(x, y)	$(2,2)$,$(14,10)$	none	none
r	7	8	9
(x, y)	$(3,1)$,$(5,3)$,$(13,9)$,$(27,19)$	$(2,4)$,$(20,14)$	$(3,0)$,$(9,6)$
r	-7	-8	9
(x, y)	$(1,2)$,$(5,4)$,$(11,8)$,$(31,22)$	$(0,2)$,$(8,6)$,$(48,34)$	$(3,3)$,$(21,15)$

Pell's Equation Intro: $D = 2$, VII

Another pattern we can observe from the examples above is that $x^2 - 2y^2 = r$ seems to have a solution if and only if $x^2 - 2y^2 = -r$ does, and that some of the solutions seem to be related.

- Let's look for a relationship between the pairs $(1, 0)$, $(3, 2)$, $(17, 12)$ from $r = +1$ and $(1, 1)$, $(7, 5)$, $(41, 29)$ from $r = -1$.
- From the larger numbers, it is not hard to spot a pattern: if (a, b) is a solution with $a^2 - 2b^2 = 1$, then $(a + 2b, a + b)$ seems to be a solution with $a^2 - 2b^2 = -1$.
- In fact, this also works the other way around: if (a, b) is a solution with $a^2 - 2b^2 = -1$, then $(a + 2b, a + b)$ is a solution with $a^2 - 2b^2 = 1$.

Pell's Equation Intro: $D = 2$, VIII

Indeed, the same pattern holds up for the other values of r (± 2, ± 4, ± 7, etc.): if (a, b) is a solution with $a^2 - 2b^2 = -r$, then $(a + 2b, a + b)$ seems to be a solution with $a^2 - 2b^2 = r$.

- Indeed, this is easy to check algebraically: if $a^2 - 2b^2 = -r$, then $(a + 2b)^2 - 2(a + b)^2 = -(a^2 - 2b^2) = r$.
- We can see quite easily that if we start with any solution to $x^2 - 2y^2 = r$ (even the "trivial" solution $(\pm 1, 0)$ to $x^2 - 2y^2 = 1$) we can generate new solutions to $x^2 - 2y^2 = \pm r$ by applying this rule mapping $(a, b) \mapsto (a + 2b, a + b)$.

Pell's Equation Intro: $D = 2$, IX

We can use this recipe $(a, b) \mapsto (a + 2b, a + b)$ to generate many more solutions starting from a single one.

- For example, starting with $(1, 0)$ we obtain $(1, 0) \mapsto (1, 1) \mapsto (3, 2) \mapsto (7, 5) \mapsto (17, 12) \mapsto (41, 29) \mapsto (99, 70) \mapsto (239, 169) \mapsto \cdots$. The odd terms in the sequence are solutions to $x^2 - 2y^2 = 1$ while the even terms are solutions to $x^2 - 2y^2 = -1$.
- If we iterate the rule twice, mapping $(a, b) \mapsto (a + 2b, a + b) \mapsto (3a + 4b, 2a + 3b)$, we obtain a recipe for generating new solutions to $x^2 - 2y^2 = r$ from old solutions.

Pell's Equation Intro: $D = 3$, I

Let's now examine the case $D = 3$ for small r: now we are seeking integer solutions to $x^2 - 3y^2 = r$.

- On the next slide, just like with $D = 2$, are the results of searching for $0 \leq x, y \leq 50$ for solutions to $x^2 - 3y^2 = r$.

Pell's Equation Intro: $D = 3$, II

Solutions to $x^2 - 3y^2 = r$ for small $|r|$ and $0 \leq x, y \leq 50$:

r	1	2	3
(x, y)	$(1,0),(2,1),(7,4),(26,15)$	none	none
r	-1	-2	-3
(x, y)	none	$(1,1),(5,3),(19,11)$	$(0,1),(3,2),(12,7),(45,26)$
r	4	5	6
(x, y)	$(2,0),(4,2),(14,8)$	none	$(3,1),(9,5),(33,19)$
r	-4	-5	-6
(x, y)	none	none	none
r	7	8	9
(x, y)	none	none	$(3,0),(6,3),(21,12)$
r	-7	-8	-9
(x, y)	none	$(2,2),(10,6),(38,22)$	none

Pell's Equation Intro: $D = 3$, III

Some things are quite similar to the case $D = 2$: for example, we can establish the nonexistence of solutions to $x^2 - 3y^2 = r$ in a number of cases using modular arithmetic.

- For example, if $r \equiv 2 \bmod 3$ (which includes the cases $r = -7, -4, -1, 2, 5, 8$), then taking $x^2 - 3y^2 = r$ modulo 3 gives $x^2 \equiv 2 \pmod 3$, which has no solution when $r \equiv 2 \pmod 3$ since 2 is not a quadratic residue.
- Similarly, if $r \equiv 3 \bmod 9$ (including the cases $r = -6, 3$) then any solution to $x^2 - 3y^2 = r$ requires x to be divisible by 3. If $x = 3a$ then cancelling the factor of 3 yields $3a^2 - y^2 = (r/3)$ so that $y^2 - 3a^2 = -(r/3) \equiv 2 \bmod 3$, but this has no solution by the above.

Pell's Equation Intro: $D = 3$, IV

Some things are quite similar to the case $D = 2$: if we have a solution to $x^2 - 3y^2 = r$ then it seems that we will always have several solutions.

- If we search for a recipe (like in the case with $D = 2$) to generate new solutions to $x^2 - 3y^2 = 1$ from old ones, we can eventually stumble upon the map $(a, b) \mapsto (2a + 3b, a + 2b)$.
- For example, this operation maps $(1, 0) \mapsto (2, 1) \mapsto (7, 4) \mapsto (26, 15) \mapsto (97, 56) \mapsto (362, 209) \mapsto \cdots$.
- This map will always yield new solutions: if $a^2 - 3b^2 = r$ then $(2a+3b)^2-3(a+2b)^2 = [4a^2+12ab+9b^2]-[3a^2+12ab+12b^2] = a^2 - 3b^2 = r$ as well.
- We can see that all of the tuples in each cell of the table are actually generated from the smallest solution in this way.

Pell's Equation Intro: $D = 3$, V

However, there is one obvious thing that is very different in the case $D = 3$: it seems that if there is a solution to $x^2 - 3y^2 = r$ then there is no solution to $x^2 - 3y^2 = -r$.

- In contrast, for $D = 2$, we saw that having a solution to $x^2 - 2y^2 = r$ always forces existence of a solution to $x^2 - 2y^2 = -r$ and vice versa.
- This is dictated by our operation $(a, b) \mapsto (a + 2b, a + b)$ constructing a solution of one equation from a solution of the other.

Pell's Equation Intro: $\mathbb{Z}[\sqrt{D}]$, I

We can explain many of the patterns witnessed above by using properties of the ring $\mathbb{Z}[\sqrt{D}] = \{a + b\sqrt{D} : a, b \in \mathbb{Z}\}$ and the associated norm map $N(a + b\sqrt{D}) = a^2 - Db^2$.

- First, notice that the norm map is always integer-valued, and also satisfies the property that $N(a + b\sqrt{D}) = (a + b\sqrt{D})(a - b\sqrt{D}) = \alpha\overline{\alpha}$.
- The other crucial fact is that the norm map is multiplicative, which follows from the much easier fact that the conjugation operation is multiplicative.
- Explicitly, suppose $\alpha, \beta \in \mathbb{Z}[\sqrt{D}]$.
- Then $N(\alpha\beta) = \alpha\beta\overline{\alpha\beta} = \alpha\beta \cdot \overline{\alpha}\overline{\beta} = \alpha\overline{\alpha} \cdot \beta\overline{\beta} = N(\alpha)N(\beta)$.

Pell's Equation Intro: $\mathbb{Z}[\sqrt{D}]$, II

The obvious connection to Pell's equation is that solving $x^2 - Dy^2 = r$ is equivalent to solving $N(x + y\sqrt{D}) = r$.

- Many of the relations we observed can be explained using the fact that the norm map is multiplicative.
- In the particular situation where $N(\beta) = \pm 1$, we can see that $N(\alpha\beta^k) = (-1)^k r$.
- Thus, multiplying the element α by β, β^2, β^3, ... will yield more solutions to $x^2 - Dy^2 = \pm r$.
- Indeed, we can generate such a sequence whenever we can find the elements in $\mathbb{Z}[\sqrt{D}]$ of norm ± 1.

Pell's Equation Intro: $\mathbb{Z}[\sqrt{D}]$, III

We have a very convenient characterization of the elements of norm ± 1 inside $\mathbb{Z}[\sqrt{D}]$:

Proposition (Units in $\mathbb{Z}[\sqrt{D}]$)

An element $\alpha = a + b\sqrt{D}$ is a unit in $\mathbb{Z}[\sqrt{D}]$ if and only its norm $N(\alpha) = a^2 - Db^2$ is 1 or -1.

Proof:

- First suppose α is a unit with inverse β, so that $\alpha\beta = 1$.
- Then $N(\alpha)N(\beta) = N(\alpha\beta) = N(1) = 1$. But $N(\alpha)$ and $N(\beta)$ are integers, so the only possibility is to have $N(\alpha) = 1$ or -1.
- Conversely, if $N(\alpha) = \pm 1$ then $\alpha\overline{\alpha} = \pm 1$, so $\alpha \cdot (\pm\overline{\alpha}) = 1$. Then $\pm\overline{\alpha}$ is a multiplicative inverse of α, so α is a unit.

Pell's Equation Intro: $\mathbb{Z}[\sqrt{D}]$, IV

As an illustration, we can see quite easily that inside $\mathbb{Z}[\sqrt{2}]$ we have $N(1+\sqrt{2}) = (1+\sqrt{2})(1-\sqrt{2}) = -1$.

- Thus, by the multiplicativity of the norm map, if $a+b\sqrt{2}$ has norm r, then $(a+b\sqrt{2})(1+\sqrt{2}) = (a+2b)+(a+b)\sqrt{2}$ will have norm $-r$.
- This is precisely the map $(a,b) \mapsto (a+2b, a+b)$ we identified above, but now it is much clearer where it comes from.

Likewise, the map $(a,b) \mapsto (2a+3b, a+2b)$ for solutions to $x^2 - 3y^2 = r$ arises from the fact that we have $N(2+\sqrt{3}) = (2+\sqrt{3})(2-\sqrt{3}) = 1$ in $\mathbb{Z}[\sqrt{3}]$.

- We can then multiply out to see that if $N(a+b\sqrt{3}) = r$, then $(a+b\sqrt{3})(2+\sqrt{3}) = (2a+3b)+(a+2b)\sqrt{2}$ will also have norm r.

Pell's Equation Intro: $\mathbb{Z}[\sqrt{D}]$, V

All of this discussion suggests that should start by looking for the solutions of $x^2 - Dy^2 = \pm 1$, which is equivalent to determining the units in $\mathbb{Z}[\sqrt{D}]$.

- Based on our (admittedly small) searches above for solutions of $x^2 - Dy^2 = \pm 1$, it would appear that the units all have the form $\pm\alpha^n$ where α is the "smallest" solution to $x^2 - Dy^2 = \pm 1$ in the sense that $\alpha = x + y\sqrt{D}$ with $x, y > 0$ and where x is minimal.
- We will in fact be able to prove that these are all of the units, but it will take a little bit more effort first.

Pell's Equation Intro: $\mathbb{Z}[\sqrt{D}]$, VI

However, under the hypothesis that the units are all powers of a single "smallest unit", we can give an explicit definition for what that smallest unit would be:

Definition

For a fixed positive squarefree D, a fundamental solution (x_1, y_1) to Pell's equation is a pair (x_1, y_1) of positive integers such that $x_1^2 - Dy_1^2 = \pm 1$ and $x_1 + y_1\sqrt{D}$ is minimal.

The fundamental unit of $\mathbb{Z}[\sqrt{D}]$ is $u = x_1 + y_1\sqrt{D}$.

Note that this fundamental solution and the fundamental unit are well defined, assuming they do exist: there will be a unique minimal positive value for $x_1 + y_1\sqrt{D}$ over all pairs (x_1, y_1) satisfying $x_1^2 - Dy_1^2 = \pm 1$.

Pell's Equation Intro: $\mathbb{Z}[\sqrt{D}]$, VII

Examples: By searching for solutions to $x^2 - Dy^2 = \pm 1$ we can generate fundamental units for various small nonsquare D:

D	2	3	5	6	7
Fund. Unit	$1+\sqrt{2}$	$2+\sqrt{3}$	$2+\sqrt{5}$	$5+2\sqrt{6}$	$8+3\sqrt{7}$
Norm	-1	1	-1	1	1
D	8	10	11	12	13
Fund. Unit	$3+\sqrt{8}$	$3+\sqrt{10}$	$10+3\sqrt{11}$	$7+2\sqrt{12}$	$18+5\sqrt{13}$
Norm	1	-1	1	1	-1
D	14	15	17	18	19
Fund. Unit	$15+4\sqrt{14}$	$4+\sqrt{15}$	$4+\sqrt{17}$	$17+4\sqrt{18}$	HW #3
Norm	1	1	-1	1	HW #3

Pell's Equation Intro: Rational Approximation, I

One of the other key ideas for solving Pell's equation is the observation that if $x^2 - Dy^2$ is small and x, y are positive, then x/y is a good approximation to $\sqrt{D}$.

- To illustrate, suppose we have a solution of $x^2 - Dy^2 = 1$.
- Dividing by y^2 yields $(x/y)^2 - D = 1/y^2$, and now solving for x/y gives $x/y = \sqrt{D + 1/y^2} = \sqrt{D} \cdot \sqrt{1 + 1/(Dy^2)} \approx \sqrt{D} \cdot (1 + 1/(2Dy^2)) = \sqrt{D} + 1/(2y^2\sqrt{D})$ using the linearization $\sqrt{1+z} \approx 1 + z/2$.
- In fact, the linearization is an overestimate since $(1 + z/2)^2 = 1 + z + z^2/4 > 1 + z$.
- Thus, we obtain the inequality $\left|\dfrac{x}{y} - \sqrt{D}\right| < \dfrac{1}{2y^2\sqrt{D}}$.

Pell's Equation Intro: Rational Approximation, II

The point is that if $x^2 - Dy^2 = 1$, then x/y is a good approximation to $\sqrt{D}$: $\left|\frac{x}{y} - \sqrt{D}\right| < \frac{1}{2y^2\sqrt{D}}$.

- In fact, the approximation is extremely good. From our results on continued fractions and rational approximation, we know that if α is irrational and p/q has the property that $|\alpha - p/q| < 1/(2q^2)$, then in fact p/q is a continued fraction convergent to α.
- So, since $\sqrt{D} > 1$, this means any solution to $x^2 - Dy^2 = 1$ must arise as a continued fraction convergent to $\sqrt{D}$.

Pell's Equation Intro: Rational Approximation, III

We can see quite explicitly that the solutions to $x^2 - 2y^2 = 1$ arise from continued fraction convergents to $\sqrt{2} = [1, \overline{2}] = [1, 2, 2, 2, \dots]$.

- The first few convergents are $1/1$, $3/2$, $7/5$, $17/12$, $41/29$, $99/70$, ... , which (as ordered pairs) have $x^2 - 2y^2$ respectively equal to -1, 1, -1, 1, -1, 1, ...
- These convergents are precisely the solutions to $x^2 - 2y^2 = \pm 1$ we identified earlier.
- We remark also that the period of the continued fraction expansion here is equal to 1 and the fundamental unit corresponds to the convergent $[1]$.

Pell's Equation Intro: Rational Approximation, IV

Let's try it out for $D = 3$.

- Here, we have $\sqrt{3} = [1, \overline{1, 2}] = [1, 1, 2, 1, 2, \dots]$ with convergents $1/1$, $2/1$, $5/3$, $7/4$, $19/11$, $26/15$, $71/41$, ...
- As ordered pairs, these convergents have $x^2 - 3y^2$ respectively equal to -2, 1, -2, 1, -2, 1, ...
- Here, we can see that we do not obtain any solutions to $x^2 - 3y^2 = -1$ (since in fact there are none as we proved earlier) but we do obtain solutions to $x^2 - 3y^2 = -2$ and $x^2 - 3y^2 = 1$.
- The period of the continued fraction expansion here is equal to 2, while the fundamental unit corresponds to the convergent $[1, 2]$.

Pell's Equation Intro: Rational Approximation, V

Let's try $D = 7$.

- Here, we have $\sqrt{7} = [2, \overline{1, 1, 1, 4}] = [2, 1, 1, 1, 4, 1, 1, 1, 4, \dots]$ with convergents $2/1$, $3/1$, $5/2$, $8/3$, $37/14$, $45/17$, $82/31$, $127/48$, $590/223$, ...
- As ordered pairs, these convergents have $x^2 - 7y^2$ respectively equal to -3, 2, -3, 1, -3, 2, -3, 1, -3, ...
- Here again we obtain no solutions to $x^2 - 3y^2 = -1$ but we do obtain solutions to $x^2 - 3y^2 = -3$, $x^2 - 3y^2 = 2$, and $x^2 - 3y^2 = 1$.
- The period of the continued fraction expansion here is equal to 4, while the fundamental unit corresponds to the convergent $[2, 1, 1, 1]$.

Pell's Equation Intro: Rational Approximation, VI

Let's try one more: $D = 13$.

- Here, $\sqrt{13} = [3, \overline{1, 1, 1, 6}] = [3, 1, 1, 1, 1, 6, \dots]$ with convergents $3/1$, $4/1$, $7/2$, $11/3$, $18/5$, $119/33$, $137/38$, $256/71$, $393/109$, $649/180$, ...
- As ordered pairs, these convergents have $x^2 - 13y^2$ respectively equal to -4, 3, -3, 4, -1, 4, -3, 3, -4, 1, ...
- Here we obtain solutions to $x^2 - 13y^2 = r$ for $r = -4, -3, -1, 1, 3, 4$.
- The period of the continued fraction expansion here is equal to 4, while the fundamental unit corresponds to the convergent $[3, 1, 1, 1, 1]$.

It appears that the fundamental unit is obtained after one period of the continued fraction expansion, regardless of whether it has norm 1 or -1. We will prove this, and some other facts, next time!

Summary

We discussed some examples and patterns in solutions to Pell's equation.

We discussed the relationships between solutions to Pell's equation and the ring structure of $\mathbb{Z}[\sqrt{D}]$.

We discussed the connection between solutions to Pell's equation and rational approximation via continued fractions.

Next lecture: Pell's Equation (part 2).