Updated: *June 16, 2010*

Diophantine approximation, irrationality and transcendence

Michel Waldschmidt

Course N°6, *May 5, 2010*

These are informal notes of my course given in April – June 2010 at IMPA (*Instituto Nacional de Matematica Pura e Aplicada*), Rio de Janeiro, Brazil.

6.3 Pell's equation

Let D be a positive integer which is not the square of an integer. It follows that $\sqrt{D}$ is an irrational number. The Diophantine equation

$$x^2 - Dy^2 = \pm 1, \tag{74}$$

where the unknowns x and y are in $\mathbf{Z}$, is called *Pell's equation.*

An introduction to the subject has been given in the colloquium lecture on April 15. We refer to

`http://seminarios.impa.br/cgi-bin/SEMINAR_palestra.cgi?id=4752`

`http://www.math.jussieu.fr/~ miw/articles/pdf/PellFermatEn2010.pdf`

and

`http://www.math.jussieu.fr/~ miw/articles/pdf/PellFermatEn2010VI.pdf`

Here we suply complete proofs of the results introduced in that lecture.

6.3.1 Examples

The three first examples below are special cases of results initiated by O. Perron and related with real quadratic fields of Richaud-Degert type.

Example 1. Take $D = a^2b^2 + 2b$ where a and b are positive integers. A solution to

$$x^2 - (a^2b^2 + 2b)y^2 = 1$$

is $(x, y) = (a^2b + 1, a)$. As we shall see, this is related with the continued fraction expansion of $\sqrt{D}$ which is

$$\sqrt{a^2b^2 + 2b} = [ab, \overline{a, 2ab}]$$

since

$$t = \sqrt{a^2b^2 + 2b} \Longleftrightarrow t = ab + \cfrac{1}{a + \cfrac{1}{t + ab}}\cdot$$

This includes the examples $D = a^2 + 2$ (take $b = 1$) and $D = b^2 + 2b$ (take $a = 1$). For $a = 1$ and $b = c - 1$ his includes the example $D = c^2 - 1$.

Example 2. Take $D = a^2b^2 + b$ where a and b are positive integers. A solution to

$$x^2 - (a^2b^2 + b)y^2 = 1$$

is $(x, y) = (2a^2b + 1, 2a)$. The continued fraction expansion of $\sqrt{D}$ is

$$\sqrt{a^2b^2 + b} = [ab, \overline{2a, 2ab}]$$

since

$$t = \sqrt{a^2b^2 + b} \Longleftrightarrow t = ab + \cfrac{1}{2a + \cfrac{1}{t + ab}}\cdot$$

This includes the example $D = b^2 + b$ (take $a = 1$).

The case $b = 1$, $D = a^2 + 1$ is special: there is an integer solution to

$$x^2 - (a^2 + 1)y^2 = -1,$$

namely $(x, y) = (a, 1)$. The continued fraction expansion of $\sqrt{D}$ is

$$\sqrt{a^2 + 1} = [a, \overline{2a}]$$

since

$$t = \sqrt{a^2 + 1} \Longleftrightarrow t = a + \frac{1}{t + a}\cdot$$

Example 3. Let a and b be two positive integers such that $b^2 + 1$ divides $2ab + 1$. For instance $b = 2$ and $a \equiv 1 \pmod 5$. Write $2ab + 1 = k(b^2 + 1)$ and take $D = a^2 + k$. The continued fraction expansion of $\sqrt{D}$ is

$$[a, \overline{b, b, 2a}]$$

since $t = \sqrt{D}$ satisfies

$$t = a + \cfrac{1}{b + \cfrac{1}{b + \cfrac{1}{a + t}}} = [a, b, b, a + z].$$

A solution to $x^2 - Dy^2 = -1$ is $x = ab^2 + a + b$, $y = b^2 + 1$.

In the case $a = 1$ and $b = 2$ (so $k = 1$), the continued fraction has period length 1 only:

$$\sqrt{5} = [1, \overline{2}].$$

Example 4. Integers which are *Polygonal numbers* in two ways are given by the solutions to quadratic equations.

Triangular numbers are numbers of the form

$$1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2} \quad \text{for } n \geq 1;$$

their sequence starts with

1, 3, 6, 10, 15, 21, 28, 36, 45, 55, 66, 78, 91, 105, 120, 136, 153, 171, ...

`http://www.research.att.com/~njas/sequences/A000217`.

Square numbers are numbers of the form

$$1 + 3 + 5 + \cdots + (2n + 1) = n^2 \quad \text{for } n \geq 1;$$

their sequence starts with

1, 4, 9, 16, 25, 36, 49, 64, 81, 100, 121, 144, 169, 196, 225, 256, 289, ...

`http://www.research.att.com/~njas/sequences/A000290`.

Pentagonal numbers are numbers of the form

$$1 + 4 + 7 + \cdots + (3n + 1) = \frac{n(3n-1)}{2} \quad \text{for } n \geq 1;$$

their sequence starts with

1, 5, 12, 22, 35, 51, 70, 92, 117, 145, 176, 210, 247, 287, 330, 376, 425, ...

`http://www.research.att.com/~njas/sequences/A000326`.

Hexagonal numbers are numbers of the form

$$1 + 5 + 9 + \cdots + (4n + 1) = n(2n - 1) \quad \text{for } n \geq 1;$$

their sequence starts with

1, 6, 15, 28, 45, 66, 91, 120, 153, 190, 231, 276, 325, 378, 435, 496, 561, ...

http://www.research.att.com/∼njas/sequences/A000384.

For instance, numbers which are at the same time triangular and squares are the numbers y^2 where (x, y) is a solution to Pell's equation with $D = 8$. Their list starts with

$$0, 1, 36, 1225, 41616, 1413721, 48024900, 1631432881, 55420693056, \ldots$$

See http://www.research.att.com/∼njas/sequences/A001110.

Example 5. Integer rectangle triangles having sides of the right angle as consecutive integers a and $a + 1$ have an hypothenuse c which satisfies $a^2 + (a + 1)^2 = c^2$. The admissible values for the hypothenuse is the set of positive integer solutions y to Pell's equation $x^2 - 2y^2 = -1$. The list of these hypothenuses starts with

$$1, 5, 29, 169, 985, 5741, 33461, 195025, 1136689, 6625109, 38613965,$$

See http://www.research.att.com/∼njas/sequences/A001653.

6.3.2 Existence of integer solutions

Let D be a positive integer which is not a square. We show that Pell's equation (74) has a non–trivial solution $(x, y) \in \mathbf{Z} \times \mathbf{Z}$, that is a solution $\neq (\pm 1, 0)$.

Proposition 75. *Given a positive integer D which is not a square, there exists $(x, y) \in \mathbf{Z}^2$ with $x > 0$ and $y > 0$ such that $x^2 - Dy^2 = 1$.*

Proof. The first step of the proof is to show that there exists a non–zero integer k such that the Diophantine equation $x^2 - Dy^2 = k$ has infinitely many solutions $(x, y) \in \mathbf{Z} \times \mathbf{Z}$. The main idea behind the proof, which will be made explicit in Lemmas 77, 78 and Corollary 79 below, is to relate the integer solutions of such a Diophantine equation with rational approximations x/y of $\sqrt{D}$.

Using the implication (i) $\Rightarrow$ (v) of the irrationality criterion 4 and the fact that $\sqrt{D}$ is irrational, we deduce that there are infinitely many $(x, y) \in \mathbf{Z} \times \mathbf{Z}$ with $y > 0$ (and hence $x > 0$) satisfying

$$\left| \sqrt{D} - \frac{x}{y} \right| < \frac{1}{y^2}.$$

For such a (x, y), we have $0 < x < y\sqrt{D} + 1 < y(\sqrt{D} + 1)$, hence

$$0 < |x^2 - Dy^2| = |x - y\sqrt{D}| \cdot |x + y\sqrt{D}| < 2\sqrt{D} + 1.$$

Since there are only finitely integers $k \neq 0$ in the range

$$-(2\sqrt{D}+1) < k < 2\sqrt{D}+1,$$

one at least of them is of the form $x^2 - Dy^2$ for infinitely many (x, y).

The second step is to notice that, since the subset of $(x, y) \pmod{k}$ in $(\mathbf{Z}/k\mathbf{Z})^2$ is finite, there is an infinite subset $E \subset \mathbf{Z} \times \mathbf{Z}$ of these solutions to $x^2 - Dy^2 = k$ having the same $(x \pmod{k}, y \pmod{k})$.

Let (u_1, v_1) and (u_2, v_2) be two distinct elements in E. Define $(x, y) \in \mathbf{Q}^2$ by

$$x + y\sqrt{D} = \frac{u_1 + v_1\sqrt{D}}{u_2 + v_2\sqrt{D}}\cdot$$

From $u_2^2 - Dv_2^2 = k$, one deduces

$$x + y\sqrt{D} = \frac{1}{k}(u_1 + v_1\sqrt{D})(u_2 - v_2\sqrt{D}),$$

hence

$$x = \frac{u_1u_2 - Dv_1v_2}{k}, \quad y = \frac{-u_1v_2 + u_2v_1}{k}\cdot$$

From $u_1 \equiv u_2 \pmod{k}$, $v_1 \equiv v_2 \pmod{k}$ and

$$u_1^2 - Dv_1^2 = k, \quad u_2^2 - Dv_2^2 = k,$$

we deduce

$$u_1u_2 - Dv_1v_2 \equiv u_1^2 - Dv_1^2 \equiv 0 \pmod{k}$$

and

$$-u_1v_2 + u_2v_1 \equiv -u_1v_1 + u_1v_1 \equiv 0 \pmod{k},$$

hence x and y are in $\mathbf{Z}$. Further,

$$\begin{aligned}
x^2 - Dy^2 &= (x + y\sqrt{D})(x - y\sqrt{D}) \\
&= \frac{(u_1 + v_1\sqrt{D})(u_1 - v_1\sqrt{D})}{(u_2 + v_2\sqrt{D})(u_2 - v_2\sqrt{D})} \\
&= \frac{u_1^2 - Dv_1^2}{u_2^2 - Dv_2^2} = 1.
\end{aligned}$$

It remains to check that $y \neq 0$. If $y = 0$ then $x = \pm 1$, $u_1v_2 = u_2v_1$, $u_1u_2 - Dv_1v_2 = \pm 1$, and

$$ku_1 = \pm u_1(u_1u_2 - Dv_1v_2) = \pm u_2(u_1^2 - Dv_1^2) = \pm ku_2,$$

which implies $(u_1, u_2) = (v_1, v_2)$, a contradiction.

Finally, if $x < 0$ (resp. $y < 0$) we replace x by $-x$ (resp. y by $-y$).

□

Once we have a non–trivial integer solution (x, y) to Pell's equation, we have infinitely many of them, obtained by considering the powers of $x+y\sqrt{D}$.

6.3.3 All integer solutions

There is a natural order for the positive integer solutions to Pell's equation: we can order them by increasing values of x, or increasing values of y, or increasing values of $x+y\sqrt{D}$ - it is easily checked that the order is the same.

It follows that there is a minimal positive integer solution[8] (x_1, y_1), which is called *the fundamental solution to Pell's equation* $x^2 - Dy^2 = \pm 1$. In the same way, there is a fundamental solution to Pell's equations $x^2 - Dy^2 = 1$. Furthermore, when the equation $x^2 - Dy^2 = -1$ has an integer solution, then there is also a fundamental solution.

Proposition 76. *Denote by (x_1, y_1) the fundamental solution to Pell's equation $x^2 - Dy^2 = \pm 1$. Then the set of all positive integer solutions to this equation is the sequence $(x_n, y_n)_{n\geq 1}$, where x_n and y_n are given by*

$$x_n + y_n\sqrt{D} = (x_1 + y_1\sqrt{D})^n, \quad (n \in \mathbf{Z}, \quad n \geq 1).$$

In other terms, x_n and y_n are defined by the recurrence formulae

$$x_{n+1} = x_n x_1 + D y_n y_1 \qquad \text{and} \qquad y_{n+1} = x_1 y_n + x_n y_1, \qquad (n \geq 1).$$

More explicitly:
• *If $x_1^2 - Dy_1^2 = 1$, then (x_1, y_1) is the fundamental solution to Pell's equation $x^2 - Dy^2 = 1$, and there is no integer solution to Pell's equation $x^2 - Dy^2 = -1$.*
• *If $x_1^2 - Dy_1^2 = -1$, then (x_1, y_1) is the fundamental solution to Pell's equation $x^2 - Dy^2 = -1$, and the fundamental solution to Pell's equation $x^2 - Dy^2 = 1$ is (x_2, y_2). The set of positive integer solutions to Pell's equation $x^2 - Dy^2 = 1$ is $\{(x_n, y_n)\ ;\ n \geq 2\ \text{ even}\}$, while the set of positive integer solutions to Pell's equation $x^2 - Dy^2 = -1$ is $\{(x_n, y_n)\ ;\ n \geq 1\ \text{ odd}\}$.*
The set of all solutions $(x, y) \in \mathbf{Z} \times \mathbf{Z}$ to Pell's equation $x^2 - Dy^2 = \pm 1$ is the set $(\pm x_n, y_n)_{n\in\mathbf{Z}}$, where x_n and y_n are given by the same formula

$$x_n + y_n\sqrt{D} = (x_1 + y_1\sqrt{D})^n, \quad (n \in \mathbf{Z}).$$

The trivial solution $(1, 0)$ is (x_0, y_0), the solution $(-1, 0)$ is a torsion element of order 2 in the group of units of the ring $\mathbf{Z}[\sqrt{D}]$.

[8] We use the letter x_1, which should not be confused with the first complete quotient in the section § SSS:InfiniteSCF on continued fractions

Proof. Let (x, y) be a positive integer solution to Pell's equation $x^2 - Dy^2 = \pm 1$. Denote by $n \geq 0$ the largest integer such that

$$(x_1 + y_1\sqrt{D})^n \leq x + y\sqrt{D}.$$

Hence $x + y\sqrt{D} < (x_1 + y_1\sqrt{D})^{n+1}$. Define $(u, v) \in \mathbf{Z} \times \mathbf{Z}$ by

$$u + v\sqrt{D} = (x + y\sqrt{D})(x_1 - y_1\sqrt{D})^n.$$

From

$$u^2 - Dv^2 = \pm 1 \quad \text{and} \quad 1 \leq u + v\sqrt{D} < x_1 + y_1\sqrt{D},$$

we deduce $u = 1$ and $v = 0$, hence $x = x_n$, $y = y_n$.

□

6.3.4 On the group of units of $\mathbf{Z}[\sqrt{D}]$

Let D be a positive integer which is not a square. The ring $\mathbf{Z}[\sqrt{D}]$ is the subring of $\mathbf{R}$ generated by $\sqrt{D}$. The map $\sigma : z = x + y\sqrt{D} \longmapsto x - y\sqrt{D}$ is the *Galois automorphism* of this ring. The *norm* $N : \mathbf{Z}[\sqrt{D}] \longrightarrow \mathbf{Z}$ is defined by $N(z) = z\sigma(z)$. Hence

$$N(x + y\sqrt{D}) = x^2 - Dy^2.$$

The restriction of N to the group of unit $\mathbf{Z}[\sqrt{D}]^\times$ of the ring $\mathbf{Z}[\sqrt{D}]$ is a homomorphism from the multiplicative group $\mathbf{Z}[\sqrt{D}]^\times$ to the group of units $\mathbf{Z}^\times$ of $\mathbf{Z}$. Since $\mathbf{Z}^\times = \{\pm 1\}$, it follows that

$$\mathbf{Z}[\sqrt{D}]^\times = \{z \in \mathbf{Z}[\sqrt{D}] \; ; \; N(z) = \pm 1\},$$

hence $\mathbf{Z}[\sqrt{D}]^\times$ is nothing else than the set of $x + y\sqrt{D}$ when (x, y) runs over the set of integer solutions to Pell's equation $x^2 - Dy^2 = \pm 1$.

Proposition 75 means that $\mathbf{Z}[\sqrt{D}]^\times$ is not reduced to the torsion subgroup ± 1, while Proposition 76 gives the more precise information that this group $\mathbf{Z}[\sqrt{D}]^\times$ is a (multiplicative) abelian group of rank 1: there exists a so–called *fundamental unit* $u \in \mathbf{Z}[\sqrt{D}]^\times$ such that

$$\mathbf{Z}[\sqrt{D}]^\times = \{\pm u^n \; ; \; n \in \mathbf{Z}\}.$$

The fundamental unit $u > 1$ is $x_1 + y_1\sqrt{D}$, where (x_1, y_1) is the fundamental solution to Pell's equation $x^2 - Dy^2 = \pm 1$. Pell's equation $x^2 - Dy^2 = \pm 1$ has integer solutions if and only if the fundamental unit has norm -1.

That the rank of $\mathbf{Z}[\sqrt{D}]^\times$ is at most 1 also follows from the fact that the image of the map

$$\begin{array}{ccc} \mathbf{Z}[\sqrt{D}]^\times & \longrightarrow & \mathbf{R}^2 \\ z & \longmapsto & \big(\log|z|,\ \log|z'|\big) \end{array}$$

is discrete in $\mathbf{R}^2$ and contained in the line $t_1 + t_2 = 0$ of $\mathbf{R}^2$. This proof is not really different from the proof we gave of Proposition 76: the proof that the discrete subgroups of $\mathbf{R}$ have rank ≤ 1 relies on Euclid's division.

6.3.5 Connection with rational approximation

Lemma 77. *Let D be a positive integer which is not a square. Let x and y be positive rational integers. The following conditions are equivalent:*
(i) $x^2 - Dy^2 = 1$.
(ii) $0 < \dfrac{x}{y} - \sqrt{D} < \dfrac{1}{2y^2\sqrt{D}}\cdot$
(iii) $0 < \dfrac{x}{y} - \sqrt{D} < \dfrac{1}{y^2\sqrt{D}+1}\cdot$

Proof. We have $\dfrac{1}{2y^2\sqrt{D}} < \dfrac{1}{y^2\sqrt{D}+1}$, hence (ii) implies (iii).
(i) implies $x^2 > Dy^2$, hence $x > y\sqrt{D}$, and consequently

$$0 < \frac{x}{y} - \sqrt{D} = \frac{1}{y(x+y\sqrt{D})} < \frac{1}{2y^2\sqrt{D}}\cdot$$

(iii) implies

$$x < y\sqrt{D} + \frac{1}{y\sqrt{D}} < y\sqrt{D} + \frac{2}{y},$$

and

$$y(x+y\sqrt{D}) < 2y^2\sqrt{D} + 2,$$

hence

$$0 < x^2 - Dy^2 = y\left(\frac{x}{y} - \sqrt{D}\right)(x+y\sqrt{D}) < 2.$$

Since $x^2 - Dy^2$ is an integer, it is equal to 1. □

The next variant will also be useful.

Lemma 78. *Let D be a positive integer which is not a square. Let x and y be positive rational integers. The following conditions are equivalent:*

(i) $x^2 - Dy^2 = -1$.

(ii) $0 < \sqrt{D} - \dfrac{x}{y} < \dfrac{1}{2y^2\sqrt{D} - 1}$.

(iii) $0 < \sqrt{D} - \dfrac{x}{y} < \dfrac{1}{y^2\sqrt{D}}$.

Proof. We have $\dfrac{1}{2y^2\sqrt{D} - 1} < \dfrac{1}{y^2\sqrt{D}}$, hence (ii) implies (iii).
The condition (i) implies $y\sqrt{D} > x$. We use the trivial estimate

$$2\sqrt{D} > 1 + 1/y^2$$

and write

$$x^2 = Dy^2 - 1 > Dy^2 - 2\sqrt{D} + 1/y^2 = (y\sqrt{D} - 1/y)^2,$$

hence $xy > y^2\sqrt{D} - 1$. From (i) one deduces

$$\begin{aligned} 1 = Dy^2 - x^2 &= (y\sqrt{D} - x)(y\sqrt{D} + x) \\ &> \left(\sqrt{D} - \frac{x}{y}\right)(y^2\sqrt{D} + xy) \\ &> \left(\sqrt{D} - \frac{x}{y}\right)(2y^2\sqrt{D} - 1). \end{aligned}$$

(iii) implies $x < y\sqrt{D}$ and

$$y(y\sqrt{D} + x) < 2y^2\sqrt{D},$$

hence

$$0 < Dy^2 - x^2 = y\left(\sqrt{D} - \frac{x}{y}\right)(y\sqrt{D} + x) < 2.$$

Since $Dy^2 - x^2$ is an integer, it is 1. □

From these two lemmas one deduces:

Corollary 79. *Let D be a positive integer which is not a square. Let x and y be positive rational integers. The following conditions are equivalent:*
(i) $x^2 - Dy^2 = \pm 1$.

(ii) $\left|\sqrt{D} - \dfrac{x}{y}\right| < \dfrac{1}{2y^2\sqrt{D} - 1}$.

(iii) $\left|\sqrt{D} - \dfrac{x}{y}\right| < \dfrac{1}{y^2\sqrt{D} + 1}$.

Proof. If $y > 1$ or $D > 3$ we have $2y^2\sqrt{D} - 1 > y^2\sqrt{D} + 1$, which means that (ii) implies trivially (iii), and we may apply Lemmas 77 and 78.

If $D = 2$ and $y = 1$, then each of the conditions (i), (ii) and (iii) is satisfied if and only if $x = 1$. This follows from

$$2 - \sqrt{2} > \frac{1}{2\sqrt{2} - 1} > \frac{1}{\sqrt{2} + 1} > \sqrt{2} - 1.$$

If $D = 3$ and $y = 1$, then each of the conditions (i), (ii) and (iii) is satisfied if and only if $x = 2$. This follows from

$$3 - \sqrt{3} > \sqrt{3} - 1 > \frac{1}{2\sqrt{3} - 1} > \frac{1}{\sqrt{3} + 1} > 2 - \sqrt{3}.$$

□

It is instructive to compare with Liouville's inequality (see § 5.2).

Lemma 80. *Let D be a positive integer which is not a square. Let x and y be positive rational integers. Then*

$$\left|\sqrt{D} - \frac{x}{y}\right| > \frac{1}{2y^2\sqrt{D} + 1}\cdot$$

Proof. If $x/y < \sqrt{D}$, then $x \le y\sqrt{D}$ and from

$$1 \le Dy^2 - x^2 = (y\sqrt{D} + x)(y\sqrt{D} - x) \le 2y\sqrt{D}(y\sqrt{D} - x),$$

one deduces

$$\sqrt{D} - \frac{x}{y} > \frac{1}{2y^2\sqrt{D}}\cdot$$

We claim that if $x/y > \sqrt{D}$, then

$$\frac{x}{y} - \sqrt{D} > \frac{1}{2y^2\sqrt{D} + 1}\cdot$$

Indeed, this estimate is true if $x - y\sqrt{D} \ge 1/y$, so we may assume $x - y\sqrt{D} < 1/y$. Our claim then follows from

$$1 \le x^2 - Dy^2 = (x + y\sqrt{D})(x - y\sqrt{D}) \le (2y\sqrt{D} + 1/y)(x - y\sqrt{D}).$$

□

This shows that a rational approximation x/y to $\sqrt{D}$, which is only slightly weaker than the limit given by Liouville's inequality, will produce a solution to Pell's equation $x^2 - Dy^2 = \pm 1$. The distance $|\sqrt{D} - x/y|$ cannot be smaller than $1/(2y^2\sqrt{D} + 1)$, but it can be as small as $1/(2y^2\sqrt{D} - 1)$, and for that it suffices that it is less than $1/(y^2\sqrt{D} + 1)$

Michel WALDSCHMIDT
Université P. et M. Curie (Paris VI)
Institut Mathématique de Jussieu
Problèmes Diophantiens, Case 247
4, Place Jussieu
75252 Paris CEDEX 05, France
miw@math.jussieu.fr
http://www.math.jussieu.fr/∼miw/

This text is available on the internet at the address
http://www.math.jussieu.fr/∼miw/enseignement.html