PELL'S EQUATION, I

KEITH CONRAD

1. Introduction

For a positive integer d that is not a square, an equation of the form

$$x^2 - dy^2 = 1$$

is called *Pell's equation*. We are interested in solutions (x, y) where x and y are integers. The term "solution" will always mean that kind of solution. The obvious solutions $(x, y) = (\pm 1, 0)$ are called the *trivial solutions*. They are the only solutions where $x = \pm 1$ or $y = 0$ (separately). Solutions where $x > 0$ and $y > 0$ will be called *positive solutions*. Every nontrivial solution can be made into a positive solution by changing the sign of x or y.

We don't consider the case when d is a square, since if $d = c^2$ with $c \in \mathbf{Z}$ then $x^2 - dy^2 = x^2 - (cy)^2$ and the only squares that differ by 1 are 0 and 1, so $x^2 - (cy)^2 = 1 \Longrightarrow x = \pm 1$ and $y = 0$. Thus $x^2 - dy^2 = 1$ for square d only has the integral solutions $(x, y) = (\pm 1, 0)$.

In Section 2 we'll show how solutions to Pell's equation can be found. In Section 3 we'll discuss an elementary problem about polygonal numbers that is equivalent to a specific Pell equation. Section 4 describes how to create new solutions of Pell's equation if we know one nontrivial solution and in Section 5 we will see how all solutions can be generated from a minimal nontrivial solution. In Section 6 a generalized Pell equation is introduced, where the right side is not 1. In Section 7 we look at the Pell-type equation with right side -1.

2. Examples of Solutions

To find a nontrivial solution of $x^2 - dy^2 = 1$ by elementary methods, rewrite the equation as $x^2 = dy^2 + 1$ and then set $y = 1, 2, 3, \ldots$ until you reach a value where $dy^2 + 1$ is a perfect square. Call that value x^2 and then we have a solution (x, y).

Example 2.1. Two positive solutions of $x^2 - 2y^2 = 1$ are $(3, 2)$ and $(17, 12)$, since $2y^2 + 1$ is a square when $y = 2$ and 12, where it has values $9 = 3^2$ and $289 = 17^2$. See below.

y	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$2y^2+1$	3	9	19	33	51	73	99	129	163	201	243	289	339	393	451
Square?	X	✓	X	X	X	X	X	X	X	X	X	✓	X	X	X

Example 2.2. Three positive solutions of $x^2 - 3y^2 = 1$ are $(2, 1)$ and $(7, 4)$, and $(26, 15)$, as shown by the table below.

y	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$3y^2+1$	4	13	28	49	76	109	148	193	244	301	364	433	508	589	676
Square?	✓	X	X	✓	X	X	X	X	X	X	X	X	X	X	✓

The table below gives a positive solution to $x^2 - dy^2 = 1$ for nonsquare d from 2 to 24 where x and y are as small as possible.

d	2	3	5	6	7	8	10	11	12	13	14	15	17	18	19	20	21	22	23	24
x	3	2	9	5	8	3	19	10	7	649	15	4	33	17	170	9	55	197	24	5
y	2	1	4	2	3	1	6	3	2	180	4	1	8	4	39	2	12	42	5	1

The theorem suggested by such data is a hard result of Lagrange.

Theorem 2.3 (Lagrange). *For all $d \in \mathbf{Z}^+$ that are not squares, the equation $x^2 - dy^2 = 1$ has a nontrivial solution.*

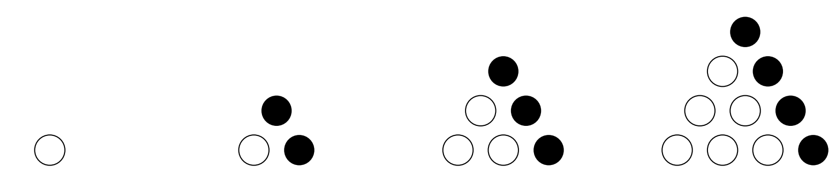
This theorem, which will be proved in Part II[1], is our hunting license to search for solutions by tabulating $dy^2 + 1$ until it takes a square value. We are guaranteed this search will eventually terminate, but we are *not* assured how long it will take. In fact, the smallest positive solution of $x^2 - dy^2 = 1$ can be unusually large compared to the size of d. The table above illustrates this if we compare the smallest positive solutions when $d = 12$, 13, and 14. As more extreme examples, see in the table below the smallest positive solutions to $x^2 - dy^2 = 1$ when $d = 61$ and 109 compared with nearby values of d.

d	60	61	62	108	109	110
x	31	1766319049	63	1351	158070671986249	21
y	4	226153980	8	130	15140424455100	2

While Lagrange was the first person to give a proof of Theorem 2.3, in 1768, a century earlier Fermat claimed to have a proof and challenged other mathematicians in Europe to prove it too. In a letter in 1657 he wrote that anyone failing this task should at least try to find a positive solution to $x^2 - 61y^2 = 1$ and $x^2 - 109y^2 = 1$, where he said he chose small coefficients "pour ne vous donner pas trop de peine" (so you don't have too much work). He clearly was being mischievous. Fermat had no idea that a nontrivial solution to $x^2 - 61y^2 = 1$ had been found in India (by Bhaskara II) 500 years before him.

3. Triangular–Square Numbers

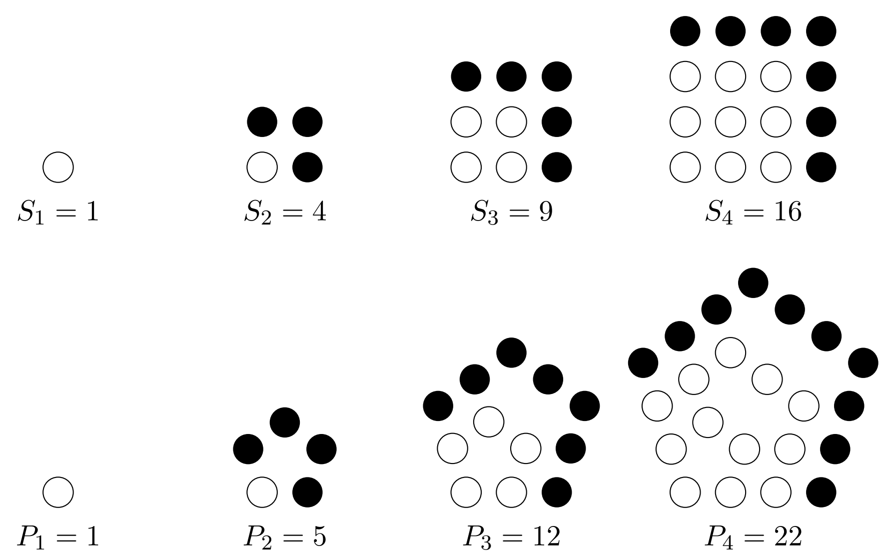
A positive integer n is called *triangular* if n dots can be arranged like an equilateral triangle. The first four triangular numbers are 1 (a degenerate case), 3, 6, and 10. In the pictures below, a new (shaded) side added to a triangular number leads to the next one.



$T_1 = 1$ $\quad$ $T_2 = 3$ $\quad$ $T_3 = 6$ $\quad$ $T_4 = 10$

For $k \geq 3$, a *k-gonal number* is a positive integer n for which n dots can be arranged to look like a regular k-gon. The first four square and pentagonal numbers, corresponding to $k = 4$ and $k = 5$, are shown below. Both sequences start with 1 as a degenerate case.

[1]See https://kconrad.math.uconn.edu/blurbs/ugradnumthy/pelleqn2.pdf.



A formula for the nth square number S_n is obvious: $S_n = n^2$. To get a formula for the nth triangular and pentagonal numbers, T_n and P_n, the first few values suggest how to write them as a sum of terms in an arithmetic progression (which are their real definitions):

$$T_n = 1 + 2 + \cdots + n = \sum_{k=1}^{n} k, \quad P_n = 1 + 4 + \cdots + (3n-2) = \sum_{k=1}^{n} (3k-2).$$

This works for square numbers too: $S_n = 1 + 3 + \cdots + (2n-1) = \sum_{k=1}^{n}(2k-1)$ is n^2. Using the formula for the sum of terms in an arithmetic progression,

$$T_n = \frac{n(n+1)}{2} \quad \text{and} \quad P_n = \frac{n(3n-1)}{2}.$$

With these formulas we fill in the table below of the first 10 triangular, square, and pentagonal numbers.

n	1	2	3	4	5	6	7	8	9	10
T_n	1	3	6	10	15	21	28	36	45	55
S_n	1	4	9	16	25	36	49	64	81	100
P_n	1	5	12	22	35	51	70	92	117	145

Besides the common value 1, we see 36 is both triangular and square: $36 = T_8 = S_6$. Call a positive integer a *triangular–square number* if it is both T_m for some m and S_n for some n. Finding these numbers is the same as solving a particular Pell equation.

Theorem 3.1. *Triangular–square numbers correspond to solutions of $x^2 - 2y^2 = 1$ in positive integers x and y.*

Proof. Using the formulas for T_m and S_n,

$$\begin{aligned} T_m = S_n &\iff \frac{m(m+1)}{2} = n^2 \\ &\iff m^2 + m = 2n^2 \\ &\iff \left(m + \frac{1}{2}\right)^2 - \frac{1}{4} = 2n^2 \\ &\iff (2m+1)^2 - 1 = 2(2n)^2 \\ &\iff (2m+1)^2 - 2(2n)^2 = 1. \end{aligned}$$

Because every step is *reversible*, finding triangular–square numbers is equivalent to solving $x^2 - 2y^2 = 1$ in positive integers x and y where $x = 2m+1$ is odd and $y = 2n$ is even: $T_{(x-1)/2} = S_{y/2}$. (While we want $x = 2m+1$ with $m \geq 1$, we can say $x > 0$ instead of $x \geq 3$ because the only solution of $x^2 - 2y^2 = 1$ with $x = 1$ has $y = 0$, which is not positive.)

In the correspondence between triangular–square numbers and solutions of $x^2 - 2y^2 = 1$, we do not need to specify that x is odd and y is even since those constraints are *forced* by the equation $x^2 - 2y^2 = 1$. Indeed, rewriting it as $x^2 = 2y^2 + 1$ shows x^2 is odd, so x is odd. Then $x = 2m + 1$ for an integer m, and feeding that into the Pell equation makes $4m^2 + 4m + 1 - 2y^2 = 1$, so $y^2 = 2m^2 + 2m$. Thus y^2 is even, so y is even. □

Example 3.2. From the solutions $(x, y) = (3, 2)$ and $(17, 12)$ of $x^2 - 2y^2 = 1$ we get the triangular–square numbers $T_1 = S_1 = 1$ and $T_8 = S_6 = 36$ by writing $x = 2m + 1$ and $y = 2n$ in each case to find m and n.

As practice with the ideas in the proof of Theorem 3.1, show that finding all square–pentagonal numbers, which are numbers of the form S_m and P_n for some positive integers m and n, is equivalent to solving $x^2 - 6y^2 = 1$ in positive integers x and y satisfying the constraint that x is one less than a multiple of 6. The first three positive integer solutions of $x^2 - 6y^2 = 1$ are $(5, 2), (49, 20), (485, 198)$, and the first and third have x being one less than a multiple of 6; they lead to the square–pentagonal numbers $1 = S_1 = P_1$ and $9801 = S_{99} = P_{81}$ if you work out the details.

4. New Solutions from Old Solutions

We found in Section 2, by making a table, that two solutions of $x^2 - 2y^2 = 1$ are $(3, 2)$ and $(17, 12)$. They are closely related when we convert the pair (x, y) into the number $x + y\sqrt{2}$:

$$17 + 12\sqrt{2} = (3 + 2\sqrt{2})^2. \tag{4.1}$$

Let's raise $3 + 2\sqrt{2}$ to a few powers beyond the second:

$$(3+2\sqrt{2})^3 = 99+70\sqrt{2}, \quad (3+2\sqrt{2})^4 = 577+408\sqrt{2}, \quad (3+2\sqrt{2})^5 = 3363+2378\sqrt{2}. \tag{4.2}$$

The coefficient pairs (99,70), (577,408), and (3363,2378) are all solutions to $x^2 - 2y^2 = 1$.

Similarly, we previously found three solutions of $x^2 - 3y^2 = 1$: $(2, 1)$, $(7, 4)$, and $(26, 15)$. When we convert the pair (x, y) into the number $x + y\sqrt{3}$ we have

$$7 + 4\sqrt{3} = (2 + \sqrt{3})^2 \quad \text{and} \quad 26 + 15\sqrt{3} = (2 + \sqrt{3})^3. \tag{4.3}$$

The key to solving $x^2 - dy^2 = 1$ in $\mathbf{Z}$ is to study numbers of the form $x + y\sqrt{d}$ where $x, y \in \mathbf{Z}$. Such numbers are *closed under multiplication*:

$$(x + y\sqrt{d})(x' + y'\sqrt{d}) = (xx' + dyy') + (xy' + yx')\sqrt{d} \tag{4.4}$$

and $xx' + dyy'$ and $xy' + yx'$ are both integers.[2] This formula is similar to the rule for multiplying complex numbers: $(x + yi)(x' + y'i) = (xx' - yy') + (xy' + yx')i$, which is the case $d = -1$ (for Pell's equation we are taking $d > 0$).

Just as a complex number $x + yi$ has a real part x and an imaginary part y, a number $x + y\sqrt{d}$ with $x, y \in \mathbf{Z}$ has coefficients x and y. The coefficients of such a number are unique: if $x+y\sqrt{d} = x'+y'\sqrt{d}$ with $x, y, x', y' \in \mathbf{Z}$ then $x = x'$ and $y = y'$. Indeed, if $y \neq y'$ then $\sqrt{d} = (x - x')/(y' - y)$ is rational, which is a contradiction (nonsquare integers have irrational square roots). Thus $y = y'$, so $x + y\sqrt{d} = x' + y\sqrt{d}$, which implies $x = x'$.

Theorem 4.1. *If $X^2 - dY^2 = 1$ has solutions (x, y) and (x', y') then the coefficients of $(x + y\sqrt{d})(x' + y'\sqrt{d})$ are also a solution.*

Proof. Using the coefficients from (4.4) we compute

$$\begin{aligned}(xx' + dyy')^2 - d(xy' + yx')^2 &= (x^2x'^2 + 2dxx'yy' + d^2y^2y'^2) - d(x^2y'^2 + 2xx'yy' + y^2x'^2)\\ &= x^2x'^2 + d^2y^2y'^2 - dx^2y'^2 - dy^2x'^2\\ &= x^2(x'^2 - dy'^2) - dy^2(x'^2 - dy'^2)\\ &= (x^2 - dy^2)(x'^2 - dy'^2)\\ &= 1.\end{aligned}$$

□

Corollary 4.2. *If $X^2 - dY^2 = 1$ has a solution (x, y) then the coefficients of $(x+y\sqrt{d})^k$ are also a solution for all $k \in \mathbf{Z}$. In particular, this Pell equation has infinitely many solutions if it has a nontrivial solution.*

Proof. The coefficients of $(x + y\sqrt{d})^k$ are solutions for $k \geq 1$ by repeated multiplication using Theorem 4.1. If $(x, y) \neq (\pm 1, 0)$ then $x + y\sqrt{d} \neq \pm 1$, so the powers $(x + y\sqrt{d})^k$ for $k \geq 1$ are distinct and give us infinitely many solutions of $X^2 - dY^2 = 1$.

To show the coefficients of $(x + y\sqrt{d})^k$ are solutions for $k < 0$, write $k = -K$ and set $(x + y\sqrt{d})^K = x_K + y_K\sqrt{d}$ with $x_K, y_K \in \mathbf{Z}$. Then $x_K^2 - dy_K^2 = 1$, so

$$\begin{aligned}(x + y\sqrt{d})^{-K} &= \frac{1}{(x + y\sqrt{d})^K}\\ &= \frac{1}{x_K + y_K\sqrt{d}}\\ &= \frac{x_K - y_K\sqrt{d}}{(x_K + y_K\sqrt{d})(x_K - y_K\sqrt{d})}\\ &= \frac{x_K - y_K\sqrt{d}}{x_K^2 - dy_K^2}\\ &= x_K - y_K\sqrt{d}\end{aligned}$$

and $(x_K, -y_K)$ is a solution. Finally, the coefficients of $(x + y\sqrt{d})^0$ are $(1, 0)$. □

Example 4.3. Since $(3 + 2\sqrt{2})^4 = 577 + 408\sqrt{2}$, we have $(3 + 2\sqrt{2})^{-4} = 577 - 408\sqrt{2}$.

If we were not dealing with solutions of Pell's equation, negative powers would not have integer coefficients. *e.g.*, $(5 + 2\sqrt{2})^{-1} = 5/17 - (2/17)\sqrt{2}$.

[2] If we use a cube root instead of a square root, such sums would not be closed under multiplication, *e.g.*, $(1 + \sqrt[3]{2})(1 - \sqrt[3]{2}) = 1 - \sqrt[3]{4} \neq x + y\sqrt[3]{2}$ for x and y in $\mathbf{Z}$.

5. All Solutions to a Pell Equation

We will describe all solutions to $x^2 - dy^2 = 1$ using inequalities on numbers $x + y\sqrt{d}$. Comparing the size of such numbers is not generally the same as comparing coefficients: $x + y\sqrt{d} < x' + y'\sqrt{d}$ is not the same as $x < x'$ and $y < y'$. Consider $1 + 2\sqrt{2} < 7 - \sqrt{2}$. But for Pell solutions, under a mild condition it is the same!

Lemma 5.1. *If $x^2 - dy^2 = 1$ in $\mathbf{Z}$ and $x + y\sqrt{d} > 1$ then $x \geq 2$ and $y \geq 1$.*

The example of $x^2 - 3y^2 = 1$ for $x = 2$ and $y = 1$ shows the lower bounds on x and y can be achieved.

Proof. The crucial point is that $1/(x + y\sqrt{d}) = x - y\sqrt{d}$ when $x^2 - dy^2 = 1$. Therefore

$$x + y\sqrt{d} > 1 > x - y\sqrt{d} > 0.$$

From $x + y\sqrt{d} > x - y\sqrt{d}$ we get $2y\sqrt{d} > 0$, so $y > 0$. Thus $y \geq 1$ since y is an integer, so $x > y\sqrt{d} \geq \sqrt{d} > 1$, which makes $x \geq 2$. $\square$

Without a hypothesis like $x^2 - dy^2 = 1$ in Lemma 5.1 there are counterexamples: we can have $x + y\sqrt{d} > 1$ when $x < 1$ or $y < 0$. For example, $-2 + 3\sqrt{2} > 1$ and $5 - \sqrt{2} > 1$.

Lemma 5.2. *Suppose $x^2 - dy^2 = 1$ and $a^2 - db^2 = 1$ in $\mathbf{Z}$ where $x, y, a, b \geq 0$. Then*

$$a + b\sqrt{d} < x + y\sqrt{d} \Longleftrightarrow a < x \text{ and } b < y \Longleftrightarrow a < x \text{ or } b < y.$$

Proof. The first ($\Leftarrow$) is obvious. To prove the first ($\Rightarrow$), we have $a \geq 1$ since $a \geq 0$ by hypothesis and from $a^2 - db^2 = 1$ we can't have $a = 0$ (why?). Similarly, $x \geq 1$. Reciprocating the inequality of positive numbers

$$a + b\sqrt{d} < x + y\sqrt{d}$$

gives us

$$x - y\sqrt{d} < a - b\sqrt{d}.$$

Adding these inequalities, we get

$$(a + x) + (b - y)\sqrt{d} < (a + x) + (y - b)\sqrt{d}.$$

Subtracting $a + x$ from both sides and dividing by $\sqrt{d}$ we get $b - y < y - b$, so $2b < 2y$ and thus $b < y$. Then $a^2 = 1 + db^2 < 1 + dy^2 = x^2$, so $a < x$ from nonnegativity of a and x.

If $a + b\sqrt{d} = x + y\sqrt{d}$ then $a = x$ and $b = y$ since $\sqrt{d}$ is irrational. If $a + b\sqrt{d} > x + y\sqrt{d}$ then $x + y\sqrt{d} < a + b\sqrt{d}$, so $x < a$ and $y < b$ by swapping the roles of a and b with x and y above. Thus the only way to have $a < x$ or $b < y$ is to have $a + b\sqrt{d} < x + y\sqrt{d}$. $\square$

Theorem 5.3. *Assume $x^2 - dy^2 = 1$ has a solution in positive integers and let (x_1, y_1) be such a solution where y_1 is minimal. Then all solutions to $x^2 - dy^2 = 1$ in integers are, up to sign, generated from (x_1, y_1) by taking integral powers of $x_1 + y_1\sqrt{d}$:*

$$x + y\sqrt{d} = \pm(x_1 + y_1\sqrt{d})^k$$

where $k \in \mathbf{Z}$. The solutions in positive integers have $k \geq 1$ and the $+$ sign out front.

Proof. By Corollary 4.2, for each $k \in \mathbf{Z}$ the coefficients of $(x_1 + y_1\sqrt{d})^k$ satisfy $x^2 - dy^2 = 1$, and clearly this is also true for coefficients of $-(x_1 + y_1\sqrt{d})^k$.

Conversely, suppose integers x and y satisfy $x^2 - dy^2 = 1$. If x and y are positive we'll show $x + y\sqrt{d} = (x_1 + y_1\sqrt{d})^k$ for some $k \geq 1$. Since $x + y\sqrt{d} > 1$ and the numbers

$(x_1+y_1\sqrt{d})^n$ for $n=0,1,2,\ldots$ are an increasing sequence that starts at $(x_1+y_1\sqrt{d})^0=1$ and tends to ∞, $x+y\sqrt{d}$ lies between two powers of $x_1+y_1\sqrt{d}$ or equals one of them:

$$(x_1+y_1\sqrt{d})^k \le x+y\sqrt{d} < (x_1+y_1\sqrt{d})^{k+1} \tag{5.1}$$

for some integer $k\ge 0$. Dividing through (5.1) by $(x_1+y_1\sqrt{d})^k$,

$$1 \le (x+y\sqrt{d})(x_1+y_1\sqrt{d})^{-k} < x_1+y_1\sqrt{d}.$$

The number $(x+y\sqrt{d})(x_1+y_1\sqrt{d})^{-k}$ has coefficients that are a Pell solution since Pell solutions are closed under multiplying and raising to integer powers (Theorem 4.1, Corollary 4.2). Setting $(x+y\sqrt{d})(x_1+y_1\sqrt{d})^{-k}=a+b\sqrt{d}$ for $a,b\in\mathbf{Z}$, we have $a^2-db^2=1$ and

$$1 \le a+b\sqrt{d} < x_1+y_1\sqrt{d}. \tag{5.2}$$

If $a+b\sqrt{d}>1$ then a and b are positive by Lemma 5.1, so $b<y_1$ by (5.2) and Lemma 5.2. This contradicts the minimality of y_1 among positive Pell solutions, so $a+b\sqrt{d}=1$. That implies $x+y\sqrt{d}=(x_1+y_1\sqrt{d})^k$. Since $x\ge 1$ and $y\ge 1$, k is not 0, so $k\ge 1$.

What if x and y are not both positive? Then $\alpha := x+y\sqrt{d}$ is not in $(1,\infty)$ by Lemma 5.1. If $\alpha\ne\pm 1$ then α is in one of the intervals $(0,1)$, $(-1,0)$, and $(-\infty,-1)$, so (exactly) one of the numbers $1/\alpha$, $-1/\alpha$, or $-\alpha$ is in $(1,\infty)$. Each of these is a Pell solution too:

$$\frac{1}{\alpha}=\frac{1}{x+y\sqrt{d}}=x-y\sqrt{d},\quad \frac{-1}{\alpha}=-(x-y\sqrt{d})=-x+y\sqrt{d},\quad -\alpha=-x-y\sqrt{d}.$$

The number among these in $(1,\infty)$ has positive coefficients, so by our previous reasoning $\pm\alpha^{\pm 1}=(x_1+y_1\sqrt{d})^K$ for some $K\ge 1$ and some signs on the left side. Thus $\alpha=x+y\sqrt{d}=\pm(x_1+y_1\sqrt{d})^{\pm K}$. If $\alpha=\pm 1$ then it also arises in this way using $K=0$. $\square$

Remark 5.4. For solutions $x+y\sqrt{d}$ with $x,y\in\mathbf{Z}^+$, the one with minimal y is also the one with minimal x, since the coefficients of $(x_1+y_1\sqrt{d})^k$ both increase with k.

Example 5.5. The positive solution of $x^2-2y^2=1$ with least y-value is $(3,2)$, so every positive solution comes from coefficients of $(3+2\sqrt{2})^n$ for $n\ge 1$.

Example 5.6. The positive solution of $x^2-3y^2=1$ with least y-value is $(2,1)$, so the positive solutions are the coefficients of $(2+\sqrt{3})^k$ for $k\ge 1$.

Example 5.7. The positive solution of $x^2-5y^2=1$ with least y-value is $(9,4)$, so every positive solution comes from coefficients of $(9+4\sqrt{5})^k$ for $k\ge 1$.

Example 5.8. For integral $a\ge 2$, the positive solution of $x^2-(a^2-1)y^2=1$ with least y-value is $(a,1)$.[3] Indeed, from $y=1$ we get $x^2=a^2-1+1=a^2$, so $x=a$. Therefore the solutions of $x^2-(a^2-1)y^2=1$ in positive integers are the coefficients of $(a+\sqrt{a^2-1})^k$ for $k\ge 1$. The first few powers are

$$\begin{aligned}
(a+\sqrt{a^2-1})^1 &= a+\sqrt{a^2-1},\\
(a+\sqrt{a^2-1})^2 &= 2a^2-1+2a\sqrt{a^2-1},\\
(a+\sqrt{a^2-1})^3 &= (4a^3-3a)+(4a^2-1)\sqrt{a^2-1},\\
(a+\sqrt{a^2-1})^4 &= (8a^4-8a^2+1)+(8a^3-4a)\sqrt{a^2-1}.
\end{aligned}$$

[3] The number a^2-1 is not a perfect square, since otherwise a^2-1 and a^2 would be consecutive positive integers that are perfect squares and that is impossible since successive positive squares spread out and the least distance between them is $2^2-1^2=3$.

Setting $(a+\sqrt{a^2-1})^k = x_k(a) + y_k(a)\sqrt{a^2-1}$, both $x_k(a)$ and $y_k(a)$ are polynomials with integral coefficients. They are called Chebyshev polynomials of the first and second kind. They grow quickly: it can be shown that $a^k \le x_k(a) \le (2a)^k$ and $(2a-1)^{k-1} \le y_k(a) \le (2a)^{k-1}$.

Properties of $x_k(a)$ and $y_k(a)$ (in terms of growth and divisibility) have an application to mathematical logic, in the solution of Hilbert's Tenth Problem [1, Sect. 10.4], [2].

6. Generalized Pell Equations

The equation $x^2 - dy^2 = n$ where $n \in \mathbf{Z} - \{0\}$ is a *generalized Pell equation*. The special case $x^2 - dy^2 = -1$ is called a *negative Pell equation*. To find a solution (in $\mathbf{Z}$) rewrite the equation as $x^2 = dy^2 + n$ and compute the right side for $y = 1, 2, \dots$ until it's a square.

In the tables below we try to solve $x^2 - 2y^2 = -1$ and $x^2 - 3y^2 = -1$. For $1 \le y \le 15$ two solutions are found for the first equation, $(1,1)$ and $(7,5)$, and none for the second.

y	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$2y^2-1$	1	7	17	31	49	71	97	127	161	199	241	287	337	391	449
Square?	✓	X	X	X	✓	X	X	X	X	X	X	X	X	X	X

y	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$3y^2-1$	2	11	26	47	74	107	146	191	242	299	362	431	506	587	674
Square?	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

A longer search in the second case would be fruitless: $x^2 - 3y^2 = -1$ has no integral solutions. To prove some generalized Pell equation has no solution, an argument by contradiction using modular arithmetic often works. Here are a few examples.

Example 6.1. If integers x and y satisfy $x^2 - 3y^2 = -1$ then reduce both sides mod 3 to get $x^2 \equiv -1 \bmod 3$. This congruence has no solution: the only squares mod 3 are 0 and 1. Thus $x^2 - 3y^2 = -1$ has no solution (in $\mathbf{Z}$).

Example 6.2. The equation $x^2 - 5y^2 = 2$ has no integral solution because reducing the equation mod 5 makes it $x^2 \equiv 2 \bmod 5$, which has no solution. By the same idea, the generalized Pell equations $x^2 - 5y^2 = 3$ and $x^2 - 5y^2 = 7$ have no solutions.

Example 6.3. The equation $x^2 - 5y^2 = 6$ has no solution, but we can't prove this by reducing both sides mod 5 to get $x^2 \equiv 6 \bmod 5$, since that congruence *has* a solution so there is no contradiction. Reduce mod 3 instead: the equation becomes $x^2 - 5y^2 \equiv 0 \bmod 3$, or $x^2 \equiv 5y^2 \equiv 2y^2 \bmod 3$. This too has a solution, namely $(0,0)$, so it doesn't seem like progress has been made. But this is progress because $(0,0)$ is the *only* solution mod 3, since the squares mod 3 are 0 and 1, and the only way one of these is twice the other mod 3 is when they're both 0. Therefore if $x^2 - 5y^2 = 6$ in $\mathbf{Z}$ then x and y are both multiples of 3. That makes $x^2 - 5y^2$ a multiple of 9, which contradicts it being 6.

We used modular arithmetic here (reducing mod 3), but in a more subtle way than in the previous two examples.

Remark 6.4. The above three examples not only have no solutions in $\mathbf{Z}$, but in fact have no solutions in $\mathbf{Q}$. In contrast, $x^2 - 34y^2 = -1$ has no solution in $\mathbf{Z}$ but has solutions in $\mathbf{Q}$, such as $(5/3, 1/3)$, $(3/5, 1/5)$, $(27/11, 5/11)$, and $(3/29, 5/29)$.

Here is a problem about sums of squares whose solution is equivalent to solving a particular generalized Pell equation.

Theorem 6.5. *Finding positive integers a and b satisfying*

$$a^2 + (a+1)^2 = b^2 + (b+1)^2 + (b+2)^2 \tag{6.1}$$

is the same as solving $x^2 - 6y^2 = 3$ in positive integers x and y other than $(3, 1)$.

Proof. Expanding the squares and combining like terms,

$$\begin{aligned} a^2 + (a+1)^2 = b^2 + (b+1)^2 + (b+2)^2 &\iff 2a^2 + 2a + 1 = 3b^2 + 6b + 5 \\ &\iff 2(a^2 + a) = 3(b^2 + 2b) + 4 \\ &\iff 2\left(\left(a + \frac{1}{2}\right)^2 - \frac{1}{4}\right) = 3((b+1)^2 - 1) + 4 \\ &\iff ((2a+1)^2 - 1) = 6((b+1)^2 - 1) + 8 \\ &\iff (2a+1)^2 - 6(b+1)^2 = 3. \end{aligned}$$

All steps are reversible, so solving (6.1) for positive integers a and b is equivalent to solving $x^2 - 6y^2 = 3$ for odd $x \geq 3$ and arbitrary $y \geq 2$. Requiring x to be odd can be dropped since it is forced by the equation $x^2 - 6y^2 = 3$: the number $x^2 = 6y^2 + 3$ must be odd, so x must be odd. □

For a Pell equation $x^2 - dy^2 = 1$, multiplying two known solutions in the form $x + y\sqrt{d}$ leads to a third solution (Theorem 4.1). For a generalized Pell equation $x^2 - dy^2 = n$, multiplying a known solution *with a solution of the Pell equation* $x^2 - dy^2 = 1$ leads to a new solution of $x^2 - dy^2 = n$. The proof is like that of Theorem 4.1. Details are left to the reader.

Example 6.6. One solution of $x^2 - 6y^2 = 3$ is $(3, 1)$. A nontrivial solution of $x^2 - 6y^2 = 1$ is $(5, 2)$. Therefore a second solution of $x^2 - 6y^2 = 3$ comes from the coefficients of

$$(3 + \sqrt{6})(5 + 2\sqrt{6}) = 27 + 11\sqrt{6}.$$

Check $27^2 - 6 \cdot 11^2 = 3$. In the context of Theorem 6.5 the solution $(27, 11)$ of $x^2 - 6y^2 = 3$ leads to a solution of (6.1): $2a + 1 = 27 \Rightarrow a = 13$ and $b + 1 = 11 \Rightarrow b = 10$, so $13^2 + 14^2 = 10^2 + 11^2 + 12^2$. This is more attractive if we swap the two sides: $10^2 + 11^2 + 12^2 = 13^2 + 14^2$.

As an application of math to art, consider the painting in Figure 1 by Bogdanov-Belsky, which is in the Tetryakov Gallery in Moscow. In it children have to calculate

$$\frac{10^2 + 11^2 + 12^2 + 13^2 + 14^2}{365}$$

in their heads. If they knew $10^2 + 11^2 + 12^2 = 13^2 + 14^2$, they could find the numerator as $2(13^2 + 14^2)$, so only two squares would have to be computed instead of five. Since $2(13^2 + 14^2) = 2(169 + 196) = 2(365)$, dividing by 365 shows the fraction on the board is 2.

The bold artistic details on the blackboard and knowledge of the period when the painting was produced help us find the deeper meaning of this work of art: it is advocating for the inclusion of generalized Pell equations in the math curriculum of 19th century peasants.

FIGURE 1. Bogdanov-Belsky's *Mental Calculation* (1895).

7. THE NEGATIVE PELL EQUATION

We shift from a generalized Pell equation $x^2 - dy^2 = n$ to the special case of the negative Pell equation $x^2 - dy^2 = -1$. While a product of two solutions of $x^2 - dy^2 = 1$ is again a solution (Theorem 4.1), this doesn't quite work in the same way for the negative Pell equation.

Theorem 7.1. *If $x^2 - dy^2 = n$ and $x'^2 - dy'^2 = n'$ then the coefficients of $(x+y\sqrt{d})(x'+y'\sqrt{d})$ are a solution of $X^2 - dY^2 = nn'$.*

Proof. By the proof of Theorem 4.1, from $(x+y\sqrt{d})(x'+y'\sqrt{d}) = (xx'+dyy')+(xy'+yx')\sqrt{d}$ we get

$$(xx' + dyy')^2 - d(xy' + yx')^2 = (x^2 - dy^2)(x'^2 - dy'^2) = nn'. \qquad \square$$

Therefore if (x, y) is a solution of $X^2 - dY^2 = -1$, the coefficients of $(x + y\sqrt{d})^k$ for $k \in \mathbf{Z}^+$ are solutions of $X^2 - dY^2 = (-1)^k$: odd powers give solutions of $X^2 - dY^2 = -1$ and even powers give solutions to $X^2 - dY^2 = 1$.

Example 7.2. The equation $x^2 - 2y^2 = -1$ has solution $1+\sqrt{2}$, whose initial powers are

$$1+\sqrt{2}, \quad (1+\sqrt{2})^2 = 3+2\sqrt{2}, \quad (1+\sqrt{2})^3 = 7+5\sqrt{2},$$

$$(1+\sqrt{2})^4 = 17+12\sqrt{2}, \quad (1+\sqrt{2})^5 = 41+29\sqrt{2}, \quad (1+\sqrt{2})^6 = 99+70\sqrt{2}.$$

The coefficients of the 1st, 3rd, and 5th powers are solutions of $x^2-2y^2=-1$, while the coefficients of the 2nd, 4th, and 6th powers are solutions of $x^2-2y^2=1$.

The main point of this is that integral solutions of $x^2-dy^2=-1$ are not closed under multiplication, but integral solutions of the equations $x^2-dy^2=\pm 1$ together are. We will prove in Theorem 7.5 below an analogue for these two equations together of Theorem 5.3 for the Pell equation $x^2-dy^2=1$.

Suppose the negative Pell equation $X^2-dY^2=-1$ has a solution (x,y) in $\mathbf{Z}$. Then $x \neq 0$ and $y \neq 0$,[4] so by changing signs on x and y there is a solution in $\mathbf{Z}^+$. The Pell equation $X^2-dY^2=1$ then also has a solution in $\mathbf{Z}^+$ by using the coefficients of $(x+y\sqrt{d})^2$.

We will show the positive integer solutions of $X^2-dY^2=-1$, when they exist, can be ordered by either the values of x or y, using the following analogues of Lemmas 5.1 and 5.2.

Lemma 7.3. *If $x^2-dy^2=-1$ in $\mathbf{Z}$ and $x+y\sqrt{d}>1$ then $x \geq 1$ and $y \geq 1$.*

The lower bounds on x and y can be achieved: $x^2-2y^2=-1$ for $x=1$ and $y=1$.

Proof. We have $1/(x+y\sqrt{d}) = -(x-y\sqrt{d}) = -x+y\sqrt{d}$, so

$$x+y\sqrt{d} > 1 > -x+y\sqrt{d} > 0.$$

From $x+y\sqrt{d} > -x+y\sqrt{d}$ we get $2x>0$, so $x>0$. Then $x \geq 1$ since x is an integer, so $y\sqrt{d} > x > 0$. Thus $y>0$, so $y \geq 1$. □

Lemma 7.4. *Suppose $x^2-dy^2=-1$ and $a^2-db^2=-1$ in $\mathbf{Z}$ where $x,y,a,b \geq 1$. Then*

$$a+b\sqrt{d} < x+y\sqrt{d} \iff a<x \text{ and } b<y \iff a<x \text{ or } b<y.$$

Proof. The first ($\Leftarrow$) is obvious. To prove the first ($\Rightarrow$), from $x+y\sqrt{d} > a+b\sqrt{d} \geq 1+\sqrt{d} > 1$ reciprocating implies

$$a+b\sqrt{d} < x+y\sqrt{d},$$

so we get

$$-x+y\sqrt{d} < -a+b\sqrt{d}.$$

Adding these inequalities gives us

$$(a-x)+(b+y)\sqrt{d} < (x-a)+(y+b)\sqrt{d}.$$

Subtracting $(b+y)\sqrt{d}$ from both sides, we get $a-x<x-a$, so $2a<2x$ and thus $a<x$. Then $db^2 = 1+a^2 < 1+x^2 = dy^2$, so $b<y$ from positivity of b and y.

That $a<x$ and $b<y$ is equivalent to $a<x$ or $b<y$ follows by the same reasoning as at the end of the proof of Lemma 5.2. □

By Lemma 7.4, the positive integer solutions to $x^2-dy^2=-1$ (when such solutions exist) can be ordered by their first coordinate or by their second coordinate. So it makes sense to speak of the solution (x_1,y_1) in positive integers of $x^2-dy^2=-1$ (when it exists) with minimal value of y_1 (or x_1). That is a basic ingredient in the next result, which is analogous to Theorem 5.3.

[4]If $d=1$ then the negative Pell equation has the solution $(x,y)=(0,1)$, but we always assume d is not a square, so in particular $d \neq 1$.

Theorem 7.5. *Assume $x^2 - dy^2 = -1$ has a solution in positive integers and let (x_1, y_1) be that solution where y_1 is minimal. The integral solutions to both equations $x^2 - dy^2 = \pm 1$ are, up to sign, generated from (x_1, y_1) by taking integral powers of $x_1 + y_1\sqrt{d}$:*

$$x + y\sqrt{d} = \pm(x_1 + y_1\sqrt{d})^k$$

for some sign and some $k \in \mathbf{Z}$. The integral solutions of $x^2 - dy^2 = -1$ have k odd and the integral solutions of $x^2 - dy^2 = 1$ have k even.

Proof. For $k \in \mathbf{Z}$, the coefficients of $(x_1 + y_1\sqrt{d})^k$ satisfy $x^2 - dy^2 = (-1)^k$, and this is also true for coefficients of $-(x_1 + y_1\sqrt{d})^k$.

To show all integral solutions of $x^2 - dy^2 = \pm 1$ are related to powers of $x_1 + y_1\sqrt{d}$, we will reduce ourselves to what was shown in Theorem 5.3 for solutions of $x^2 - dy^2 = 1$ in $\mathbf{Z}^+$.

The Pell equation $x^2 - dy^2 = 1$ has a solution in positive integers, such as the coefficients of $(x_1 + y_1\sqrt{d})^2$. Let (X_1, Y_1) be the solution of $x^2 - dy^2 = 1$ in positive integers with minimal Y_1. We will show $X_1 + Y_1\sqrt{d} = (x_1 + y_1\sqrt{d})^2$.

By Theorem 5.3 and minimality of (X_1, Y_1), $(x_1+y_1\sqrt{d})^2 = (X_1+Y_1\sqrt{d})^k$ for some $k \geq 1$. If $k = 2\ell$ is even then $(x_1 + y_1\sqrt{d})^2 = (X_1 + Y_1\sqrt{d})^{2\ell}$, so $x_1 + y_1\sqrt{d} = \pm(X_1 + Y_1\sqrt{d})^\ell$. However, since $X_1^2 - dY_1^2 = 1$, the coefficients of 1 and $\sqrt{d}$ in $\pm(X_1+Y_1\sqrt{d})^\ell$ fit the equation $x^2 - dy^2 = 1$ while $x_1^2 - dy_1^2 = -1$. This is a contradiction, so k is odd: $k = 2\ell + 1$ where $\ell \geq 0$. Then

$$\begin{aligned}(x_1 + y_1\sqrt{d})^2 = (X_1 + Y_1\sqrt{d})^{2\ell+1} \Longrightarrow X_1 + Y_1\sqrt{d} &= (x_1 + y_1\sqrt{d})^2(X_1 + Y_1\sqrt{d})^{-2\ell} \\ &= ((x_1 + y_1\sqrt{d})(X_1 - Y_1\sqrt{d})^\ell)^2.\end{aligned}$$

Set $(x_1 + y_1\sqrt{d})(X_1 - Y_1\sqrt{d})^\ell = a + b\sqrt{d}$, so $a^2 - db^2 = -1$ and

$$X_1 + Y_1\sqrt{d} = (a + b\sqrt{d})^2. \tag{7.1}$$

That implies $2ab = Y_1 > 0$, so a and b are nonzero with the same sign. By replacing a with $-a$ and b with $-b$ in case a and b are negative, (7.1) holds for some $a, b \in \mathbf{Z}^+$.

If $\ell > 0$ then

$$(x_1 + y_1\sqrt{d})^2 = (X_1 + Y_1\sqrt{d})^{2\ell+1} > X_1 + Y_1\sqrt{d} = (a + b\sqrt{d})^2 > 1,$$

so $x_1 + y_1\sqrt{d} > a + b\sqrt{d}$ by taking positive square roots. This implies $y_1 > b$ by Lemma 7.4, which contradicts the minimality of y_1 among solutions to $x^2 - dy^2 = -1$ in $\mathbf{Z}^+$. Thus $\ell = 0$, so

$$(x_1 + y_1\sqrt{d})^2 = (X_1 + Y_1\sqrt{d})^{2\ell+1} = X_1 + Y_1\sqrt{d}.$$

Now we are ready to relate the solutions of $x^2 - dy^2 = \pm 1$ in positive integers to integral powers of $x_1 + y_1\sqrt{d}$. Afterwards we'll remove the condition that $x > 0$ and $y > 0$.

<u>Case 1</u>: if $x^2 - dy^2 = -1$ for $x, y \in \mathbf{Z}^+$ then $(x+y\sqrt{d})^2$ has coefficients that are a solution of the Pell equation (right side being 1), so by Theorem 5.3

$$(x + y\sqrt{d})^2 = (X_1 + Y_1\sqrt{d})^k = (x_1 + y_1\sqrt{d})^{2k}$$

for some $k \in \mathbf{Z}^+$. Taking positive square roots of both sides,

$$x + y\sqrt{d} = (x_1 + y_1\sqrt{d})^k.$$

Since $x^2 - dy^2 = -1$ and $x_1^2 - dy_1^2 = -1$, we have $-1 = (-1)^k$ by Theorem 7.1, so k is odd: $x + y\sqrt{d}$ is an odd power of $x_1 + y_1\sqrt{d}$.

Case 2: if $x^2 - dy^2 = 1$ for $x, y \in \mathbf{Z}^+$ then by Theorem 5.3

$$x + y\sqrt{d} = (X_1 + Y_1\sqrt{d})^k = (x_1 + y_1\sqrt{d})^{2k}$$

for some $k \in \mathbf{Z}^+$, so $x + y\sqrt{d}$ is an even power of $x_1 + y_1\sqrt{d}$.

Combining Cases 1 and 2, the solutions to $x^2 - dy^2 = \pm 1$ in positive integers are the coefficients of $(x_1 + y_1\sqrt{d})^K$ for $K \in \mathbf{Z}^+$.

Now suppose $x^2 - dy^2 = \varepsilon$ where $\varepsilon = \pm 1$ and the integers x and y are not both positive. By Lemmas 5.1 and 7.3, $\alpha := x + y\sqrt{d}$ is not in $(1, \infty)$. If $\alpha \neq \pm 1$ then α is in the interval $(0, 1)$, $(-1, 0)$, or $(-\infty, -1)$, which makes (exactly) one of the numbers $1/\alpha$, $-1/\alpha$, or $-\alpha$ lie in $(1, \infty)$. Each of these has coefficients that equal x and y up to sign, and therefore satisfy $X^2 - dY^2 = \varepsilon$:

$$\frac{1}{\alpha} = \varepsilon(x - y\sqrt{d}), \quad \frac{-1}{\alpha} = -\varepsilon(x - y\sqrt{d}) = \varepsilon(-x + y\sqrt{d}), \quad -\alpha = -x - y\sqrt{d}.$$

The number among these that lies in $(1, \infty)$ has positive coefficients (Lemmas 5.1 and 7.3 again), so by our previous reasoning $\pm\alpha^{\pm 1} = (x_1 + y_1\sqrt{d})^K$ for some $K \geq 1$ and some signs on the left side. Thus $\alpha = x + y\sqrt{d} = \pm(x_1 + y_1\sqrt{d})^{\pm K}$ and $\pm K \neq 0$. If $\alpha = \pm 1$ then $\alpha = \pm(x_1 + y_1\sqrt{d})^0$. $\square$

Example 7.6. The solution of $x^2 - 5y^2 = -1$ in $\mathbf{Z}^+$ with least y is $(2, 1)$, so the solutions in positive integers are the coefficients of $(2 + \sqrt{5})^k$ for odd $k \geq 1$. The initial odd powers after $(2 + \sqrt{(}5))^1$ are

$$(2 + \sqrt{5})^3 = 38 + 17\sqrt{5}, \quad (2 + \sqrt{5})^5 = 682 + 305\sqrt{5}, \quad (2 + \sqrt{5})^7 = 12238 + 5473\sqrt{5}.$$

Example 7.7. For integral $a \geq 1$, the positive solution of $x^2 - (a^2 + 1)y^2 = -1$ with least y is $(a, 1)$.[5] Therefore the solutions of $x^2 - (a^2 + 1)y^2 = \pm 1$ in positive integers are the coefficients of $(a + \sqrt{a^2 + 1})^k$ for $k \geq 1$. Here are the first few powers:

$$\begin{aligned}
(a + \sqrt{a^2 + 1})^1 &= a + \sqrt{a^2 + 1}, \\
(a + \sqrt{a^2 + 1})^2 &= 2a^2 + 1 + 2a\sqrt{a^2 + 1}, \\
(a + \sqrt{a^2 + 1})^3 &= (4a^3 + 3a) + (4a^2 + 1)\sqrt{a^2 + 1}, \\
(a + \sqrt{a^2 + 1})^4 &= (8a^4 + 8a^2 + 1) + (8a^3 + 4a)\sqrt{a^2 + 1}.
\end{aligned}$$

The coefficients here are similar to those in powers of $a + \sqrt{a^2 - 1}$ in Example 5.8, except the powers of a on the right side don't have negative coefficients. When $a = 1$, this is Example 7.2.

References

[1] R. E. Hodel, An Introduction to Mathematical Logic, PWS Publishing Co., New York, 1995.

[2] J. P. Jones, Y. V. Matiyasevich, "Proof of Recursive Unsolvability of Hilbert's Tenth Problem," *Amer. Math. Monthly* **98** (1991), 689–709.

[5] The number $a^2 + 1$ is not a perfect square, since otherwise a^2 and $a^2 + 1$ would be consecutive positive integers that are perfect squares and that's impossible.