The cakravāla method

S. Raghavan

The object of this article spurred by a book review in Current Science (1996, 70, 753–754) is to try to place in proper perspective the well-known work of Indian mathematicians especially their kuttaka, bhāvānā and cakravāla methods evolved in connection with the solution in integers of certain indeterminate equations of degree one or two. In this effort, we base ourselves completely on André Weil’s masterly, unbiased and incisive analysis of this topic in his beautiful book Number Theory – An Approach Through History – From Hammurapi to Legendre. Any help needed to understand related results concerning continued fractions can be readily secured from the book An Introduction to the Theory of Numbers (especially chapter 7), although we endeavour to skirt a reference thereto by the reader by providing as self-contained an account as possible of facts related to simple continued fractions (in the sequel).

Solving indeterminate equations of the first degree, say $ax + by + cz + \cdots = m$ in the variables $x, y, z, \ldots$ with integral coefficients $a, b, c, \ldots$ for integer values of the variables has been the key to cracking puzzles or finding integral solutions of simultaneous simple linear congruences, e.g. $x \equiv k (\text{mod } r)$, $y \equiv l (\text{mod } s)$ requiring $x - k, y - l$ to be divisible by integers $r, s$ respectively for given integers $k, l$. A single linear congruence $px - qy = m$ in the variable $x$ for the modulus $q$ is equivalent to a linear equation $px - qy = m$ in two variables $x, y$. Euclid’s algorithm for finding the greatest common divisor of the integers $p$ and $q$ provides a method of solving the last-mentioned equation; expanding the rational number $p/q$ (for $q \neq 0$) as a simple continued fraction gives an alternative (but equivalent) approach to the same problem. A clear description of the general solution of this linear equation in $x, y$ can be found in the Sanskrit text Aryabhatiya of the 5th or 6th century A.D. In subsequent Sanskrit treatises, this method came to be known as the ‘kuttaka’ (=‘pulveriser’) and is indeed a kind of forerunner of Fermat’s powerful principle of ‘infinite descent’. It is also the ‘first ever explicit description’ of the general solution from anywhere, not taking into account China. Indian astronomy at that time was under the influence of Greek sources and yet perhaps one cannot with certainty attribute ‘kuttaka’ to Greek mathematics. In utter disregard or (possible) ignorance of this Indian dimension to the ‘kuttaka’ and of the connection with the seventh book of Euclid’s Elements, Bachet inserted a strong claim to the method as his own, in the second edition of his book on Problèmes plaisants et délectables.

Let $N$ be a natural number which is not the square of an integer; the simplest example is $N = 2$. In view of the connection with finding close approximations to the irrational number $\sqrt{2}$ by rational numbers, indeterminate equations of degree 2 like $x^2 - Ny^2 = \pm m$ for given integers $m$ must have been indeed investigated by the Greek mathematicians. Special cases of the ‘composition formulae’

$$ (x^2 - Ny^2)(x'^2 - N'y'^2) = (x \pm Nyt)^2 - N(xt \pm yz)^2, \quad (1) $$

e.g. for $N = 3$, $z = 5$, $t = 3$ may have been applied by Archimedes for finding rational approximations to $\sqrt{3}$. However, the composition formula (1) occurs explicitly in the work of Brahmagupta (in the 7th century) while seeking the solution in integers $x, y$ of equations of the form $x^2 - Ny^2 = \pm m$ for any fixed natural number $m$. For $m = 1$, the name Pell’s equation has come to stay for the diophantine equation

$$ x^2 - Ny^2 = 1. \quad (2) $$

Such equations ‘do occur in Diophantus..., but it is a rational solution that is asked for, even when accidentally a solution in integers is obtained...’ It may be reasonable to suppose that Archimedes was interested in equations of this type or, to be very optimistic, that he had even found a general method of solving equations like (2).

In the book entitled Algebra with Arithmetic and Mensuration, from the Sanskrit of Brahmagupta and Bhāscara by H. T. Colebrooke, one finds an entire section dealing with Brahmagupta’s investigations (in the seventh century) on Vargaprakṛti – the solution of equations $Ny^2 + m = x^2$ in integers $x, y$ with $N$ as above and $m$, a non-zero integer; $N$ is called ‘gunaka’ (or ‘prakṛti’) and $m$ the ksepa (=additive). If the triple $(x, y; m)$ stands for a solution in integers $x, y$ of the equation $x^2 - Ny^2 = m$, keeping $N$ fixed all the while, the composition formula (1) is given by Brahmagupta in the form

$$ (x, y; m) \cdot (x, y; n) \rightarrow (xz \pm Nyt, xt \pm yz; mn). $$

These laws acquired, in the post-Brahmagupta manuscripts in India, the name bhāvānti (=‘production’ rules); in modern parlance, this is just the ‘multiplicativity of the norm’. Brahmagupta shows, how composition of $(x, y; m)$ with a triple $(p, q; 1)$ gives a triple $(x', y'; m)$ for the same additive $m$. Under composition with itself, any triple $(x, y; m)$ yields a solution in rational numbers $x'lm, y'lm$ of the equation $(x'lm)^2 - N(y'lm)^2 = 1$ and indeed a triple $(x'lm, y'lm; 1)$ if $x'lm$ and $y'lm$ are actually integers. Applying then his bhāvānti, Brahmagupta solves equation (2) for several cases of $N$ including ones like $N = 92$ or $N = 83$. For $m = -1$ or $\pm 2$, composition of a triple $(p, q; m)$ with itself leads to $(p^2 + Nq^2, 2pq; 1)$ or $(p^2 + Nq^2)/2, pq; 1)$ respectively.

Despite the remarkable results of Brahmagupta, the general solution of (2) is still not at hand. Actually, the cakravāla (‘cyclic’ method) for getting the general solution of equation (2) is to be found much later, around the twelfth century, in the work of Bhāskara; nearly the same description of the cakravāla is provided in a commentary of the eleventh century by ‘an otherwise unknown author’ Jayadeva, leaving one to guess who was the true inventor of the cakravāla. Following Weil, we can see how the brilliant cakravāla arises in a
natural manner from the work of earlier Indian mathematicians.

Starting from a triple \((p_0, q_0, m_0)\) with 'small' \(m_0\), the idea is to use the bhāvānā to get a triple \((p_1, q_1; m_1)\) with \(m_1\) also 'small' and eventually hope to hit upon a triple \((u, v; 1)\) giving a non-trivial solution of equation (2), of course. First we can assume, without loss of generality that the greatest common divisor \(d\) of \(p_0\) and \(q_0\) in the initial triple is already equal to 1, since if \(d > 1\), we could start instead from \((p_0/d, q_0/d; m_0d^2)\). Since \(p_0\) and \(q_0\) have greatest common divisor 1, so have \(m_0\) and \(m_0\) clearly. Then the kuttaka readily enables us to find an integer \(x_0\) such that \(m_0 \mid p_0 + q_0x_0\) (i.e. solve the arithmetical congruence \(q_0x_0 \equiv -p_0 \pmod{m_0}\)). If, in addition, the chosen \(x_0\) is fixed in its residue class modulo \(m_0\), so as to satisfy the inequalities \(x_0 \leq \sqrt{N} \leq x_0 + lm_0\), then we see that \(\sqrt{N} + x_0 \neq 0\) cannot be negative provided that \(lm_0 < 2\sqrt{N}\) (since \(\sqrt{N} + x_0 < 0\) would imply that \(2\sqrt{N} < \sqrt{N} + x_0 + lm_0 < 0\) and consequently for \(lm_0 < 2\sqrt{N}\),

\[
0 < N - x_0^2 = (\sqrt{N} - x_0)(\sqrt{N} + x_0)
\]

\[
< lm_0(\sqrt{N} + x_0) = 2lm_0 < 2\sqrt{N} \quad . \quad (3)
\]

Composition of \((p_0, q_0; m_0)\) with \((x_0, 1; x_0^2 - N)\) gives rise to the triple \((p_1, q_1; m_1)\)

\[
p_1 := (p_0x_0 + Nq_0/\gcd(m_0, q_0)) \quad ,
\]

\[
q_1 := (p_0 + q_0x_0) \quad ,
\]

\[
m_1 := (x_0^2 - N) \quad ,
\]

(4)

are clearly integers. (In fact, by composition, \((m_0q_0)^2 - N(m_0q_0)^2 = m_0(x_0^2 - N) = m_0m_1\quad \text{and} \quad q_0^2 (x_0^2 - N) = \frac{q_0^2 - N}{p_0^2 + m_0} = m_0((q_0x_0 - p_0)/(q_0x_0 + p_0)/m_0) + 1)\) is divisible by \(m_0\) i.e. \(x_0^2 - N\) is divisible by \(m_0\) in view of the greatest common divisor \(m_0\) and \(q_0\) being 1 by our assumption above. Thus \(m_1\) is an integer while the congruence condition on \(x_0\) implies that \(q_1\) is an integer as well and so \(p_1^2 = Nq_1^2 + 1\) and \(p_1\) too as a consequence! Moreover, \(\gcd(m_1q_1) = \gcd(N, x_0^2 - 2lm_0\sqrt{N})\), by (3) and therefore we have

\[
\|m_1\| < 2\sqrt{N} \quad . \quad (5)
\]

Starting with the triple \((p_0, q_0; m_0)\), the passage to \((p_1, q_1; m_1)\) as above gives an inductive construction of the triples \((p_n, q_n; m_n)\) as follows.

It is convenient to take \(q_0 = 1\) and \(P_0 := \lfloor \sqrt{N} \rfloor\), the largest integer not exceeding the (positive) square root \(\lfloor \sqrt{N} \rfloor\) of \(N\), so that \(0 < \sqrt{N} - p_0 < 1\). The congruence condition on \(x_0\) now looks simpler, viz. \(x_0 \equiv -p_0 \pmod{m_0}\) with \(m_0 := p_0^2 - N \geq 0\), i.e. \(x_0\) is any integer such that \(x_0 + p_0\) is divisible by \(m_0\) (but subject to the additional conditions \(x_0 < \sqrt{N} < x_0 + lm_0\) of course!). Due to the special choice of \(q_0\), the kuttaka does not need to be invoked here (for solving a congruence for \(x_0\)) but also at every subsequent step under the induction, as nicely emphasized by Weil. We reproduce his comments in this regard verbatim.‘Strangely enough, this does not seem to have been noticed by any of our Indian authors (nor even by their later commentators, down to the sixteenth century); they make no mention of it, and invariably refer to the kuttaka for the choice of \(x_0\), even though their abundant numerical evidence could easily have convinced them that this was unnecessary.’

Let us assume the triples \((p_n, q_n; m_n)\), \(s_j\) for \(0 \leq j \leq i\) constructed inductively with \(\|m_j\| < 2\sqrt{N}\), \(s_j = s_{j-1} \pmod{m_j}\) i.e. \(s_j + s_{j-1}\) divisible by \(m_j\), \(s_j < \sqrt{N} < s_{j-1}\) + \(\|m_j\|\). \(m_j := 1\), \(x_{j-1} := p_0\) and \(x_{j+1} - N = m_jm_{j+1}\). Then composition of \((p_n, q_n; m_n)\) with the triple \((x_0, 1; x_0^2 - N)\) leads to the definition of \(p_{i+1}, q_{i+1}, m_{i+1}\), viz., \(p_{i+1} := (p_0x_i + Nq_i) \quad , \quad q_{i+1} := (q_0x_i + p_0)m_i\) and \(m_{i+1} := (x_0^2 - N)/m_i\). The congruence condition on \(s_j\) above coupled with the relation \(-q_{i+1}s_j + p_i = -s_{i+1}q_j + q_{i+1}s_{i+1}p_0/m_{i+1} + (p_iq_{i+1} + q_0s_{i+1})q_{i+1}/m_{i+1} = q_{i+1}\) for \(N = x_0^2\) \(m_{i+1} = -q_{i+1}m_i\) ensures that \(q_{i+1}\) is an integer. Since \(m_{i+1}\) is an integer by the same congruence condition on \(s_{i+1}\), \(s_{i+1}\) is an integer too. The same kind of argument applied to derive the bound \((5)\) leads to \(\|m_{i+1}\| < 2\sqrt{N}\) and further ensures that \((\neg m_{i+1}, \sqrt{N} \pm x)\) are all positive (and so \(N - x^2 > 0\) for all \(i\)). We have thus, on hand, an infinite sequence of integers (the ‘additives’) \(m_0, m_1, m_2, \ldots\) bounded by \(2\sqrt{N}\) in absolute value. Hence infinitely many among them must coincide by Dirichlet’s box principle. But, as we will see presently, more is true, namely, (i) there exists an integer \(s\) such that \(m_{j+1} = m_j\) for \(j \geq 1\) and (ii) \(m_j = 1\) for an integer \(j\), leading to a solution of equation (2).

In other words, the \(m_j\) repeat themselves in a periodic fashion (actually, corresponding to the periodicity of the infinite simple continued fraction expansion for the quadratic irrationality \(\sqrt{N}\)).

Before moving on to indicate proofs for assertions (i) and (ii) above, it will be quite in order to quote some interesting observations by Weil (see ref. 1 pp. 23, 24, 94–97, 230–232) on the construction of \((p_n, q_n; m_n)\). ‘The Indian prescription’ for the choice of \(x_0\) within its congruence class modulo \(m_0\) is not quite the one described above, since their rule is ‘to make \(N - x^2\) “small”’ (i.e. in actual practice, as small as possible), but as the context shows, in absolute value’ or in other words, to replace \(x_j\) by \(y_j := x_j + lm_j\) if \(y_j - N \leq \sqrt{N}\) turns out to be less than \(N - x_j^2\). ‘It can be shown that this has merely the effect of abbreviating the procedure somewhat when that is the case,’ but though ‘numerically useful’, can ‘make the theoretical discussion much more cumbersome.’ Moreover, the above rigorous treatment for constructing \((p_n, q_n; m_n)\) ‘may have been known to the Indians only experimentally; there is nothing to indicate whether they had proofs for them, or even for part of them’. ‘In order to carry out the cakravala’, a ‘starting point’ which ‘invariably they choose’ is the triple \((p_0; 1; m_0)\) of which \(p_0^2\) is the closest square to \(N\), above or below.’ Finally we are told to iterate the process only till we find an “additive” \(m_0\) with one of the values ±1, ±2, ±4 and then to make use of the bhāvānā, i.e. Brahmagupta’s procedure for that case. Actually this is no more than a shortcut, since it can be shown that the cakravala applied in a straightforward manner, would inevitably lead to a triple \((p, q; 1)\) as desired; while this shortcut is quite effective from the point of view of the numerical solution, it destroys the ‘cyclic’ character of the method, which otherwise would appear from the fact that the additives … would repeat themselves periodically corresponding to the periodicity of the continued fraction of \(\sqrt{N}\).

‘For the Indians, of course, the effectiveness of “cakravala” could be no more than an experimental fact, based on their treatment of a great many special cases, some of them of considerable..."
HISTORICAL NOTES

complexity and involving (to their delight, no doubt) quite large numbers. Fermat was the first to perceive the need for a general proof and Lagrange was the first to publish one. Nevertheless, to have developed the calculus and to have applied it successfully to such difficult cases as $N = 61$ or $N = 67$ had been no mean achievement.

We now go on to the promised proofs for assertions (i) and (ii) above. The triples $(y, q; m)$ and $x_i$ for $i \geq 0$ as constructed above enable us to obtain an infinite simple continued fraction for \( \sqrt{N} \). Let $\xi_0 := \sqrt{N} + p_0$, $a_0 := 2p_0$, $a_i := (1)^i(x_{i+1} + x_{i-1})/m_{i-1}$ for $i \geq 1$ and $\eta_i$ be defined by $\sqrt{N} = x_i + \eta_i$ for $j \geq 0$.

Then, in view of the relation $x_{j+1}^2 - N = m_j m_{j+1}$ for $j \geq 0$, we can derive the following:

$$
\xi_0 = a_0 + \frac{1}{\xi_1}
$$

with

$$
\xi_1 = \frac{\sqrt{N} + p_0}{\sqrt{N} - p_0} = -m_0
$$

and so on for $\xi_0 = \sqrt{N} + [\sqrt{N}]$ we can even show that $l = 0$, i.e.

$$
\xi_0 = (a_0, a_1, a_2, \ldots)
$$

$\xi_0$ is periodic and $\xi_1 = (a_1, a_2, \ldots)$ so $\xi_1 \equiv \xi_0$.

Now $a_i \geq 1$ for every $i$, $(-1)^{m_i-1}$, $\sqrt{N} - x_{i-2}$ are all positive (by construction) and so, $\xi_0 < 0$, leading to $1/\xi_{i+1} = a_i - 1 < 0$, i.e. $-1 < \xi_{i+1} < 0$ for every $i \geq 1$. Then, using (10), we have $0 < (1/\xi_{i+1}) - a_i < 1$ so that $a_i$ is just the largest integer $[-1/\xi_{i+1}]$ not exceeding the positive real number $-1/\xi_{i+1}$. Since, for $k > 1$, $\xi_k = \xi_1$ by (8), the conjugates $\xi_k, \xi_k, \ldots$ coincide and so $a_{k+1} = a_{k-1}$. Thus from $\xi_k = \xi_1$ we conclude that $\xi_{k+1} = a_{k-1} + 1/\xi_{k+1} = a_{k-1} + 1/\xi_{k-1} = \xi_{k-1}$. Iteration yields $\xi_k = \xi_1 = \xi_0$ proving (9).

Now for any $j \geq 1$, $x_j^2$ has the same continued fraction expansion

$$
(\xi_0, a_1, a_2, \ldots)
$$

with the natural numbers $a_0, a_1, a_2, \ldots$ occurring as partial quotients.

There exists, as we shall see now, a minimal positive integer $r$ such that $a_{i+r} = a_i$ for all $i$ from a certain index $l$, say. Now $\xi_l = (\sqrt{N} + x_{l-2})/(-1)^{m_{l-1}}$ as derived inductively in (7) and coupled with the bound $\ln m \leq 2\sqrt{N}$ for all $i \geq 0$, the relation $x_i^2 = N + m_i m_{i+1}$ leads to the bound $l x_l^2 < 5N$ for all $i \geq 0$. The number of distinct pairs $(x_{i+1}, (-1)^{m_{i+1}})$ of integers for $i = 1, 2, 3, \ldots, j$ subject to such fixed bounds (depending only on the given non-square natural number $N$) can only be finite. Thus there exist indices $l$ and $k > l$ such that

$$
(a_0, a_1, a_2, \ldots) = (a_0, a_1, a_2, \ldots) = (a_0, a_1, a_2, \ldots)
$$

and therefore

$$
(a_0, a_1, a_2, \ldots) = (a_0, a_1, a_2, \ldots)
$$

is a periodic continued fraction with period $r = k - l$, proving that $a_{i+r} = a_i$ for all $i \geq l$. Clearly $r$ can be chosen to be minimal.

In the case $\xi_0 = \sqrt{N} + [\sqrt{N}]$, we can even show that $l = 0$, i.e.

$$
\xi_0 = (a_0, a_1, a_2, \ldots)
$$

and so is 'purely periodic'. For this purpose, we note first that while $\xi_0 = p_0 + \sqrt{N}$, its 'conjugate' $\xi_0 := p_0 - \sqrt{N}$ satisfies the inequalities $1 < \xi_0 < 0$. Denoting for $\xi_1 = (\sqrt{N} + x_{i+2})/(-1)^{m_{i+1}}$ its 'conjugate' $(-\sqrt{N} + x_{i+2})/(-1)^{m_{i+1}}$ by $\xi_1$, we have

$$
\xi_1 = a_1 + 1/\xi_{i+1}, \quad \xi_1 = a_1 + 1/\xi_{i+1}.
$$

Now $a_1 \geq 1$ for every $i$, $(-1)^{m_{i-1}}$, $\sqrt{N} - x_{i-2}$ are all positive (by construction) and so, $\xi_1 < 0$, leading to $1/\xi_{i+1} = a_1 - 1 < 0$. Thus, using (10), we have $0 < (1/\xi_{i+1}) - a_1 < 1$ so that $a_i$ is just the largest integer $[-1/\xi_{i+1}]$ not exceeding the positive real number $-1/\xi_{i+1}$. Since, for $k > 1$, $\xi_k = \xi_1$ by (8), the conjugates $\xi_k, \xi_k, \ldots$ coincide and so $a_{k+1} = a_{k-1}$. Thus from $\xi_k = \xi_1$ we conclude that $\xi_k = a_{k-1} + 1/\xi_k = a_{k-1} + 1/\xi_{k-1} = \xi_{k-1}$. Iteration yields $\xi_k = \xi_{k-1} = \xi_1$ proving (9).

Now for any $j \geq 1$, $\xi_j$ has the same continued fraction expansion

$$
(\xi_0, a_1, a_2, \ldots) = (\xi_0, a_1, a_2, \ldots)
$$

as $\xi_0$. Hence

$$
\sqrt{N} + x_{j-2} = \xi_j = \xi_0 = \sqrt{N} + p_0.
$$

When $r = 1$ is even (e.g., $r = 2$ for odd $r$ and $j = 1$ for even integers $r$), we have $m_{j+1} = 1$ and the triple $(p_{j+1}, q_{j+1}, 1)$ gives a solution of the diophantine equation (2) proving assertion (ii). Assertion (i) also follows from above easily.

An example of how unspanning a critical analysis can tend to be (when intended or called for) may be found in Weil's remarks on Euler's contribution to the topic of 'Pell's equation' and the related continued fraction algorithm, on pages 232–233 (ref. 1). 'While Euler drew attention' to the periodicity and 'palindromic' proper of the partial quotients $a_i$ in the 'continued fractions for square roots $\sqrt{N}$, as well as to their use in solving Pell's equation, there is no sign that he (Euler) ever sought to back up his findings by anything more than experimental evidence. He (Euler) did mention that the values obtained by his process for the integers $B, A, m_i$ are necessarily bounded...; from this he (Euler) could at least have derived the conclusion that the sequence $(m_i)$ is periodic from a certain point onwards, but he failed to mention this, or did not bother to do so'. When in his later years (after Lagrange gave a 'definitive treatment' of the subject, based on the continued fraction algorithm...) Euler came back to the topic of Pell's equation, he added nothing of substance to what by that time was already public knowledge on that subject.

Fermat must have been in the dark about the contribution of the Indian mathematicians to the solution of (2) and also possibly about Archimedes' Problema bovinum. He offered (in 1657) the problem of solving equation (2) (in integers, of course!) as a challenge to the English mathematicians and all others. In a personal letter to Huygens, a few months later, commenting on the solution by Wallis and Brouncker, he observed that 'the English had failed to give a general proof'; such a ('general') proof, according to Fermat, could only be 'obtained by descent'. But perhaps 'Fermat's method of solution (for (2)) did not greatly differ.
Addendum

Fifty years of the exact solution of the two-dimensional Ising model by Onsager

Somendra M. Bhattacharjee and Avinash Khare


We would like to make the following corrections/additions which we learnt after the paper was written.

W. Lenz was working in Rostock University (not Rostalk as mentioned in the text).

E. Ising was a teacher (and later became the headmaster) of a Jewish school in Germany from 1934 to November 1938. When his school was damaged, he managed to leave Germany. It was as late as 1949 that he realized that his name had become famous.

It seems that Ising also agrees that the model should be named the Lenz–Ising model. The name 'Ising model' became popular following the title of Peierls' paper.

It appears that Heisenberg in his 1928 paper also thought that at least eight nearest neighbours are needed for a phase transition. We are not sure whether this comment by Heisenberg refers to the Ising model or to the new model he proposed in that paper.

We thank Prof. Sigismund Kobe of University of Dresden for several clarification.