Dilip Das
18.704
Rogalski

Hasse's Theorem and Rational Points on the General Conic

Diophantine Equations are equations to which only integer solutions or alternatively rational solutions are considered. The study of these equations has fascinated man for thousands of years. The ancient Babylonians enumerated Pythagorean triples, integer solutions to the equation:

$$X^2 + Y^2 = Z^2$$

The study of Diophantine equations continued through Greece and the Renaissance with Diophantos, Fermat, Gauss and many others, and up to the present day in which Wiles' very recent proof of Fermat's last theorem still astonishes mathematicians.

The 'natural' first case in which the problem of finding solutions to Diophantine equations becomes nontrivial is the conic. Curves are classified by their genus, a certain very useful topological invariant. The Diophantine theory of curves of genus zero is nontrivial but complete. The Diophantine theory of curves of genus greater than zero including elliptic curves is highly nontrivial and incomplete. However, many general theorems exist and the theory is in constant development. In this paper we will detail the theory for conics.

We define a point in the plane to be rational if its coordinates are rational. We define a curve to be rational if it is given by an equation with rational coefficients. In this paper we are concerned with finding rational points on rational curves.

The first case that one might wish to consider is that of the linear equation. Linear equations pose no problems, especially if rational points are all that is desired, in which case the set of rational points on a rational line (a line with rational coefficients) is obviously in bijective correspondence with the rationals themselves by projection. In the case of a conic C with a known rational point, say p, the picture is almost as simple. One simply notes that for any other rational point on C, the line passing through it and p is a rational line. Similarly any rational line passing through p must intersect C in a second (not necessarily distinct) rational point in the projective plane, since solving for this intersection point amounts to solving a quadratic equation with one known rational root. In this paper we assume a basic notion of the ideas of projective geometry. For background see the appendix of Silverman and Tate. Thus, we may parameterize all of the rational points on C by taking our rational parameter to be the slope of the line passing through p. This rational parameterization gives an 'almost' (with the exception of finitely many points) bijective correspondence between the rational points on the curve and the rational points on the x-axis. We say that a curve whose rational points satisfy such a correspondence is *birationally equivalent* to the line. But the question of the existence of a rational point on a conic still remains. It turns out that over **Q**, a curve of genus zero is birationally equivalent to either a line or a conic. So, in fact, the characterization of the rational points on curves of genus zero is reduced to determining whether a given conic has a rational point.

To see that not all conics have rational points, consider the curve

$$C: \quad x^2 + y^2 = 3.$$

Converting to homogenous coordinates we have $X^2 + Y^2 = 3Z^2$. Again see the appendix of Silverman and Tate if this process is not familiar to you. For any integer triple (X, Y, Z) satisfying this homogenous equation, we can assume $\gcd(X, Y, Z) = 1$. If $3|X$ then we see that $3|Y$, so that finally $3|Z$. This is a contradiction. It follows that

$X^2 \equiv Y^2 \equiv 1 \pmod 3$. But this implies the left hand side of the homogenous equation is congruent to 2 (mod 3) while the right hand side is congruent to 0 (mod 3). It follows there are no rational points on C.

In order to understand the formulation of Hasse's theorem, the theorem that dominates the theory of curves of genus zero, we will first need to develop some of the basic properties of p-adic numbers. Recall that the usual absolute value is a *valuation* on **Q**. A valuation on a general field F is a function $|.|: F \rightarrow \mathbf{R}$ with the following properties:

(i) $|x| > 0$ for all $x \in F - \{0_F\}$, and $|0_F| = 0$

(ii) $|xy| = |x||y|$, $\forall\, x,y \in F$

(iii) $|x + y| \leq |x| + |y|$, $\forall\, x,y \in F$.

Now in F, $(-1)^2 = 1$, because $((-1)^{-1} + 1)(-1) = 1 + (-1) = 0$, so $(-1)^{-1} = -1$. Now we have that $|1|^2 = |1|$, so $|1| = 1$, and $|-1|^2 = |1| = 1$, so $|-1| = 1$ by property (ii) and more generally $|-r| = |r|$ $\forall\, r \in F$.

There are other valuations on **Q** other than the standard absolute value. We claim that the function $|.|_p$, defined by $|r|_p = p^{-\rho}$ for $r \neq 0$, where r has the unique expression $r = p^\rho u/v$, with $\rho, u, v \in \mathbf{Z}$, and $p \nmid u$, $p \nmid v$, and $|0|_p = 0$ is a valuation. It is easily seen that this definition satisfies (i) and (ii). To prove it satisfies (iii), let $|r|_p = p^{-\rho}$ and $|s|_p = p^{-\sigma}$, say $r = p^\rho u/v$ and $s = p^\sigma m/n$, and without loss of generality $\sigma \geq \rho$. Then

$$r + s = p^\rho(un + mvp^{\sigma - \rho})/vn \qquad (1)$$

so that $|r + s|_p \leq p^{-\rho}$. It is clear that in general we have $\forall\, r, s \in \mathbf{Q}$,

$$|r + s|_p \leq \max\{|r|_p, |s|_p\} \qquad (2)$$

If $|s|_p < |r|_p$ or $|r|_p < |s|_p$ then we actually have

$$|r + s|_p = \max\{|r|_p, |s|_p\} \qquad (3)$$

since the numerator of (1)—assuming without loss of generality $|s|_p < |r|_p$—cannot be divisible by p. A valuation satisfying this *ultrametric inequality* (2), which is clearly stronger than the triangle inequality (iii) is said to be *nonarchimedian*. This valuation may seem to be very counterintuitive at first, but will soon prove to be tremendously useful. Just remember that a reduced rational number is p-adically small if its numerator is divisible by a high power of p and p-adically large if it's denominator is divisible by a high power of p.

One of the many constructions of **R** from **Q** involves essentially extending **Q** by defining the reals to be equivalence classes of Cauchy sequences where the classes are defined by convergent. Recall that a Cauchy sequence is a sequence $\{a_n\}$ with the property that for every $\varepsilon > 0$ there exists $N > 0$ such that $n \geq m \geq N$ implies $|a_m - a_n| < \varepsilon$, and a field is complete if every such sequence converges. Notice that the definition of Cauchy sequence depends on the chosen valuation. This construction can be applied in much the same way for each p to our valuation $|.|_p$ to construct a p-adic field $\mathbf{Q}_p$ that is a *completion* of the field **Q** that is quite distinct from the usual completion **R**. Formally a field K with valuation $||.||$ is a completion of a field F with valuation $|.|$ if there exists a field monomorphism

$$\lambda: F \rightarrow K$$

with the property that $||\lambda a|| = |a| \ \forall \ a \in F$, and the properties:

(i) K is complete with respect to $||.||$

(ii) K is the closure of $\lambda(F)$ in the topology on K generated by the metric $d(x, y) = ||x - y||$.

We will assume that Q_p satisfies the above properties. It can be shown that the completion of a field always exists and is unique up to isomorphism. For more details on the construction of Q_p from Q see pp. 11-12 in Cassel's book. From now on we identify Q with a subfield of Q_p. Now that we have seen how the construction of Q_p goes, let us derive some of its fundamental properties.

We would hope that many of the properties of Q_p would follow through from the valuation $|.|_p$ on Q, just as is the case for R and Q with respect to the usual absolute value. This is in fact the case, and we will denote the valuation on Q_p by $|.|_p$ as suggestive notation and in light of the fact that completions preserve valuation on the subfield. Notice that by definition of closure of a topological space, for any a and $b \in Q_p$, we can find for any $\varepsilon > 0$, r and $s \in Q$ such that $|a - r|_p < \varepsilon$ and $|b - s|_p < \varepsilon$. So we can see that inequalities (2) and (3) must hold over Q_p by relating these statements about elements of Q_p to statements about sufficiently close rationals, for which we have seen (2) and (3) do hold. Note that the argument requires both the triangle inequality, as well as the fact that Q is dense in Q_p. We leave the details of the epsilon argument as an exercise to the reader. Given $a \in Q_p$ with $a \neq 0$, we can take $r \in Q$ such that $|r - a|_p < |a|_p$. It follows that $|r|_p = |a|_p$ by (2) and the contrapositive of (3). Therefore, the values that $|.|_p$ takes on Q_p are the same as those it takes on Q. We define the ring of *p-adic integers* Z_p to be $\{a \in Q_p \mid |a|_p \leq 1\}$. We leave it to the reader to prove the ring properties. We term the set of elements of Q_p with valuation 1 the *p-adic units*.

Dilip Das
18.704
Rogalski

We now go back to the problem of deciding whether a conic has a rational point. We express the equation of the conic in homogenous coordinates. The equation takes the form:

$$C: \quad F(\mathbf{X}) = \sum f_{ij}X_iX_j = 0$$

where $\mathbf{X} = (X_1, X_2, X_3)$, and in the sum, $1 \leq i, j \leq 3$. Also define $f_{ji} = f_{ij}$ for $1 \leq i < j \leq 3$. In other words, if $i \neq j$, f_{ij} is one half the coefficient of X_iX_j in the expanded form of F. Furthermore assume without loss of generality that the conic is nonsingular, or that $\det(\mathbf{f}) \neq 0$ where $\mathbf{f}$ is the 3×3 matrix with $\mathbf{f}_{ij} = f_{ij}$ as defined above. We are free to do this because the singular point of a rational conic must be rational point, so the question of the existence of a rational point is settled in this case. We leave it to the reader to show the equivalence of the singularity of $\mathbf{f}$ as a matrix and the existence of a point of singularity of C. The following theorem completely characterizing the existence of a rational point on a conic is due to Hasse, and is the goal of the remainder of the paper.

Hasse's Theorem: A conic C defined over **Q** has a rational point if and only if it has a point over **R** and over $\mathbf{Q}_p$ for every prime p.

Notice that necessity is trivial by the inclusion of **Q** in **R** and $\mathbf{Q}_p$. To show sufficiency we begin by performing a linear change of variables to diagonalize the matrix $\mathbf{f}$. Our linear transformation T is of the form

$$\mathbf{TY} = \mathbf{X}$$

where $\mathbf{T}_{ij} \in \mathbf{Q}$ and $\det(\mathbf{T}) \neq 0$. Then $\mathbf{Y}$ is a rational point on C_1 if and only if $\mathbf{TY} = \mathbf{X}$ is a rational point on C, where

$$C_1: \quad G(\mathbf{Y}) = \sum f_{ij}\,(\mathbf{TY})_i\,(\mathbf{TY})_j.$$

So if Hasse's theorem holds for C, then it holds for C_1. We can express our quadratic form as:

$$F(\mathbf{X}) = \mathbf{X}^T \cdot \mathbf{f} \cdot \mathbf{X}.$$

Our linear change of variables transforms our form into:

$$G(\mathbf{Y}) = \mathbf{Y}^T \cdot (\mathbf{T}^T \cdot \mathbf{f} \cdot \mathbf{T}) \cdot \mathbf{Y}.$$

The theory of bilinear forms says that there is a basis with respect to which the form F is orthogonal (see for example Artin). In other words we can transform F into the form $F(\mathbf{X}) = f_1X_1^2 + f_2X_2^2 + f_3X_3^2$, with f_i rational and nonzero with an appropriately chosen linear change of variable. We may now clear denominators so that $f_1, f_2, f_3 \in \mathbf{Z}$. Now we can divide through by any common prime factors so that $\gcd(f_1, f_2, f_3) = 1$. Finally if p divides exactly two of the f_i, we can replace the variable corresponding to the other coefficient, say X_j, by pX_j and then divide through by p. In this way we can reduce F to the form above where $\prod f_i$ is square free.

Suppose as in the hypothesis of Hasse's theorem that for each prime p there is a 3-vector $\mathbf{a} = (a_1, a_2, a_3)$ over $\mathbf{Q}_p$ such that $F(\mathbf{a}) = 0$. Since we know that the valuation $|.|_p$ assumes the same values over $\mathbf{Q}_p$ as over $\mathbf{Q}$ we multiply $\mathbf{a}$ by a sufficient power of p so that $\max |a_i|_p = 1$. We consider three cases, and derive conclusions in each that shall soon be put to use in proving Hasse's theorem.

Case 1: $p \neq 2$, $p \mid f_1f_2f_3$. Assume without loss of generality that $p \mid f_1$, so that $p \nmid f_2$ and $p \nmid f_3$. Then we have that $|f_1a_1^2|_p < 1$. Suppose that $|a_2|_p < 1$. Then we have that $|f_1a_1^2|_p = |f_2a_2^2 + f_3a_3^2|_p < 1$, so that $|f_3a_3^2|_p < 1$ by (2) and (3) and $|a_3|_p < 1$. But then we have that $|f_1a_1^2|_p = |f_2a_2^2 + f_3a_3^2|_p \leq p^{-2}$, so that $|a_1|_p < 1$, contrary to the condition that $\max |a_i|_p = 1$. It follows that $|a_2|_p = |a_3|_p = 1$. Now we have that

$|f_2a_2^2 + f_3a_3^2|_p = |f_2 + f_3(a_3/a_2)^2|_p < 1$, where $|a_3/a_2|_p = 1$. By definition of closure, there exists $s \in \mathbf{Q}$ such that $|f_2 + f_3s^2|_p < 1$ with $|s|_p = 1$. It is easy to prove using the fact that F_p is a field that for each such s, there is a unique integer $n \in \{0, 1, \ldots , p-1\}$ such that $|s - n|_p < 1$. Take n to be this integer. Then $|f_2 + f_3n^2|_p \le |f_2 + f_3s^2|_p + | f_3n^2 - f_3s^2|_p \le 1/p + 1/p^2 < 1$. So there is some $n \in \{0, 1, \ldots , p-1\}$ such that $f_2 + f_3n^2 \equiv 0 \pmod{p}$.

Case 2: $p = 2$, $2 \nmid f_1f_2f_3$. Suppose without loss of generality that a_3 is a unit. It follows that at least one of a_1 and a_2 is a unit since $1 = |f_3a_3^2|_2 = |f_1a_1^2 + f_2a_2^2|_2$, so one of $f_2a_1^2$ and $f_3a_2^2$ must have valuation 1 by (2). Suppose both are units. For all rational units r and s in $\mathbf{Q}_2$ we have that $|f_1r + f_2s|_2 < 1$, by looking at the parity of the numerator and denominator. It follows that the same result must hold for all units in $\mathbf{Q}_2$ by a simple epsilon argument. Therefore $1 = |f_3a_3^2|_2 = |f_1a_1^2 + f_2a_2^2|_2 < 1$, a contradiction. It follows that precisely two of the a_i are units, and so without loss of generality a_2 and a_3 are, and $|a_1|_p \le \frac{1}{2}$, so that $|f_2a_2^2 + f_3a_3^2|_p \le \frac{1}{4}$. By the same logic as in case 1 we have that $f_2 + f_3n^2 \equiv 0 \pmod{4}$ for some $n \in \{0, 1\}$. Since f_2 and f_3 are odd, we have

$$f_2 + f_3 \equiv 0 \pmod{4}.$$

Case 3: $p = 2$, $2 | f_1f_2f_3$. Without loss of generality $2|f_1$. If a_2 is not a unit, then we have that $|f_3a_3^2|_2 = |f_1a_1^2 + f_2a_2^2|_2 \le \frac{1}{2}$, so a_3 is not a unit. But then, $|f_1a_1^2|_2 = |f_2a_2^2 + f_3a_3^2|_2 \le \frac{1}{4}$, so a_1 is not a unit, contradictory to normalization. Therefore, without loss of generality a_2 and a_3 are units. If a_1 is not a unit, then we have that $|f_1a_1^2|_2 = |f_2a_2^2 + f_3a_3^2|_2 \le \frac{1}{8}$, so by the same logic as in cases 1 and 2, we have

$$f_2 + f_3 \equiv 0 \pmod{8}.$$

If a_1 is a unit, then for some rational units r, s, and t we have $|f_1r^2 + f_2s^2 + f_3t^2|_2 \leq 1/8$ by applying the definition of closure and using a simple epsilon argument. By clearing denominators and using the fact that $a^2 \equiv 1 \pmod 8$ for all integers a, we see that

$$f_1 + f_2 + f_3 \equiv 0 \pmod 8.$$

We now state without proof a theorem of Minkowski that is useful in proving many number theoretic theorems such as the four squares theorem, and whose proof can therefore be found in a number of elementary texts. See for instance Jones and Jones pp. 211.

Theorem: Let Λ be a subgroup of index m in the additive group $\mathbb{Z}^n$. Let $\mathbf{C}$ be a symmetric convex set of volume

$$V(\mathbf{C}) > 2^n m.$$

Then $\mathbf{C}$ and Λ have a nonzero point in common.

We begin by defining a subgroup Λ of $\mathbb{Z}^3$ by imposing various congruence conditions according to the conditions derived supposing the hypotheses of Hasse's theorem. Let $\mathbf{x} = (x_1, x_2, x_3)$

Case 1: $p \neq 2$, $p \mid f_1f_2f_3$. Recall without loss of generality $p|f_1$. We saw there exists an integer n such that $f_2 + f_3n^2 \equiv 0 \pmod p$. Impose the condition $x_3 \equiv nx_2 \pmod p$. We then have $F(\mathbf{x}) = f_1x_1^2 + f_2x_2^2 + f_3x_3^2 \equiv 0 \pmod p$.

Case 2: $p = 2$, $2 \nmid f_1f_2f_3$. Recall without loss of generality $f_2 + f_3 \equiv 0 \pmod 4$. Impose the conditions $x_1 \equiv 0 \pmod 2$, $x_2 \equiv x_3 \pmod 2$. This imples $F(\mathbf{x}) \equiv 0 \pmod 4$.

Case 3: $p = 2$, $2 \mid f_1f_2f_3$. Recall without loss of generality $2|f_1$ and either

$f_2 + f_3 \equiv 0 \pmod 8$ or $f_1 + f_2 + f_3 \equiv 0 \pmod 8$. So $s^2 f_1 + f_2 + f_3 \equiv 0 \pmod 8$, where s is either 0 or 1. Impose the conditions $x_2 \equiv x_3 \pmod 4$ and $x_1 \equiv sx_2 \pmod 2$. These imply $F(\mathbf{x}) \equiv 0 \pmod 8$.

The subgroup Λ obtained by imposing all of these congruence conditions is of index $4|f_1f_2f_3|$ in $\mathbb{Z}^3$. This is because the each of the congruences in case 1 is with respect to different primes and gives a subgroup of index p in $\mathbb{Z}^3$, so by the Chinese Remainder Theorem, the simultaneous imposition of these conditions gives a subgroup of index $\prod p_i$ in $\mathbb{Z}^3$, where the p_i range over all primes not equal to 2 dividing $f_1f_2f_3$. In the case that $2 \nmid f_1f_2f_3$ the conditions in case 2 clearly give an index 4 subgroup of $\mathbb{Z}^3$, and so by the Chinese remainder theorem the simultaneous imposition of all the conditions in cases 1 and 2 give a subgroup of index $4|f_1f_2f_3|$ in $\mathbb{Z}^3$. In the case that $2|\ f_1f_2f_3$, the conditions in case 3 clearly give an index 8 subgroup of $\mathbb{Z}^3$, and so by the Chinese remainder theorem the simultaneous imposition of all the conditions in cases 1 and 3 give a subgroup of index $4|f_1f_2f_3|$ in $\mathbb{Z}^3$. For all $\mathbf{x} \in \Lambda$ we have that $F(\mathbf{x}) \equiv 0 \pmod{4|f_1f_2f_3|}$.

We now apply Minkowski's theorem to Λ and the set

$$\mathbf{C}\text{: } |f_1|x_1^2 + |f_2|x_2^2 + |f_3|x_3^2 < 4|f_1f_2f_3|.$$

We see that $\mathbf{C}$ is clearly symmetric, and convex because it is an ellipsoid. Using the volume formula for the ellipsoid we have $V(\mathbf{C}) = 32\pi/3\sqrt{}(f_1f_2)\ (f_1f_3)\ (f_2f_3) > 2^3|4f_1f_2f_3|$. Applying Minkowski's theorem, we see that there exists a nonzero point $\mathbf{y}$ in $\Lambda \cap \mathbf{C}$. We have $F(\mathbf{y}) \equiv 0 \pmod{4|f_1f_2f_3|}$ and $|F(\mathbf{y})| \le |f_1|y_1^2 + |f_2|y_2^2 + |f_3|y_3^2 < 4|f_1f_2f_3|$, so that $F(\mathbf{y}) = 0$ as desired. This proves the existence of a projective point on our conic C given the hypotheses of Hasse's theorem, and concludes the proof of Hasse's theorem.

Notice that in addition to proving the existence of such a point, we have provided for any conic C with a rational point an ellipsoid in which a corresponding homogenous point can be found. So one can determine whether a given conic has a rational point in finite time. Also, we have not used the hypotheses of Hasse's theorem for $p \nmid 2f_1f_2f_3$. The reader may check that there always exist p-adic points on C for such primes and so we gain no additional information by considering them. Finally, we have not used the hypothesis of the existence of a point on C over $\mathbf{R}$. In fact, this is implied by the existence of points over $\mathbf{Q}_p$ for every p. P-adic extensions are widely used in the theory of rational points on curves of genus 1 and greater. The reader will thus find the present paper to not only conclude the case of genus 0, but also to introduce valuable techniques for further study of Diophantine equations.

Dilip Das
18.704
Rogalski

References

1. Cassels, J.W.S., Lectures on Elliptic Curves. Cambridge, England : Cambridge University Press, 1991.

2. Silverman, Joseph and John Tate. Rational Points on Elliptic Curves. New York: Springer-Verlag Inc., 1992.

3. Jones, Gareth and Mary Jones. Elementary Number Theory. London: Springer-Verlag Inc., 1998.