The Class Number Problem

1. Binary Quadratic Forms

Rajat Tandon

Rajat Tandon received his Ph.D. from Yale University, USA. After an initial period as a visiting post-doctoral fellow at TIFR, he joined the faculty of the North Eastern Hill University, where he taught for a few years. For about two decades now, he has been with the University of Hyderabad. His interests are in the area of number theory.

This two part article introduces the reader to the notion of 'class numbers'. The first part defines class numbers the way they arose in the study of 'binary quadratic equations'.

Remember the formula $\frac{-b\pm\sqrt{b^2-4ac}}{2a}$ that we all learnt in school. Indians have a long history of work on quadratics. The high point seems to have been when Brahmagupta in the early seventh century gave a method by which, knowing one solution (x, y) in integers of the equation $cX^2 + 1 = Y^2$ (Pell's Equation!), where c is a constant integer, he could generate an infinite family of solutions. But I am interested here in the above formula. I will always assume that a, b, c are integers. The quantity $b^2 - 4ac$ under the square root sign gives us information about the quadratic $aX^2 + bX + c$. For instance, it tells us whether the quadratic has any real roots – it must be positive for this to be so. It tells us whether the quadratic has any rational roots – it must be a perfect square for this to be so. We call it the discriminant of the quadratic $aX^2 + bX + c$. The class number problem is concerned with the following questions:

1) Given an integer Δ, are there any quadratics $F(X) = aX^2 + bX + c, a, b, c$ integers, whose discriminant $b^2 - 4ac$ equals Δ ?

2) If so, how many such quadratics exist? Can we classify them in any way?

It is obvious that if $\Delta = b^2 - 4ac$, a, b, c integers, then 4 divides Δ or 4 divides $\Delta - 1$,i.e., $\Delta \equiv 0$ or $1 (\mathrm{mod}\ 4)$. This is a necessary condition for there to be an integral quadratic with discriminant Δ. It is a simple exercise to show that it is also sufficient. So we have a complete answer to the

first question. The second question is considerably more complex.

Before proceeding further let me give a quick recap of the notion of an *equivalence relation.* A relation '$\sim$' on a set S is called an equivalence relation if it is reflexive ($x \sim x$), symmetric ($x \sim y \Rightarrow y \sim x$) and transitive ($x \sim y$ and $y \sim z \Rightarrow x \sim z$). Let $[x]$ denote the subset of S consisting of elements equivalent to x. It is called an equivalence class; note that any two equivalence classes are either identical or disjoint. Then S is the disjoint union of the distinct equivalence classes. We denote the set of equivalence classes by $S/\sim$.

Suppose we replace X by $X+1$ in the quadratic aX^2+bX+c. We have $a(X+1)^2+b(X+1)+c = aX^2+(b+2a)X+(a+b+c)$. The discriminant of this is $(b+2a)^2-4a(a+b+c) = b^2-4ac$. So the discriminant does not change if we replace $F(X) = aX^2 + bX + c$ by $F(X+1)$ and hence by $F(X+2), F(X+3)$, .. Similarly for $F(X-1), F(X-2)$ etc. Notice also that $\Delta = b^2 - 4ac$ is symmetric in a and c, i.e., if we replace $aX^2 + bX + c$ by $cX^2 + bX + a$ then the discriminant does not change. This indicates that it might be better to replace $F(X) = aX^2 + bX + c$ by the corresponding homogeneous polynomial in 2 variables $F(X,Y) = aX^2 + bXY + cY^2$. Then instead of the transformation $X \to X+1$ we take the transformation $X \to X+Y \;\; Y \to Y$

Let $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $W = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. T and W are members of $SL(2,\mathbf{Z})$, the group of 2×2 matrices with integer coefficients and determinant 1. If $A = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in SL(2,\mathbf{Z})$ and F is a homogeneous quadratic polynomial in two variables, we denote by $A \;\; F$ the polynomial obtained by replacing X by $\alpha X + \beta Y$ and Y by $\gamma X + \delta Y$ Observe that if $A, B \in SL(2,\mathbf{Z})$ then $A \cdot (B \cdot F) = AB \cdot F$. It is easy to check that if F has discriminant Δ, then so does $A \;\; F$ for any $A \in SL(2,\mathbf{Z})$. Denote by $S(\Delta)$ the set of all integral homogeneous quadratic polynomials in two variables of

discriminant Δ.

We define an equivalence relation on $S(\Delta)$ by $F \sim G$ if either $F = G$ or there exists a chain $F_1, F_2, \ldots, F_n$ in $S(\Delta)$ such that $F = F_1$, $G = F_n$ and each F_{i+1} is either $T \cdot F_i$ or T^{-1} F_i or W F_i; such a chain is called a chain from F to G. It is easy to see that this gives an equivalence relation on $S(\Delta)$. Hence $S(\Delta)$ can be partitioned into equivalence classes. We remark that it can be shown that $SL(2, \mathbf{Z})$ is generated by T and W and hence two forms F and G are equivalent if and only if there exists an $A \in SL(2, \mathbf{Z})$ such that A $F = G$.

Assume from now on that $\Delta < 0$. This is not because the case $\Delta > 0$ is uninteresting but because it is more difficult and less is known in this case. If the discriminant of $F(X, Y) = aX^2 + bXY + cY^2$ is Δ then it is also so for $-F$. Note that $\Delta < 0$ implies that a and c have the same sign. We define $S_1(\Delta)$ to be the subset of $S(\Delta)$ consisting of those forms F for which a and c are positive and $S_2(\Delta)$ its complement. Then $F \mapsto -F$ is a bijection from $S_1(\Delta)$ to $S_2(\Delta)$. It is also easy to see that no member of $S_1(\Delta)$ can be equivalent to any member of $S_2(\Delta)$. We restrict ourselves to $S_1(\Delta)$.

Definition: The form $F(X, Y) = aX^2 + bXY + cY^2$ of $S_1(\Delta)$ is said to be *almost reduced* if $|b| \leq a \leq c$.

Theorem: Each equivalence class in $S_1(\Delta)$ has at least one almost reduced form.

Proof: Consider an equivalence class with an element $F(X, Y) = aX^2 + bXY + cY^2$ in it. If $a > c$ replace F by W $F = F_1$ (say). Then $F_1(X, Y) = a_1X^2 + b_1XY + c_1Y^2$ with $a_1 = c$ and $c_1 = a$, and so $a_1 \leq c_1$. Notice $a > a_1$. If now $|b_1| \leq a_1$, F_1 is reduced. If not, find an integer n such that $|b_1 + 2a_1n| \leq a_1$. Replace F_1 by $F_2 = T^n \cdot F_1$. Then

$$\begin{aligned} F_2(X, Y) &= a_1(X + nY)^2 + b_1(X + nY)Y + c_1Y^2 \\ &= a_1X^2 + (b_1 + 2a_1n)XY + (a_1n^2 + b_1n + c_1)Y^2 \\ &= a_2X^2 + b_2XY + c_2Y^2 \end{aligned}$$

(say) with $|b_2| \leq a_2$ and $a_2 = a_1$. But now a_2 may not be less than or equal to c_2. If so, again apply W and continue as before. After a finite number of steps we get an almost reduced form (finite because $a \geq a_1 \geq a_2 \ldots > 0$).

Corollary: The number of equivalence classes in $S_1(\Delta)$ is finite.

Proof: It suffices to show that the number of almost reduced forms is finite. If $aX^2 + bXY + cY^2$ is almost reduced of discriminant Δ then

$$a \leq c \Rightarrow 4a^2 \leq 4ac = b^2 - \Delta \leq a^2 - \Delta.$$

Hence $3a^2 \leq |\Delta|$. Since a is a positive integer there are only finitely many possible values of a and hence of b. Once a and b are given, c is uniquely determined.

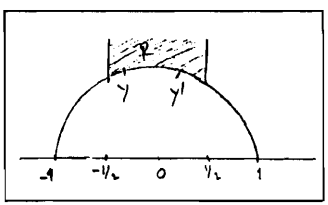
A natural question to ask is: is there precisely one almost reduced form in each equivalence class? The answer is – almost but not quite.

We know that $X^2 + \frac{b}{a}X + \frac{c}{a}$ has two non-real roots (because $\Delta < 0$), say τ and $\bar{\tau}$. One of these (say τ) will lie in the upper half plane $\{x+iy : x, y \in \mathbf{R}, y > 0\}$. Hence $F(X,Y) = aX^2+bXY+cY^2 = a(X+\tau Y)(X+\bar{\tau}Y)$ with $b = a(\tau+\bar{\tau})$ and $c = a\tau\bar{\tau}$. Hence to say that F is almost reduced is equivalent to saying that $|\tau + \bar{\tau}| \leq 1$ and $\tau\bar{\tau} \geq 1$, i.e., $\tau \in \mathbf{R}$ where $\mathbf{R}$ is the region shown in *Figure 1* (including the boundary).

Notice that if τ is on the left vertical boundary of $\mathbf{R}$ then $\tau + 1$ is on the right vertical boundary which is also in $\mathbf{R}$. Similarly, if τ is on the curve at Y then $\frac{-1}{\tau}$ is at Y'. In view of this we make the following definition:

Definition: We say that $F(X,Y) = aX^2 + bXY + cY^2$ is *reduced* if the corresponding $\tau \in \mathbf{R}$ but $\tau \notin$ the left boundary of $\mathbf{R}$, i.e., the left vertical boundary and curve Y This is equivalent to saying that $|b| \leq a \leq c$, and in case $a = |b|$ then $b > 0$, and in case $a = c$ then $b \geq 0$. We now have the expected theorem.

Figure 1.



Theorem: In each equivalence class of $S_1(\Delta)$ there is pre-

Δ	possible a	possible b, c	reduced forms	$\underline{h}(\Delta)$
-3	$a=1$	$b=1, c=1$	X^2+XY+Y^2	1
-4	$a=1$	$b=0, c=1$	X^2+Y^2	1
-7	$a=1$	$b=1, c=2$	$X^2+XY+2Y^2$	1
-8	$a=1$	$b=0, c=2$	X^2+2Y^2	1
-11	$a=1$	$b=1, c=3$	$X^2+XY+3Y^2$	1
-12	$a=1$ or 2	$b=0, c=3$ if $a=1$	X^2+3Y^2	2
		$b=2, c=2$ if $a=2$	$2(X^2+XY+Y^2)$	
-15	$a=1$ or 2	$b=1, c=4$ if $a=1$	$X^2+XY+4Y^2$	2
		$b=1, c=2$ if $a=2$	$2X^2+XY+2Y^2$	
-16	$a=1$ or 2	$b=0, c=4$ if $a=1$	X^2+4Y^2	2
		$b=0, c=2$ if $a=2$	$2(X^2+Y^2)$	
-19	$a=1$ or 2	$b=1, c=5$ if $a=1$	$X^2+XY+5Y^2$	1
-20	$a=1$ or 2	$b=0, c=5$ if $a=1$	X^2+5Y^2	2
		$b=2, c=3$ if $a=2$	$2X^2+2XY+3Y^2$	
-23	$a=1$ or 2	$b=1, c=6$ if $a=1$	$X^2+XY+6Y^2$	3
		$b=1, c=3$ if $a=2$	$2X^2+XY+3Y^2$	
		$b=-1, c=3$ if $a=2$	$2X^2-XY+3Y^2$	

cisely one reduced form.

In the chart given is a list of reduced forms for low values of $|\Delta|$; $\underline{h}(\Delta)$ is the number of reduced forms.

We notice that some forms in the list are constant multiples of forms which came earlier in the list.

Definition A form $aX^2+bXY+cY^2$ is said to be *primitive* if $(a,b,c)=1$.

We let $h(\Delta)$ be the number of primitive reduced forms of discriminant Δ in $S_1(\Delta)$. $h(\Delta)$ is known as the *class number* of the forms with discriminant Δ. Notice that $h(\Delta)$ is 1 for $\Delta=-3,-4,-7,-8,-11,-12,-16,-19$ in the list.

Definition: An integer $\Delta \equiv 0$ or $1 (\mathrm{mod}\ 4)$ is said to be a *fundamental discriminant* if it is not of the form $\Delta_0 n^2$ where Δ_0 is a discriminant and n an integer.

For instance -12 and -16 are not fundamental discriminants. Notice that if Δ is fundamental, then a form of discriminant Δ is always primitive. Notice also that if Δ is

In 1954, an amateur mathematician Heegner, in Germany had proved the same result but his proof had some gaps which were responsible for mathematicians expressing reservations about the proof. But later it was shown by Stark that the arguments of Heegner can be made rigorous and he managed to make Heegner's proof work. In fact, Heegner's ideas, in particular his construction of what are now called *Heegner points*, have proved to be very fruitful in later work on elliptic curves.

fundamental, then it cannot have an odd square factor. We will see later that if Δ is fundamental then it has another interpretation.

In 1934, Heilbronn showed that $h(\Delta) \to \infty$ as $\Delta \to -\infty$ from which it follows (how?) that given any natural number N there are only a finite number of negative fundamental discriminants Δ for which the class number, $h(\Delta) = N$. One of the questions that suggests itself from the above is: what are the negative fundamental Δ for which $h(\Delta)$ is 1? Above we have given six such Δ's. Here are three more: $\Delta = -43, -67, -163$. In 1800, Gauss conjectured that there were no more.

In 1936, Siegel showed that for every $\epsilon > 0$ there exists a positive constant C_ϵ such that $h(\Delta) \geq C_\epsilon |\Delta|^{\frac{1}{2}-\epsilon}$. However, the result showed the existence of C_ϵ but not how to compute it. But his proof showed that there cannot be two 'large' values of $|\Delta|$'s for which $h(\Delta)$ is small. From this it was proved that there is possibly just one other Δ (call it Δ_{10}) for which $h(\Delta) = 1$ and this Δ must be very large indeed. In 1966, Harold Stark, in his thesis, showed that Δ_{10} does not exist [1]. The same methods were applied to the negative Δ for which $h(\Delta) = 2$ and it was found that there are 18 such Δ's, the largest value of $|\Delta|$ being 427 (Baker, Stark, Montgomery etc). In 1986 using powerful methods in algebraic geometry D Goldfeld, B H Gross and D Zagier solved the problem of fundamental negative Δ with $h(\Delta) = 3$.

Remark: The Δ for which $h(\Delta) = 1$ have remarkable properties. For instance, if p is a positive prime number which is congruent to 3(mod 4) and $h(-p) = 1$ then $x^2 + x + \frac{p+1}{4}$ is a prime number for all x such that $0 \leq x \leq \frac{p-7}{4}$.

Address for Correspondence
Rajat Tandon
Department of Mathematics
University of Hyderabad
Central University P.O.
Hyderabad 500 046, India

Suggested Reading

- **H M Stark. The complete determination of the complex quadratic fields of Class number one. *Michigan Math J.* 14. 1–27,1967.**
- **J P Serre. *A Course in Arithmetic.* Narosa Publishing House. New Delhi, 1979.**
- **D Flath. *Introduction to Number Theory.* John Wiley and Sons. New York, 1989.**