The Ubiquitous modular group

B.Sury
Indian Statistical Institute
Bangalore, India
sury@isibang.ac.in
Indian Institute of Science Education and Research
Pune, India
Lectures in October 2010

October 23, 2010

No one would have had the imagination to invent them

$$\frac{e^{-2\pi/5}}{1+}\,\frac{e^{-2\pi}}{1+}\,\frac{e^{-4\pi}}{1+}\,\frac{e^{-6\pi}}{1+}\cdots = \sqrt{\frac{5+\sqrt{5}}{2}} - \frac{\sqrt{5}+1}{2}.$$

No one would have had the imagination to invent them

$$\frac{e^{-2\pi/5}}{1+}\,\frac{e^{-2\pi}}{1+}\,\frac{e^{-4\pi}}{1+}\,\frac{e^{-6\pi}}{1+}\cdots = \sqrt{\frac{5+\sqrt{5}}{2}} - \frac{\sqrt{5}+1}{2}.$$

This continued fraction appeared in Ramanujan's first letter to Hardy written on January 16, 1913. Of this and some other formulae in that letter, Hardy said in 1937:

"They defeated me completely. I had never seen anything in the least like them before. A single look at them is enough to show that they could only be written down by a mathematician of the highest class. They must be true because, if they were not true, no one would have had the imagination to invent them."

"They defeated me completely. I had never seen anything in the least like them before. A single look at them is enough to show that they could only be written down by a mathematician of the highest class. They must be true because, if they were not true, no one would have had the imagination to invent them."

The continued fraction quoted in the beginning can be proved using the so-called Rogers-Ramanujan identities which are, in turn, intimately connected to the theory of partitions to which Ramanujan made fundamental contributions.

What is the exact number of partitions?

$$p(n) = \frac{1}{\sqrt{2}\pi} \sum_{q=1}^{\infty} A_q(n) \sqrt{q} [\frac{d}{dx} \frac{\sinh((\pi/q)(\frac{2(x-1/24)}{3})^{1/2})}{(x-1/24)^{1/2}}]_{x=n}$$

where $A_q(n) = \sum \omega_{p,q} e^{-2np\pi i/q}$, the last sum being over p's prime to q and less than it, $\omega_{p,q}$ is a certain $24q$-th root of unity.

What is the exact number of partitions?

$$p(n) = \frac{1}{\sqrt{2}\pi} \sum_{q=1}^{\infty} A_q(n)\sqrt{q}[\frac{d}{dx} \frac{\sinh((\pi/q)(\frac{2(x-1/24)}{3})^{1/2})}{(x-1/24)^{1/2}}]_{x=n}$$

where $A_q(n) = \sum \omega_{p,q} e^{-2np\pi i/q}$, the last sum being over p's prime to q and less than it, $\omega_{p,q}$ is a certain $24q$-th root of unity.

What a bizarre expression relating $p(n)$ with 24-th roots of unity!

Let p be a prime of the form $4k+1$. Then, the set of values assumed by $px^2 + 2(\frac{p-1}{2})!xy + qy^2$ at integer values of x, y coincides with those assumed by $x^2 + y^2$ where

$$((\frac{p-1}{2})!)^2 + 1 = pq$$

In particular, p is a sum of two squares.

$e^{\pi\sqrt{163}}$ is 'almost' an integer

This inriguing title has the more precise formulation

$$e^{\pi\sqrt{163}} - integer = 196884\ e^{-\pi\sqrt{163}} + 21493760\ e^{-2\pi\sqrt{163}} \ldots \approx 0!$$

Sums of powers of divisors

If $\sigma_r(n)$ is the sum of the r-th powers of divisors of n, then we have relations like:

Sums of powers of divisors

If $\sigma_r(n)$ is the sum of the r-th powers of divisors of n, then we have relations like:

$$\sigma_7(n) = \sigma_3(n) + 120 \sum_{m=1}^{n-1} \sigma_3(m)\sigma_3(n-m)$$

$$11\sigma_9(n) = 21\sigma_5(n) - 10\sigma_3(n) + 5040 \sum_{m=1}^{n-1} \sigma_3(m)\sigma_5(n-m)$$

Expressions as sums of squares

If $r_k(n)$ is the number of ways a positive integer can be expressed as $x_1^2 + \cdots + x_k^2$ for integral x_i's, then

Expressions as sums of squares

If $r_k(n)$ is the number of ways a positive integer can be expressed as $x_1^2 + \cdots + x_k^2$ for integral x_i's, then

$$r_4(n) = 8 \sum \{d : d|n, 4 \not| d\}$$

$$r_8(n) = 16 \sum_{d|n} (-1)^{n+d} d^3$$

Shortest continued fraction for rationals

Suppose one considers continued fraction expansions of a rational number permitting negative integers.

Shortest continued fraction for rationals

Suppose one considers continued fraction expansions of a rational number permitting negative integers.

For instance, $\frac{31}{13} = [2; 2, 1, 1, 2] = [2; 3, -3, 2]$ are two C.F.s where the second one is obtained by choosing a nearest integer at each stage.

Is it true that a continued fraction expansion of a rational number obtained by choosing the nearest integer at each step, is a shortest one?

Is it true that a continued fraction expansion of a rational number obtained by choosing the nearest integer at each step, is a shortest one?

Is it true that a C.F. $[b_0; b_1, \cdots, b_n]$ for a rational number is a shortest one if and only if $|b_i| \geq 2$ for all $i \neq 0$ and $b_1, \cdots, b_n$ does not have substring of the form

$$2, -3, 3, -3, 3, \cdots, -3, 3, -2?$$

Is it true that a continued fraction expansion of a rational number obtained by choosing the nearest integer at each step, is a shortest one?

Is it true that a C.F. $[b_0; b_1, \cdots, b_n]$ for a rational number is a shortest one if and only if $|b_i| \geq 2$ for all $i \neq 0$ and $b_1, \cdots, b_n$ does not have substring of the form

$$2, -3, 3, -3, 3, \cdots, -3, 3, -2?$$

The answer to both questions turn out to be affirmative.

Integer programming

There is a polynomial time algorithm to determine whether a triangle on the plane contains a lattice point or not.

There is a polynomial time algorithm to determine whether a triangle on the plane contains a lattice point or not.

More generally, one can decide in polynomial time (in the length of A, B) whether an integral solution vector exists or not for the system of inequalities $AX \leq B$ where A, B are integer matrices of sizes $m \times n$ and $m \times 1$ respectively, wherein we look for solutions for $n \times 1$ integer-tuples X.

The common thread in these rather different statements is the modular group Γ.

The common thread in these rather different statements is the modular group Γ.

Depending on the situation Γ is the group of 2×2 integer matrices of determinant ± 1 or the subgroup of matrices with determinant 1 or the quotient of that by the center $\pm I$.

The common thread in these rather different statements is the modular group Γ.

Depending on the situation Γ is the group of 2×2 integer matrices of determinant ± 1 or the subgroup of matrices with determinant 1 or the quotient of that by the center $\pm I$.

The heroine of our story is Γ and we accompany her in a journey touching several aspects of elementary number theory including the above ones.

What is the number of partitions?

Start with partitions:

$$p(n) = \frac{1}{\sqrt{2}\pi} \sum_{q=1}^{\infty} A_q(n) \sqrt{q} [\frac{d}{dx} \frac{\sinh((\pi/q)(\frac{2(x-1/24)}{3})^{1/2})}{(x-1/24)^{1/2}}]_{x=n}$$

What is the number of partitions?

Start with partitions:

$$p(n) = \frac{1}{\sqrt{2}\pi} \sum_{q=1}^{\infty} A_q(n)\sqrt{q}[\frac{d}{dx} \frac{\sinh((\pi/q)(\frac{2(x-1/24)}{3})^{1/2})}{(x-1/24)^{1/2}}]_{x=n}$$

where $A_q(n) = \sum \omega_{p,q} e^{-2np\pi i/q}$, the last sum being over p's prime to q and less than it, $\omega_{p,q}$ is a certain $24q$-th root of unity.

What is the number of partitions?

Start with partitions:

$$p(n) = \frac{1}{\sqrt{2}\pi} \sum_{q=1}^{\infty} A_q(n)\sqrt{q}[\frac{d}{dx} \frac{\sinh((\pi/q)(\frac{2(x-1/24)}{3})^{1/2})}{(x-1/24)^{1/2}}]_{x=n}$$

where $A_q(n) = \sum \omega_{p,q} e^{-2np\pi i/q}$, the last sum being over p's prime to q and less than it, $\omega_{p,q}$ is a certain $24q$-th root of unity.

The exact formula above is due to Rademacher but it comes out of an asymptotic formula due to Hardy and Ramanujan.

They show that $p(n)$ is the integer nearest to $\frac{1}{2\sqrt{2}}\sum_{q=1}^{v}\sqrt{q}A_q(n)\psi_q(n)$, where $A_q(n)=\sum\omega_{p,q}e^{-2np\pi i/q}$, the last sum being over p's prime to q and less than it, $\omega_{p,q}$ is a certain $24q$-th root of unity, v is of the order of $\sqrt{n}$, and

$$\psi_q(n)=\frac{d}{dn}(exp\{C\sqrt{n-\frac{1}{24}}/q\}\ ,\ C=\pi\sqrt{2/3}.$$

We may take $v=4$ when $n=100$. For $n=200$, we may take $v=5$.

They show that $p(n)$ is the integer nearest to $\frac{1}{2\sqrt{2}}\sum_{q=1}^{v}\sqrt{q}A_q(n)\psi_q(n)$, where $A_q(n)=\sum\omega_{p,q}e^{-2np\pi i/q}$, the last sum being over p's prime to q and less than it, $\omega_{p,q}$ is a certain $24q$-th root of unity, v is of the order of $\sqrt{n}$, and

$$\psi_q(n)=\frac{d}{dn}(exp\{C\sqrt{n-\frac{1}{24}}/q\}\ ,\ C=\pi\sqrt{2/3}.$$

We may take $v=4$ when $n=100$. For $n=200$, we may take $v=5$.

While reviewing the collected works of Ramanujan in the Mathematical Gazette, Littlewood says of this latter paper:

“The reader does not need to be told that this is a very astonishing theorem, and he will readily believe that the methods by which it was established involve a new and important principle, which has been found very fruitful in other fields. The story of the theorem is a romantic one.

“The reader does not need to be told that this is a very astonishing theorem, and he will readily believe that the methods by which it was established involve a new and important principle, which has been found very fruitful in other fields. The story of the theorem is a romantic one.

To do it justice I must infringe a little the rules about collaboration. I therefore add that Prof. Hardy confirms and permits my statements of bare fact.

One of Ramanujan's Indian conjectures was that the first term of the sum was a very good approximation to $p(n)$; this was established without great difficulty. From this point the real attack begins. The next step in development was to treat the above sum as an "asymptotic" series, of which a fixed number of terms were to be taken, the error being of the order of the next term. But from now to the very end Ramanujan always insisted that much more was true than had been established: "there must be a formula with error $O(1)$."

One of Ramanujan's Indian conjectures was that the first term of the sum was a very good approximation to $p(n)$; this was established without great difficulty. From this point the real attack begins. The next step in development was to treat the above sum as an "asymptotic" series, of which a fixed number of terms were to be taken, the error being of the order of the next term. But from now to the very end Ramanujan always insisted that much more was true than had been established: "there must be a formula with error $O(1)$."

This was his most important contribution; it was both absolutely essential and most extraordinary."

The exact formula for $p(n)$ written down by Rademacher essentially follows from the fact that the so-called Dedekind eta function is a modular form.

The exact formula for $p(n)$ written down by Rademacher essentially follows from the fact that the so-called Dedekind eta function is a modular form.

A modular form f is roughly a function defined on the upper half-plane such that $f((az+b)/(cz+d))$ is of the form $(cz+d)^k f(z)$ for $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ in the modular group.

The exact formula for $p(n)$ written down by Rademacher essentially follows from the fact that the so-called Dedekind eta function is a modular form.

A modular form f is roughly a function defined on the upper half-plane such that $f((az+b)/(cz+d))$ is of the form $(cz+d)^k f(z)$ for $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ in the modular group.

We mention in passing that Rademacher used what are known as Ford circles which are related to Farey fractions which also have connections with the modular group as we shall see!)

$$\sum_{n=0}^{\infty} p(n)q^n = \prod_{n=1}^{\infty}(1-q^n)^{-1}$$

$$\sum_{n=0}^{\infty} p(n)q^n = \prod_{n=1}^{\infty}(1-q^n)^{-1}$$

The eta function defined on the upper half-plane by

$$\eta(z) = e^{2i\pi z/24}\prod_{n=1}^{\infty}(1-e^{2i\pi nz})$$

is related to it by

$$\eta(z) = e^{2i\pi z/24}/\sum_{n\geq 0} p(n)e^{2i\pi nz}.$$

$$\sum_{n=0}^{\infty} p(n)q^n = \prod_{n=1}^{\infty}(1-q^n)^{-1}$$

The eta function defined on the upper half-plane by

$$\eta(z) = e^{2i\pi z/24}\prod_{n=1}^{\infty}(1-e^{2i\pi nz})$$

is related to it by

$$\eta(z) = e^{2i\pi z/24}/\sum_{n\geq 0} p(n)e^{2i\pi nz}.$$

The eta function satisfies

$$\eta(-1/z) = \sqrt{\frac{z}{i}}\eta(z).$$

Here, the square-root is the branch having nonnegative real part.

Using the transformation above and one for $\eta(z+1)$, one gets a transformation formula for $\eta((az+b)/(cz+d))$ for any $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ in $SL(2, \mathbf{Z})$ easily since this group can be generated by the two nice matrices $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ and $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$.

Using the transformation above and one for $\eta(z+1)$, one gets a transformation formula for $\eta((az+b)/(cz+d))$ for any $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ in $SL(2,\mathbf{Z})$ easily since this group can be generated by the two nice matrices $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ and $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$.

The transformation for $c > 0$ is:

$$\log \eta((az+b)/(cz+d)) = \log \eta(z) - \frac{i\pi}{4}$$

$$+\frac{1}{2}\log(cz+d) - i\pi s(d,c) + \frac{i\pi(a+d)}{12c}$$

where $s(d,c)$ is a so-called Dedekind sum.

Using the transformation above and one for $\eta(z+1)$, one gets a transformation formula for $\eta((az+b)/(cz+d))$ for any $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ in $SL(2,\mathbf{Z})$ easily since this group can be generated by the two nice matrices $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ and $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$.

The transformation for $c > 0$ is:

$$\log \eta((az+b)/(cz+d)) = \log \eta(z) - \frac{i\pi}{4}$$

$$+\frac{1}{2}\log(cz+d) - i\pi s(d,c) + \frac{i\pi(a+d)}{12c}$$

where $s(d,c)$ is a so-called Dedekind sum.

This expression quite easily gives the explicit expression for $p(n)$. We won't talk more about this but refer the interested reader to the book 'Theory of Partitions', by George Andrews.

The modular group and hyperbolic geometry have intimate connections with continued fractions.

Continued fractions and geodesics in hyperbolic geometry

The modular group and hyperbolic geometry have intimate connections with continued fractions.

$\frac{31}{13} = [2; 2, 1, 1, 2] = [2; 3, -3, 2]$ are two continued fractions of which the second, which is shorter, is obtained by choosing the nearest integer at each stage.

Continued fractions and geodesics in hyperbolic geometry

The modular group and hyperbolic geometry have intimate connections with continued fractions.

$\frac{31}{13} = [2; 2, 1, 1, 2] = [2; 3, -3, 2]$ are two continued fractions of which the second, which is shorter, is obtained by choosing the nearest integer at each stage.

Associated to a continued fraction, it turns out that one has a path in the so-called Farey graph and, this is a geodesic on the Poincaré hyperbolic upper half-plane if the C.F. is a shortest one. Facts about continued fractions can often be proven using the geometry of the hyperbolic plane aided by the modular group.

Consider for any (positive or negative) integer a the transformation $S_a : z \mapsto \frac{az+1}{z}$ on the upper half-plane. One also considers the action on 0 and at $i\infty$, the 'point at infinity' where the definition is

$$S_a(0) = \infty, S_a(\infty) = a.$$

Consider for any (positive or negative) integer a the transformation $S_a : z \mapsto \frac{az+1}{z}$ on the upper half-plane. One also considers the action on 0 and at $i\infty$, the 'point at infinity' where the definition is

$$S_a(0) = \infty, S_a(\infty) = a.$$

Now, $S_a = T^a S$ where $Tz = z + 1, Sz = 1/z$.

Consider for any (positive or negative) integer a the transformation $S_a : z \mapsto \frac{az+1}{z}$ on the upper half-plane. One also considers the action on 0 and at $i\infty$, the 'point at infinity' where the definition is

$$S_a(0) = \infty, S_a(\infty) = a.$$

Now, $S_a = T^a S$ where $Tz = z + 1, Sz = 1/z$.

$$S_{a_0} \circ S_{a_1} \circ \cdots \circ S_{a_n}(\infty) = [a_0; a_1, \cdots , a_n]$$

Consider for any (positive or negative) integer a the transformation $S_a : z \mapsto \frac{az+1}{z}$ on the upper half-plane. One also considers the action on 0 and at $i\infty$, the 'point at infinity' where the definition is

$$S_a(0) = \infty, S_a(\infty) = a.$$

Now, $S_a = T^a S$ where $Tz = z + 1, Sz = 1/z$.

$$S_{a_0} \circ S_{a_1} \circ \cdots \circ S_{a_n}(\infty) = [a_0; a_1, \cdots, a_n]$$

S_a's generate $GL(2, \mathbf{Z})$.
Thus, we have a bijective correspondence between finite continued fractions of integers and words in T, S.

The Farey graph is formed with vertices as rational numbers and a/b and c/d joined by a semicircle if $ad - bc = \pm 1$.

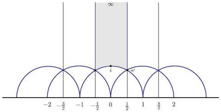
The Farey graph is formed with vertices as rational numbers and a/b and c/d joined by a semicircle if $ad - bc = \pm 1$.

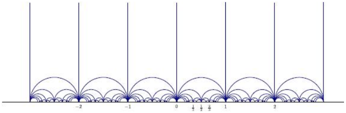
The geodesics on the upper half-plane for the hyperbolic metric are vertical lines and semicircles perpendicular to the X-axis.

The Farey graph is formed with vertices as rational numbers and a/b and c/d joined by a semicircle if $ad - bc = \pm 1$.

The geodesics on the upper half-plane for the hyperbolic metric are vertical lines and semicircles perpendicular to the X-axis.

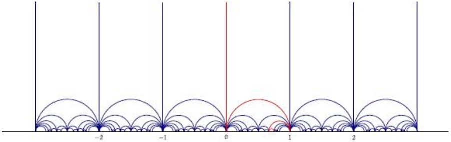
We have a bijection between finite continued fractions of integers and finite paths from ∞ in the Farey graph.


∞
i
ω
−2
−3/2
−1
−1/2
0
1/2
1
3/2
2



Farey graph

fareygraph2.jpg


−2
−1
0
1
2

Thus, the question whether the ‘nearest-integer’ continued fraction of a rational number is a shortest one is equivalent to the question as to whether the corresponding path in the Farey graph is a geodesic. This interpretation helps in answering the question affirmatively.

Thus, the question whether the 'nearest-integer' continued fraction of a rational number is a shortest one is equivalent to the question as to whether the corresponding path in the Farey graph is a geodesic. This interpretation helps in answering the question affirmatively.

Not only that; the recognition of paths which are geodesics leads to the statement on continued fractions made earlier; viz.

Thus, the question whether the 'nearest-integer' continued fraction of a rational number is a shortest one is equivalent to the question as to whether the corresponding path in the Farey graph is a geodesic. This interpretation helps in answering the question affirmatively.

Not only that; the recognition of paths which are geodesics leads to the statement on continued fractions made earlier; viz.

A C.F. $[b_0; b_1, \cdots, b_n]$ for a rational number is a shortest one if and only if $|b_i| \geq 2$ for all $i \neq 0$ and $b_1, \cdots, b_n$ does **not** have a substring of the form

$$2, -3, 3, -3, 3, \cdots, -3, 3, -2.$$

Lest someone think that continued fractions is old hat, here is an open problem !

Lest someone think that continued fractions is old hat, here is an open problem !

Zaremba's conjecture:
There exists an $N > 0$ such that every $f > 0$ arises as the denominator of a rational number $\frac{a}{f} = [a_0, \cdots, a_n]$ with $1 \leq a_i \leq N$.

Given two rationals $\frac{a}{b} < \frac{c}{d} \in [0,1]$, let us make the mistake that a child might make while learning the addition of fractions. Think of the sum of these two fractions as $\frac{a+c}{b+d}$. This mistake turns out to be a fruitful one!

What is in a question? - a Farey tale

Given two rationals $\frac{a}{b} < \frac{c}{d} \in [0,1]$, let us make the mistake that a child might make while learning the addition of fractions. Think of the sum of these two fractions as $\frac{a+c}{b+d}$. This mistake turns out to be a fruitful one!

Call $\frac{a+c}{b+d}$ the *mediant* of the two fractions $\frac{a}{b}$ and $\frac{c}{d}$.

Given two rationals $\frac{a}{b} < \frac{c}{d} \in [0, 1]$, let us make the mistake that a child might make while learning the addition of fractions. Think of the sum of these two fractions as $\frac{a+c}{b+d}$. This mistake turns out to be a fruitful one!

Call $\frac{a+c}{b+d}$ the *mediant* of the two fractions $\frac{a}{b}$ and $\frac{c}{d}$.

Starting with the extremes $\frac{0}{1}, \frac{1}{1}$, we have the mediant $\frac{1}{2}$ which lies in between; that is,

$$\frac{0}{1} < \frac{1}{2} < \frac{1}{1}.$$

Given two rationals $\frac{a}{b} < \frac{c}{d} \in [0,1]$, let us make the mistake that a child might make while learning the addition of fractions. Think of the sum of these two fractions as $\frac{a+c}{b+d}$. This mistake turns out to be a fruitful one!

Call $\frac{a+c}{b+d}$ the *mediant* of the two fractions $\frac{a}{b}$ and $\frac{c}{d}$.

Starting with the extremes $\frac{0}{1}, \frac{1}{1}$, we have the mediant $\frac{1}{2}$ which lies in between; that is,

$$\frac{0}{1} < \frac{1}{2} < \frac{1}{1}.$$

At the next stage, the mediant of $\frac{0}{1}$ and $\frac{1}{2}$ is $\frac{1}{3}$ and that of $\frac{1}{2}$ and $\frac{1}{1}$ is $\frac{2}{3}$.

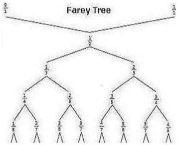
We can place the mediants in between their respective progenitors and get

$$\frac{0}{1}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{1}{1}$$

We can place the mediants in between their respective progenitors and get

$$\frac{0}{1}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{1}{1}$$

In this manner, we can iterate and get strings of fractions. This produces a tree - known as the Farey tree which looks as follows.


Farey Tree

Neighbouring fractions at any level of the tree have the property: $\frac{a}{b} < \frac{c}{d}$ are neighbours, then $bc - ad = 1$!

Neighbouring fractions at any level of the tree have the property: $\frac{a}{b} < \frac{c}{d}$ are neighbours, then $bc - ad = 1$!

If this property holds for the above neighbours, then go to the next level below on the tree and show that the property holds for the neighbours

$$\frac{a}{b} < \frac{a+c}{b+d} < \frac{c}{d}$$

This is clear!

Neighbouring fractions at any level of the tree have the property: $\frac{a}{b} < \frac{c}{d}$ are neighbours, then $bc - ad = 1$!

If this property holds for the above neighbours, then go to the next level below on the tree and show that the property holds for the neighbours

$$\frac{a}{b} < \frac{a+c}{b+d} < \frac{c}{d}$$

This is clear!

The matrix $\begin{pmatrix} c & a \\ d & b \end{pmatrix} \in SL(2, \mathbf{Z})$.

The *unimodularity* also shows easily the inequalities

$$\frac{a}{b} < \frac{a+c}{b+d} < \frac{c}{d}.$$

Neighbouring fractions at any level of the tree have the property: $\frac{a}{b} < \frac{c}{d}$ are neighbours, then $bc - ad = 1$!

If this property holds for the above neighbours, then go to the next level below on the tree and show that the property holds for the neighbours

$$\frac{a}{b} < \frac{a+c}{b+d} < \frac{c}{d}$$

This is clear!

The matrix $\begin{pmatrix} c & a \\ d & b \end{pmatrix} \in SL(2, \mathbf{Z})$.
The *unimodularity* also shows easily the inequalities

$$\frac{a}{b} < \frac{a+c}{b+d} < \frac{c}{d}.$$

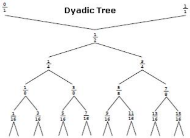
So, each row in the Farey tree is in increasing order and the tree is in bijective correspondence with the set of rationals in $[0, 1]$.

Just like the Farey tree, one has another binary tree of the same shape, known as the dyadic tree. In this tree, the top level has $\frac{0}{1}$ and $\frac{1}{1}$.

Just like the Farey tree, one has another binary tree of the same shape, known as the dyadic tree. In this tree, the top level has $\frac{0}{1}$ and $\frac{1}{1}$.

In general, at the n-th level, the fractions are

$$\frac{1}{2^{n-1}}, \frac{3}{2^{n-1}}, \frac{5}{2^{n-1}}, \cdots, \frac{2^{n-1}-1}{2^{n-1}}$$


Dyadic Tree

Minkowski's question mark function

Minkowski defined a function $?(x)$ from the Farey tree to the dyadic tree matching the labels.

Minkowski's question mark function

Minkowski defined a function $?(x)$ from the Farey tree to the dyadic tree matching the labels.

For instance, $?(1/3) = 1/4$. Due to the recursive definitions:

$$?(\frac{a+c}{b+d}) = \frac{?(\frac{a}{b} + \frac{c}{d})}{2}$$

Minkowski's question mark function

Minkowski defined a function $?(x)$ from the Farey tree to the dyadic tree matching the labels.

For instance, $?(1/3) = 1/4$. Due to the recursive definitions:

$$?(\frac{a+c}{b+d}) = \frac{?(\frac{a}{b} + \frac{c}{d})}{2}$$

As the rational numbers in $[0,1]$ as well as the dyadics in this interval are both dense subsets of $[0,1]$, the question mark function extends to a continuous function.

Minkowski's question mark function

Minkowski defined a function $?(x)$ from the Farey tree to the dyadic tree matching the labels.

For instance, $?(1/3) = 1/4$. Due to the recursive definitions:

$$?(\frac{a+c}{b+d}) = \frac{?(\frac{a}{b} + \frac{c}{d})}{2}$$

As the rational numbers in $[0,1]$ as well as the dyadics in this interval are both dense subsets of $[0,1]$, the question mark function extends to a continuous function.

$?(x)$ is a monotonically increasing function as the fractions were ordered. It is continuous everywhere but has zero derivative almost everywhere.

Using continued fractions, the question mark function can be defined more explicitly on any real number in $[0,1]$ as follows. Look at a continued fraction expansion:

$$x = [a_1, a_2, \cdots] := \frac{1}{a_1+}\frac{1}{a_2+}\frac{1}{a_3+}\cdots$$

Using continued fractions, the question mark function can be defined more explicitly on any real number in $[0,1]$ as follows. Look at a continued fraction expansion:

$$x = [a_1, a_2, \cdots] := \frac{1}{a_1+}\,\frac{1}{a_2+}\,\frac{1}{a_3+}\cdots$$

$$?(x) = \sum_{k=1}^{N} (-1)^{k-1} 2^{1-(a_1+a_2+\cdots+a_k)}$$

where N is the length of the continued fraction (this is infinite if x is irrational).

Using continued fractions, the question mark function can be defined more explicitly on any real number in $[0,1]$ as follows. Look at a continued fraction expansion:

$$x = [a_1, a_2, \cdots] := \frac{1}{a_1+}\frac{1}{a_2+}\frac{1}{a_3+}\cdots$$

$$?(x) = \sum_{k=1}^{N} (-1)^{k-1} 2^{1-(a_1+a_2+\cdots+a_k)}$$

where N is the length of the continued fraction (this is infinite if x is irrational).

The definition can be visualized in terms of the binary expansion of $?(x)$ as

$$?(x) = 0.\underbrace{0\cdots0}_{a_1-1}\underbrace{1\cdots1}_{a_2}\underbrace{0\cdots0}_{a_3}\underbrace{1\cdots1}_{a_4}\cdots$$

When N is finite, the expansion after the string of a_N zeroes (if N is odd) or a_N ones (if N is even) is a string of all ones (respectively, all zeroes).

Γ to answer questions

$$?(1-x) = 1 - ?(x).$$

$?(1-x) = 1-?(x)$.

From the explicit definition above, we also have

$$?\left(\frac{x}{1+x}\right) = \frac{?(x)}{2}.$$

$?(1-x) = 1-?(x)$.

From the explicit definition above, we also have

$$?(\frac{x}{1+x}) = \frac{?(x)}{2}.$$

This is because the continued fraction expansion of $\frac{x}{1+x}$ is $[a_1+1, a_2, \cdots]$.

Note that $x \mapsto 1-x$ and $x \mapsto \frac{x}{1+x}$ are fractional linear transformations and that Γ is the group of all fractional linear transformations.

The Γ generation

Let us show that the matrices $S = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ and $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ generate the whole of Γ.

The Γ generation

Let us show that the matrices $S = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ and $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ generate the whole of Γ.

Look at the elements $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ of Γ as fractional linear transformations $z \mapsto \frac{az+b}{cz+d}$.

Let us show that the matrices $S = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ and $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ generate the whole of Γ.

Look at the elements $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ of Γ as fractional linear transformations $z \mapsto \frac{az+b}{cz+d}$.

Notice that T acts as the translation $z \mapsto z+1$ and S acts as $z \mapsto -1/z$ (note that the minus sign ensures that we land again inside the upper half-plane).

For $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$,

observe that $\mathrm{Im}\ \gamma z = \frac{\mathrm{Im}\ z}{|cz+d|^2}$.

For $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$,
observe that $\mathrm{Im}\ \gamma z = \frac{\mathrm{Im}\ z}{|cz+d|^2}$.

Fix any point z in the upper half-plane.

For $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$,
observe that $\mathrm{Im}\ \gamma z = \frac{\mathrm{Im}\ z}{|cz+d|^2}$.

Fix any point z in the upper half-plane.

As c, d vary, there is a disc around the origin which contains no non-zero lattice point. Thus there is some γ in the subgroup of Γ generated by S, T such that $\mathrm{Im}\ \gamma z$ is maximal.

If necessary, we may replace γ by $T^j\gamma$ for some j, and assume that $-1/2 \leq \mathrm{Re}\ \gamma z \leq 1/2$.

If necessary, we may replace γ by $T^j\gamma$ for some j, and assume that $-1/2 \leq \mathrm{Re}\ \gamma z \leq 1/2$.

$$\mathrm{Im}\ S\gamma z = \frac{\mathrm{Im}\ \gamma z}{|\gamma z|^2} > \mathrm{Im}\ \gamma z \quad \text{if } |\gamma z| < 1.$$

If necessary, we may replace γ by $T^j\gamma$ for some j, and assume that $-1/2 \leq \mathrm{Re}\ \gamma z \leq 1/2$.

$$\mathrm{Im}\ S\gamma z = \frac{\mathrm{Im}\ \gamma z}{|\gamma z|^2} > \mathrm{Im}\ \gamma z \quad \text{if } |\gamma z| < 1.$$

This contradicts the choice of γ. Hence we have

$$|\gamma z| \geq 1.$$

Thus for each $z \in \mathbb{H}$, we have a $\gamma \in < S, T >$ so that

$$\frac{1}{2} \leq \mathrm{Re}\ \gamma z \leq \frac{1}{2} \qquad \text{and} \quad |\gamma z| \geq 1.$$

Thus for each $z \in \mathbb{H}$, we have a $\gamma \in < S, T >$ so that

$$\frac{1}{2} \leq \mathrm{Re}\ \gamma z \leq \frac{1}{2} \qquad \text{and} \quad |\gamma z| \geq 1.$$

This leads us to consider the closed region

$$\mathcal{F} = \left\{ z \in \mathbb{H} : \frac{1}{2} \leq \mathrm{Re}\ z \leq \frac{1}{2} \qquad \text{and} \quad |z| \geq 1 \right\}. \tag{1}$$

Thus for each $z \in \mathbb{H}$, we have a $\gamma \in < S, T >$ so that

$$\frac{1}{2} \leq \mathrm{Re}\ \gamma z \leq \frac{1}{2} \qquad \text{and} \quad |\gamma z| \geq 1.$$

This leads us to consider the closed region

$$\mathcal{F} = \left\{ z \in \mathbb{H} : \frac{1}{2} \leq \mathrm{Re}\ z \leq \frac{1}{2} \qquad \text{and} \quad |z| \geq 1 \right\}. \qquad (1)$$

Every point $z \in \mathbb{H}$ get transformed into an equivalent point $g(z) \in \mathcal{F}$ for some $g \in < S, T >$. If z_0 is an interior point of $\mathcal{F}$ and $\gamma \in \Gamma$, then get $g \in < S, T >$ with $g^{-1}\gamma(z_0) \in \mathcal{F}$. This shows $g = \gamma$ (so $\Gamma = < S, T >$) by the argument below.

We prove that if $z_1, z_2 \in \mathcal{F}$ are Γ-equivalent then they are on the boundary. We shall prove more.

We prove that if $z_1, z_2 \in \mathcal{F}$ are Γ-equivalent then they are on the boundary. We shall prove more.

Let $\mathrm{Im}\ z_2 \geq \mathrm{Im}\ z_1$. Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$ be such that $z_2 = Az_1$.
Then $1 \geq |cz_1 + d| \geq c^2 + d^2 - cd$ which gives either $c = \pm 1, d = 0$ or $d = 1, c = 0$ or $d = c = \pm 1$.

We prove that if $z_1, z_2 \in \mathcal{F}$ are Γ-equivalent then they are on the boundary. We shall prove more.

Let $\mathrm{Im}\ z_2 \geq \mathrm{Im}\ z_1$. Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$ be such that $z_2 = Az_1$.
Then $1 \geq |cz_1 + d| \geq c^2 + d^2 - cd$ which gives either $c = \pm 1, d = 0$ or $d = 1, c = 0$ or $d = c = \pm 1$.

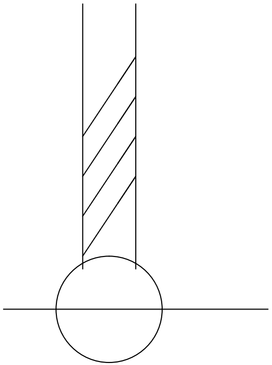
This forces not only z_1 to be on the boundary but also that $A = T, S, TS$ or ST.

Hence, we see that $\mathcal{F}$ is a closed region in $\mathcal{H}$ satisfying:
(i) Each $z \in \mathcal{H}$ is Γ-equivalent to a point in $\mathcal{F}$.
(ii) No two interior (distinct) points are Γ-equivalent.

Hence, we see that $\mathcal{F}$ is a closed region in $\mathcal{H}$ satisfying:
(i) Each $z \in \mathcal{H}$ is Γ-equivalent to a point in $\mathcal{F}$.
(ii) No two interior (distinct) points are Γ-equivalent.

Such a set is called a *fundamental domain*. Since the matrices S and T generate $SL_2(\mathbf{Z})$, we have thus constructed a fundamental domain for $SL_2(\mathbf{Z})$ in $\mathcal{H}$.

The finding of a fundamental domain goes under the name of reduction theory for $SL(2, \mathbf{Z})$. What we have done amounts to finding a complement to $SL(2, \mathbf{Z})$ in $SL(2, \mathbf{R})$.



Group structure from fundamental domain

Fundamental domains can be very useful in many ways; for example, they give even a presentation for the group.

Group structure from fundamental domain

Fundamental domains can be very useful in many ways; for example, they give even a presentation for the group.

Indeed, the above fundamental domain gives the presentation $< x, y | x^2, y^3 >$ for the group $PSL(2, \mathbf{Z}) = SL(2, \mathbf{Z})/\{\pm I\}$; that is, $PSL(2, \mathbf{Z})$ is a free product of cyclic groups of orders 2 and 3. The modular group $SL(2, \mathbf{Z})$ itself is thus an amalgamated free product of cyclic groups of orders 4 and 6 amalgamated along a subgroup of order 2.

A fundamental domain is written in terms of the so-called Iwasawa decomposition of $SL(2, \mathbf{R})$. The latter is simply a statement from linear algebra - the Gram-Schmidt process.

A fundamental domain is written in terms of the so-called Iwasawa decomposition of $SL(2,\mathbf{R})$. The latter is simply a statement from linear algebra - the Gram-Schmidt process.

For any $g \in GL(2,\mathbf{R})$, the canonical basis vectors e_1, e_2 for $\mathbf{R}^2$ are carried to another basis $\{ge_1, ge_2\}$. The Gram-Schmidt process reduces to a basis ke_1, ke_2 of orthonormal vectors. As the process amounts to multiplying by an invertible upper triangular matrix (which can be written as a product of a diagonal one and a matrix of the form $\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}$), we have a unique decomposition of $g \in GL(2,\mathbf{R})$ as a product kan where k is an orthogonal matrix.

One has $SL(2,\mathbf{R}) = KAN$ in the same way where K is the 'special orthogonal group' of rotation matrices $\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$,
$A = \{diag(a, a^{-1}) : a \in \mathbf{R}^*\}$, $N = \{\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} : x \in \mathbf{R}\}$.

One has $SL(2,\mathbf{R}) = KAN$ in the same way where K is the 'special orthogonal group' of rotation matrices $\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$,
$A = \{diag(a, a^{-1}) : a \in \mathbf{R}^*\}$, $N = \{\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} : x \in \mathbf{R}\}$.

The reduction theory for $SL(2,\mathbf{Z})$ says
$SL(2,\mathbf{R}) = KA_{\frac{2}{\sqrt{3}}} N_{\frac{1}{2}} SL(2,\mathbf{Z})$.
Here $A_t = \{diag(a_1, a_2) \in SL(2,\mathbf{R}) : a_i > 0 \text{ and } \frac{a_1}{a_2} \leq t\}$ and
$N_u = \{\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \in N : |x| \leq u\}$.

Modular group and quadratic forms

Consider a positive definite, binary quadratic form $f(x,y) = ax^2 + bxy + cy^2$ with $a, b, c \in \mathbf{Z}$; it takes only positive values except when $x = y = 0$.

Consider a positive definite, binary quadratic form $f(x, y) = ax^2 + bxy + cy^2$ with $a, b, c \in \mathbf{Z}$; it takes only positive values except when $x = y = 0$.

Two forms f and g are said to be equivalent if $\exists\ A = \begin{pmatrix} p & q \\ r & s \end{pmatrix} \in SL(2, \mathbf{Z})$ such that $f(x, y) = g(px + qy, rx + sy)$.

Modular group and quadratic forms

Consider a positive definite, binary quadratic form $f(x,y) = ax^2 + bxy + cy^2$ with $a, b, c \in \mathbf{Z}$; it takes only positive values except when $x = y = 0$.

Two forms f and g are said to be equivalent if $\exists\ A = \begin{pmatrix} p & q \\ r & s \end{pmatrix} \in SL(2, \mathbf{Z})$ such that $f(x,y) = g(px+qy, rx+sy)$.

Obviously, equivalent forms represent the same values. Indeed, this is the reason for the definition of equivalence.

One defines the discriminant of f to be $\mathrm{disc}(f) = b^2 - 4ac$.
Note that if f is positive-definite, the discriminant D must be < 0 because $4a(ax^2 + bxy + cy^2) = (2ax + by)^2 - Dy^2$ represents positive as well as negative numbers if $D > 0$.

One defines the discriminant of f to be $\mathrm{disc}(f) = b^2 - 4ac$. Note that if f is positive-definite, the discriminant D must be < 0 because $4a(ax^2 + bxy + cy^2) = (2ax + by)^2 - Dy^2$ represents positive as well as negative numbers if $D > 0$.

Further, f is said to be primitive if $(a, b, c) = 1$.

A primitive, positive definite, binary quadratic form $f(x,y) = ax^2 + bxy + cy^2$ is said to be *reduced* if $|b| \leq a \leq c$ and $b \geq 0$ if either $a = c$ or $|b| = a$.
These clearly imply

$$0 < a \leq \sqrt{\frac{|D|}{3}}.$$

A primitive, positive definite, binary quadratic form $f(x, y) = ax^2 + bxy + cy^2$ is said to be *reduced* if $|b| \leq a \leq c$ and $b \geq 0$ if either $a = c$ or $|b| = a$.
These clearly imply

$$0 < a \leq \sqrt{\frac{|D|}{3}}.$$

For example, the only reduced form of discriminant $D = -4$ is $x^2 + y^2$.
The only two reduced forms of discriminant $D = -20$ are $x^2 + 5y^2$ and $2x^2 + 2xy + 3y^2$.

$GL(2, \mathbf{R})$ acts on the space S of $+$ve-definite, binary quadratic forms as follows:
Each $P \in S$ can be represented by a $+$ve-definite, symmetric matrix; the corresponding form is $p_{11}x^2 + 2p_{12}xy + p_{22}y^2$.

$GL(2,\mathbf{R})$ acts on the space S of $+$ve-definite, binary quadratic forms as follows:
Each $P \in S$ can be represented by a $+$ve-definite, symmetric matrix; the corresponding form is $p_{11}x^2 + 2p_{12}xy + p_{22}y^2$.

The action of $g \in GL(2,\mathbf{R})$ takes P to ${}^t gPg \in S$.
This action is transitive and the isotropy at $I \in S$ is $O(2)$. In other words, S can be identified with $GL(2,\mathbf{R})/O(2)$ i.e.
$S = \{{}^t gg : g \in GL(2,\mathbf{R})\}$.

$GL(2,\mathbf{R})$ acts on the space S of $+$ve-definite, binary quadratic forms as follows:
Each $P \in S$ can be represented by a $+$ve-definite, symmetric matrix; the corresponding form is $p_{11}x^2 + 2p_{12}xy + p_{22}y^2$.

The action of $g \in GL(2,\mathbf{R})$ takes P to ${}^t gPg \in S$.
This action is transitive and the isotropy at $I \in S$ is $O(2)$. In other words, S can be identified with $GL(2,\mathbf{R})/O(2)$ i.e.
$S = \{{}^t gg : g \in GL(2,\mathbf{R})\}$.

In general, this works for $+$ve-definite quadratic forms in n variables.

The reduction theory for $SL(2, \mathbf{Z})$ shows that each $+$ve definite, binary quadratic form is equivalent to a unique reduced form.

The reduction theory for $SL(2, \mathbf{Z})$ shows that each +ve definite, binary quadratic form is equivalent to a unique reduced form.

Indeed, writing $f = {}^t gg$ and $g = kan\gamma$, ${}^t gg = {}^t\gamma {}^t na^2 n\gamma$ with $u \in U_{1/2}$ and $a^2 \in A_{4/3}$.
So ${}^t ua^2 u$ is a reduced form equivalent to f.

The reduction theory for $SL(2, \mathbf{Z})$ shows that each $+$ve definite, binary quadratic form is equivalent to a unique reduced form.

Indeed, writing $f = {}^t gg$ and $g = kan\gamma$, ${}^t gg = {}^t\gamma {}^t na^2 n\gamma$ with $u \in U_{1/2}$ and $a^2 \in A_{4/3}$.
So ${}^t ua^2 u$ is a reduced form equivalent to f.

$u = \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}$ with $|n| \leq 1/2$ and $a = diag(a_1, a_2)$ with $a_1/a_1 \leq 2/\sqrt{3}$, we have:
${}^t ua^2 u = \begin{pmatrix} a_1^2 & a_1^2 n \\ a_1^2 n & a_1^2 n + a_2^2 \end{pmatrix}$.
So, the corresponding form is $a_1^2 x^2 + 2a_1^2 nxy + (a_1^2 n + a_2^2) y^2$.
This is reduced because, if $0 < n \leq 1/2$ (the only non-obvious case)

$$2a_1^2 n \leq a_1^2 \leq a_1^2 n + a_2^2$$

the last inequality following as $a_1^2/a_2^2 \leq 1/(1 - n^2) \leq 4/3$.

To see an application, let us prove a beautiful discovery of Fermat, viz., that any prime number $p \equiv 1 \bmod 4$ is expressible as a sum of two squares.

To see an application, let us prove a beautiful discovery of Fermat, viz., that any prime number $p \equiv 1 \bmod 4$ is expressible as a sum of two squares.

Since $(p-1)! \equiv -1 \bmod p$ and since $(p-1)/2$ is even, it follows that $(\frac{p-1}{2}!)^2 \equiv -1 \bmod p$ i.e.,

$$((\frac{p-1}{2})!)^2 + 1 = pq$$

for some natural number q.

To see an application, let us prove a beautiful discovery of Fermat, viz., that any prime number $p \equiv 1 \bmod 4$ is expressible as a sum of two squares.

Since $(p-1)! \equiv -1 \bmod p$ and since $(p-1)/2$ is even, it follows that $(\frac{p-1}{2}!)^2 \equiv -1 \bmod p$ i.e.,

$$((\frac{p-1}{2})!)^2 + 1 = pq$$

for some natural number q.

The form $px^2 + 2(\frac{p-1}{2})!xy + qy^2$ is $+$ve definite and has discriminant -4 and must be equivalent to the reduced form $x^2 + y^2$.
As the former form has p as the value at $(1,0)$, the latter also takes the value p for some integers x, y.

In fact, reduction theory can also be used to show :
For any $D < 0$, there are only finitely many classes of primitive, positive-definite forms of discriminant D.
The number of classes alluded to is the class number $h(D)$ of the field $\mathbf{Q}(\sqrt{D})$; an isomorphism is obtained by sending $f(x,y)$ to the ideal $a\mathbf{Z} + \frac{-b+\sqrt{D}}{2}\mathbf{Z}$.

When we study some number-theoretic sequence (the same thing as an arithmetic function), it is often useful to look at the generating function which encodes the sequence. From the analytic or algebraic properties of the generating function, one can often draw number-theoretic conclusions.

Modular forms generate arithmetic functions

When we study some number-theoretic sequence (the same thing as an arithmetic function), it is often useful to look at the generating function which encodes the sequence. From the analytic or algebraic properties of the generating function, one can often draw number-theoretic conclusions.

The functions $\sigma_r(n) := \sum_{d|n} d^r$ and $r_k(n) =$ number of ways of writing n as a sum of k squares, have nice generating functions which are studied very conveniently with the help of our Γ.

The theta function is defined as
$\theta(z) = \sum_{n \in \mathbf{Z}} e^{i\pi n^2 z}$ for $\Re(z) > 0$.
So, $\theta(z)^k = \sum_n r_k(n) e^{i\pi n z}$.

The theta function is defined as
$\theta(z) = \sum_{n \in \mathbf{Z}} e^{i\pi n^2 z}$ for $\Re(z) > 0$.
So, $\theta(z)^k = \sum_n r_k(n) e^{i\pi n z}$.

The theta function has nice transformation properties with respect to the changes $z \mapsto \frac{az+b}{cz+d}$ for $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$.
This leads to beautiful expressions such as:

The theta function is defined as
$\theta(z) = \sum_{n \in \mathbf{Z}} e^{i\pi n^2 z}$ for $\Re(z) > 0$.
So, $\theta(z)^k = \sum_n r_k(n) e^{i\pi n z}$.

The theta function has nice transformation properties with respect to the changes $z \mapsto \frac{az+b}{cz+d}$ for $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$.
This leads to beautiful expressions such as:

$$r_4(n) = 8 \sum \{d : d|n, 4 \not| d\}$$

$$r_8(n) = 16 \sum_{d|n} (-1)^{n+d} d^3$$

We shall return to this later.

Sum of divisors via Eisenstein

Let $G_k(z) = \sum'_{c,d}(cz+d)^{-2k}$ where the sum is over integer pairs $(c,d) \neq (0,0)$ and $k \geq 2$ is an integer.
It is easy to show that the series $\sum'_{c,d}(cz+d)^{-\alpha}$ converges when $\alpha > 2$.

Sum of divisors via Eisenstein

Let $G_k(z) = \sum'_{c,d}(cz+d)^{-2k}$ where the sum is over integer pairs $(c,d) \neq (0,0)$ and $k \geq 2$ is an integer.
It is easy to show that the series $\sum'_{c,d}(cz+d)^{-\alpha}$ converges when $\alpha > 2$.

We prove now that

$$G_k(z) = 2\zeta(2k) + \frac{2(-2\pi)^k}{(2k-1)!} \sum_{d \geq 1} \sigma_{2k-1}(d) e^{2id\pi z}$$

where $\zeta(l) = \sum_{n \geq 1} \frac{1}{n^l}$ for $l > 1$.

We have $\pi \cot \pi z = \lim_{m \to \infty} \sum_{n=-m}^{m} \frac{1}{z+n}$.

We have $\pi \cot \pi z = \lim_{m\to\infty} \sum_{n=-m}^{m} \frac{1}{z+n}$.

One may deduce that

$$\frac{d^{r-1}}{dz^{r-1}}(\pi \cot \pi z) = (-1)^{r-1}(r-1)! \sum_{n\in\mathbf{Z}} \frac{1}{(z+n)^r}.$$

We have $\pi \cot \pi z = \lim_{m\to\infty} \sum_{n=-m}^{m} \frac{1}{z+n}$.

One may deduce that

$$\frac{d^{r-1}}{dz^{r-1}}(\pi \cot \pi z) = (-1)^{r-1}(r-1)! \sum_{n\in \mathbf{Z}} \frac{1}{(z+n)^r}.$$

$$\pi \cot \pi z = -i\pi \frac{1+e^{2i\pi z}}{1-e^{2i\pi z}} = -i\pi(1+2\sum_{d\geq 1} e^{2id\pi z})$$

can be differentiated term by term to give

$$\frac{d^{r-1}}{dz^{r-1}}(\pi \cot \pi z = (-2i\pi)^r \sum_{d\geq 1} d^{r-1} e^{2id\pi z}.$$

Comparing the two expressions we have:

$$\sum_{n\in\mathbf{Z}} \frac{1}{(z+n)^r} = \frac{(-2i\pi)^r}{(r-1)!} \sum_{d\geq 1} d^{r-1} e^{2id\pi z}.$$

Comparing the two expressions we have:

$$\sum_{n \in \mathbf{Z}} \frac{1}{(z+n)^r} = \frac{(-2i\pi)^r}{(r-1)!} \sum_{d \geq 1} d^{r-1} e^{2id\pi z}.$$

We break up the double series $\sum'_{c,d}(cz+d)^{-2k}$ for G_k into the three sums corresponding to $c = 0, c > 0, c < 0$.

Comparing the two expressions we have:

$$\sum_{n\in\mathbf{Z}} \frac{1}{(z+n)^r} = \frac{(-2i\pi)^r}{(r-1)!} \sum_{d\geq 1} d^{r-1} e^{2id\pi z}.$$

We break up the double series $\sum'_{c,d}(cz+d)^{-2k}$ for G_k into the three sums corresponding to $c=0, c>0, c<0$.

The first sum gives $2\zeta(2k)$ and the other two are equal as seen by putting $-c,-d$ in place of c,d.

$$G_k(z) = \sum_{c,d}{}' (cz+d)^{-2k} = 2\zeta(2k) + 2\sum_{c\geq 1}\sum_{d\in\mathbf{Z}} (cz+d)^{-2k}.$$

Using the expression (proved earlier)

$$\sum_{n \in \mathbf{Z}} \frac{1}{(z+n)^r} = \frac{(-2i\pi)^r}{(r-1)!} \sum_{d \geq 1} d^{r-1} e^{2id\pi z}$$

Using the expression (proved earlier)

$$\sum_{n\in\mathbf{Z}}\frac{1}{(z+n)^r}=\frac{(-2i\pi)^r}{(r-1)!}\sum_{d\geq 1}d^{r-1}e^{2id\pi z}$$

$$G_k(z)=2\zeta(2k)+\frac{2(-1)^k(2\pi)^{2k}}{(2k-1)!}\sum_{c\geq 1}\sum_{n\geq 1}n^{2k-1}e^{2i\pi ncz}$$

$$=2\zeta(2k)+\frac{2(-1)^k(2\pi)^{2k}}{(2k-1)!}\sum_{n}\sigma_{2k-1}(n)e^{2in\pi z}$$

One can also expand $z \cot z$ as a series $1 + \sum_{n \geq 1} \frac{B_{2n}(2z)^{2n}}{(2n)!}$ where B_k are the so-called Bernoulli numbers. One has

$$\zeta(2n) = \frac{B_{2n}(2\pi)^{2n}}{2(2n)!}$$

One can also expand $z \cot z$ as a series $1 + \sum_{n \geq 1} \frac{B_{2n}(2z)^{2n}}{(2n)!}$ where B_k are the so-called Bernoulli numbers. One has

$$\zeta(2n) = \frac{B_{2n}(2\pi)^{2n}}{2(2n)!}$$

We have the 'normalized Eisenstein series'

$$E_k(z) := \frac{1}{\zeta(2k)} G_k(z) = 1 + \frac{(-1)^k 4k}{B_{2k}} \sum_{n \geq 1} \sigma_{2k-1}(n) e^{2in\pi z}.$$

This normalized series has the property that it has the value 1 at 'infinity'.

One can also expand $z \cot z$ as a series $1 + \sum_{n \geq 1} \frac{B_{2n}(2z)^{2n}}{(2n)!}$ where B_k are the so-called Bernoulli numbers. One has

$$\zeta(2n) = \frac{B_{2n}(2\pi)^{2n}}{2(2n)!}$$

We have the 'normalized Eisenstein series'

$$E_k(z) := \frac{1}{\zeta(2k)} G_k(z) = 1 + \frac{(-1)^k 4k}{B_{2k}} \sum_{n \geq 1} \sigma_{2k-1}(n) e^{2in\pi z}.$$

This normalized series has the property that it has the value 1 at 'infinity'.

The Eisenstein series above are examples of modular forms. A word about why it is natural to study modular forms.

The fractional linear transformation $z \mapsto \frac{az+b}{cz+d}$ is invertible which means that its Jacobian (the amount by which the transformation distorts volumes) is non-zero everywhere.
A simple calculation shows that the Jacobian is $(cz+d)^{-2}$.

The fractional linear transformation $z \mapsto \frac{az+b}{cz+d}$ is invertible which means that its Jacobian (the amount by which the transformation distorts volumes) is non-zero everywhere.
A simple calculation shows that the Jacobian is $(cz+d)^{-2}$.

A function f for which $f((az+b)/(cz+d)) = (cz+d)^t f(z)$ (in particular, a modular form) has the property that the functions $z \mapsto f(z)$ and $z \mapsto f((az+b)/(cz+d))$ have the same zeroes and poles.

The fractional linear transformation $z \mapsto \frac{az+b}{cz+d}$ is invertible which means that its Jacobian (the amount by which the transformation distorts volumes) is non-zero everywhere.
A simple calculation shows that the Jacobian is $(cz+d)^{-2}$.

A function f for which $f((az+b)/(cz+d)) = (cz+d)^t f(z)$ (in particular, a modular form) has the property that the functions $z \mapsto f(z)$ and $z \mapsto f((az+b)/(cz+d))$ have the same zeroes and poles.

This condition is not as strong a condition as asking that f be invariant; that is, asking that $f((az+b)/(cz+d)) = f(z)$ and, hence it is more likely that one has several modular forms even if there were no invariant functions.

Fortunately, the modular forms of a given weight and the subspace of cusp forms (those which 'vanish at cusps') are form a finite-dimensional vector spaces. In fact, in some weights, there are no non-zero cusp forms.

Fortunately, the modular forms of a given weight and the subspace of cusp forms (those which 'vanish at cusps') are form a finite-dimensional vector spaces. In fact, in some weights, there are no non-zero cusp forms.

This gives relationships involving different Eisenstein series. For instance, the equalities $E_2^2 = E_4$, $E_2 E_3 = E_5$ follow from the fact that there are no cusp forms of weights $4, 10$.
They imply:

$$\sigma_7(n) = \sigma_3(n) + 120 \sum_{m=1}^{n-1} \sigma_3(m)\sigma_3(n-m)$$

$$11\sigma_9(n) = 21\sigma_5(n) - 10\sigma_3(n) + 5040 \sum_{m=1}^{n-1} \sigma_3(m)\sigma_5(n-m)$$

Sums of squares via generalized Eisenstein

The Eisenstein series above are not quite enough to study the 'sum of squares' function as this requires a generalized form which are modular forms not for the whole of Γ but for a slightly smaller group.

Sums of squares via generalized Eisenstein

The Eisenstein series above are not quite enough to study the 'sum of squares' function as this requires a generalized form which are modular forms not for the whole of Γ but for a slightly smaller group.

For a natural number q, look at the group $\Gamma(q)$ of those elements of Γ which look like the identity matrix when their entries are considered modulo q. This is known as the principal congruence subgroup of level q.

Sums of squares via generalized Eisenstein

The Eisenstein series above are not quite enough to study the 'sum of squares' function as this requires a generalized form which are modular forms not for the whole of Γ but for a slightly smaller group.

For a natural number q, look at the group $\Gamma(q)$ of those elements of Γ which look like the identity matrix when their entries are considered modulo q. This is known as the principal congruence subgroup of level q.

Define the general Eisenstein series
$G_k(z : c, d, q) := \sum (mz + n)^{-2k}$ where the sum is over
$(m, n) \neq (0, 0)$ such that $(m, n) \equiv (c, d) \bmod q$.
Define the restricted Eisenstein series
$G_k^*(z : c, d, q) := \sum (mz + n)^{-2k}$ where the sum is over
$GCD(m, n) = 1$ such that $(m, n) \equiv (c, d) \bmod q$.

These series are related as follows.

These series are related as follows.

$$G_k(z : c, d, q) = \sum_{(a,q)=1} \sum_{(m,n)\equiv(c,d);GCD(m,n)=a} (mz+n)^{-2k}$$

$$= \sum_{(a,q)=1} \frac{1}{a^{2k}} G_k^*(z : a^{-1}c, a^{-1}d, q)$$

$$= \sum_{(a,q)=1,a\leq q} \Big(\sum_{na\equiv 1(q)} \frac{1}{n^{2k}}\Big) G_k^*(z : ac, ad, q).$$

These series are related as follows.

$$G_k(z:c,d,q) = \sum_{(a,q)=1} \sum_{(m,n)\equiv(c,d);GCD(m,n)=a} (mz+n)^{-2k}$$

$$= \sum_{(a,q)=1} \frac{1}{a^{2k}} G_k^*(z:a^{-1}c, a^{-1}d, q)$$

$$= \sum_{(a,q)=1,a\leq q} \Big(\sum_{na\equiv 1(q)} \frac{1}{n^{2k}}\Big) G_k^*(z:ac, ad, q).$$

Using Mobius inversion, we get

$$G_k^*(z:c,d,q) = \sum_{(a,q)=1} \frac{\mu(a)}{a^{2k}} G_k(z:a^{-1}c, a^{-1}d, q)$$

$$= \sum_{(a,q)=1,a\leq q} \Big(\sum_{na\equiv 1(q)} \frac{\mu(n)}{n^{2k}}\Big) G_k^*(z:ac, ad, q).$$

Similar to the way we obtained the Fourier expansion of Eisenstein series, we get

$$G_k(z : c, d, q) = \sum_{n \equiv c(q)} \frac{1}{n^{2k}} + \frac{(-2\pi)^k}{q^{2k}(2k-1)!} \sum_r \sigma_{2k-1}(r, c, d) e^{2ir\pi z/q}$$

where the first term is not there if q does not divide c and $\sigma_l(r, c, d) = \sum_{u|r, r/u \equiv c(q)} u^l \cos(2i\pi ud/q)$.

Now, we just say a final word about the sums of squares function. The function $r_{4k}(n)$ is obtained from $\theta(z)^{4k}$ which is a modular form for $\Gamma(2)$ of weight k.

Now, we just say a final word about the sums of squares function. The function $r_{4k}(n)$ is obtained from $\theta(z)^{4k}$ which is a modular form for $\Gamma(2)$ of weight k.

Determining the values of $G_k^*(z;0,1,2)$ and $G_k^*(z;1,0,2)$ at the cusps of $\Gamma(2)$, one obtains the fact that

$$\theta(z)^{4k} - G_k^*(z;0,1,2) - (-1)^k G_k^*(z;1,0,2)$$

vanishes at all the cusps.

Now, we just say a final word about the sums of squares function. The function $r_{4k}(n)$ is obtained from $\theta(z)^{4k}$ which is a modular form for $\Gamma(2)$ of weight k.

Determining the values of $G_k^*(z;0,1,2)$ and $G_k^*(z;1,0,2)$ at the cusps of $\Gamma(2)$, one obtains the fact that

$$\theta(z)^{4k} - G_k^*(z;0,1,2) - (-1)^k G_k^*(z;1,0,2)$$

vanishes at all the cusps.

For $k = 2$, one knows that there are no such non-zero functions which means that we have an equality

$$\theta(z)^8 = G_2^*(z;0,1,2) - G_2^*(z;1,0,2)$$

From this, one obtains $r_8(n) = 16\sum_{d|n}(-1)^{n+d}d^3$.

For general k, the above argument gives $r_{4k}(n)$

$$= \frac{4k}{(2^{2k}-1)B_{2k}}\Big(\sum_{d|n,n/d=1(2)} d^{2k-1} + (-1)^k \sum_{d|n,n/d=0(2)} (-1)^d d^{2k-1}\Big) + a_n$$

where $a_n = O(n^k)$ is the n-th Fourier coefficient of a cusp form and B_{2k} is the Bernoulli number.

The modular function

$$G_2(\tau) = 60 \sum_{m,n}{}' \frac{1}{(m+n\tau)^4} \left(= \frac{(2\pi)^4}{12} \left(1 + \sum_{n=1}^{\infty} \sigma_3(n) e^{2\pi i n\tau} \right) \right)$$

$$G_3(\tau) = 140 \sum_{m,n}{}' \frac{1}{(m+n\tau)^6} \left(= \frac{(2\pi)^6}{12} \left(1 + \sum_{n=1}^{\infty} \sigma_5(n) e^{2\pi i n\tau} \right) \right).$$

The modular function

$$G_2(\tau) = 60 \sum_{m,n}{}' \frac{1}{(m+n\tau)^4} \left(= \frac{(2\pi)^4}{12} \left(1 + \sum_{n=1}^{\infty} \sigma_3(n) e^{2\pi i n \tau} \right) \right)$$

$$G_3(\tau) = 140 \sum_{m,n}{}' \frac{1}{(m+n\tau)^6} \left(= \frac{(2\pi)^6}{12} \left(1 + \sum_{n=1}^{\infty} \sigma_5(n) e^{2\pi i n \tau} \right) \right).$$

$\mathfrak{p}'(z)^2 = 4\mathfrak{p}(z)^3 - G_2(\tau)\mathfrak{p}(z) - G_3(\tau)$ where the Weierstrass $\mathfrak{p}$-function on $\mathbf{Z} + \mathbf{Z}\tau$ is the doubly periodic meromorphic function given by $\mathfrak{p}(z) = \frac{1}{z^2} + \sum_w (\frac{1}{(z-w)^2} - \frac{1}{w^2})$.

It can be shown that $\Delta(\tau) \stackrel{d}{=} G_2(\tau)^3 - 27G_3(\tau)^2 \neq 0$ for any τ on the upper half-plane.

It can be shown that $\Delta(\tau) \stackrel{d}{=} G_2(\tau)^3 - 27G_3(\tau)^2 \neq 0$ for any τ on the upper half-plane.

The elliptic modular function $j : \mathbf{H} \to \mathbf{C}$ is defined by

$$j(\tau) = 12^3 \cdot \frac{G_2(\tau)^3}{\Delta(\tau)}.$$

It can be shown that $\Delta(\tau) \stackrel{d}{=} G_2(\tau)^3 - 27G_3(\tau)^2 \neq 0$ for any τ on the upper half-plane.

The elliptic modular function $j : \mathbf{H} \to \mathbf{C}$ is defined by

$$j(\tau) = 12^3 \cdot \frac{G_2(\tau)^3}{\Delta(\tau)}.$$

The adjective 'modular' accompanies the j-function because of the invariance property:

$$j(\tau) = j(\tau') \Leftrightarrow \tau' \in SL(2,\mathbf{Z})(\tau) \stackrel{d}{=} \left\{ \frac{a\tau + b}{c\tau + d} : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2,\mathbf{Z}) \right\}.$$

Theorem
(i) j is holomorphic on $\mathbf{H}$.
(ii) j has the invariance property above.
(iii) $j : \mathbf{H} \to \mathbf{C}$ is onto.

Theorem
(i) j is holomorphic on $\mathbf{H}$.
(ii) j has the invariance property above.
(iii) $j : \mathbf{H} \to \mathbf{C}$ is onto.

The proof of (iii) needs the fundamental domain of $SL(2, \mathbf{Z})$ we referred to earlier.

Theorem
(i) j is holomorphic on $\mathbf{H}$.
(ii) j has the invariance property above.
(iii) $j : \mathbf{H} \rightarrow \mathbf{C}$ is onto.

The proof of (iii) needs the fundamental domain of $SL(2, \mathbf{Z})$ we referred to earlier.

That fact that $\mathfrak{p}$ satisfies the equation

$$\mathfrak{p}'(z)^2 = 4\mathfrak{p}(z)^3 - G_2(\tau)\mathfrak{p}(z) - G_3(\tau)$$

implies, by the theorem, that the j-function, gives an isomorphism from the coset space$SL(2, \mathbf{Z})\backslash\mathbf{H}$ to the set all 'complex elliptic curves' $\mathbf{C}/\mathbf{Z} + \mathbf{Z}\tau$.

In fact, one has bijective correspondences between :
(i) lattices $L = \mathbf{Z} + \mathbf{Z}\tau \subset \mathbf{C}$ upto scalar multiplication,

In fact, one has bijective correspondences between :
(i) lattices $L = \mathbf{Z} + \mathbf{Z}\tau \subset \mathbf{C}$ upto scalar multiplication,

(ii) complex elliptic curves $\mathbf{C}/L$ upto isomorphism,

In fact, one has bijective correspondences between :
(i) lattices $L = \mathbf{Z} + \mathbf{Z}\tau \subset \mathbf{C}$ upto scalar multiplication,

(ii) complex elliptic curves $\mathbf{C}/L$ upto isomorphism,

(iii) the numbers $j(\tau)$, and

In fact, one has bijective correspondences between :
(i) lattices $L = \mathbf{Z} + \mathbf{Z}\tau \subset \mathbf{C}$ upto scalar multiplication,

(ii) complex elliptic curves $\mathbf{C}/L$ upto isomorphism,

(iii) the numbers $j(\tau)$, and

(iv) Riemann surfaces of genus 1 upto complex analytic isomorphism.

As a matter of fact, $SL(2, \mathbf{Z}) \backslash \mathbf{H}$ is the so-called (coarse) moduli space of elliptic curves over $\mathbf{C}$.

As a matter of fact, $SL(2, \mathbf{Z}) \backslash \mathbf{H}$ is the so-called (coarse) moduli space of elliptic curves over $\mathbf{C}$.

In general, various subgroups of $SL(2, \mathbf{Z})$ describe other moduli problems for elliptic curves. This description has been vastly exploited in modern number theory.

As a matter of fact, $SL(2, \mathbf{Z})\backslash\mathbf{H}$ is the so-called (coarse) moduli space of elliptic curves over $\mathbf{C}$.

In general, various subgroups of $SL(2, \mathbf{Z})$ describe other moduli problems for elliptic curves. This description has been vastly exploited in modern number theory.

For instance, complex spaces like $\Gamma_0(N)\backslash\mathbf{H}$ have algebraic models over $\mathbf{Q}$ called Shimura varieties.

The Taniyama-Shimura-Weil conjecture (which implies Fermat's Last Theorem) says that any elliptic curve over $\mathbf{Q}$ admits a surjective, algebraic map defined over $\mathbf{Q}$ from a projectivised model of $\Gamma_0(N)\backslash\mathbf{H}$ onto it.

The Taniyama-Shimura-Weil conjecture (which implies Fermat's Last Theorem) says that any elliptic curve over $\mathbf{Q}$ admits a surjective, algebraic map defined over $\mathbf{Q}$ from a projectivised model of $\Gamma_0(N)\backslash\mathbf{H}$ onto it.

The point of this is that functions on $\Gamma_0(N)\backslash\mathbf{H}$ or even on $SL(2,\mathbf{Z})\backslash\mathbf{H}$ with nice analytic properties are essentially modular forms and conjectures like Taniyama-Shimura-Weil say essentially that 'nice geometric objects over $\mathbf{Q}$ come from modular forms'.

As $j : \mathbf{H} \to \mathbf{C}$ is $SL(2, \mathbf{Z})$ - invariant, one has $j(\tau+1) = j(\tau)$. So $j(\tau)$ is a holomorphic function in the variable $q = e^{2\pi i \tau}$, in the region $0 < |q| < 1$.

Thus, $j(\tau) = \sum_{n=-\infty}^{\infty} c_n q^n$ is a Laurent expansion i.e., all but finitely many $c_n (n < 0)$ vanish.

As $j : \mathbf{H} \to \mathbf{C}$ is $SL(2, \mathbf{Z})$ - invariant, one has $j(\tau + 1) = j(\tau)$. So $j(\tau)$ is a holomorphic function in the variable $q = e^{2\pi i \tau}$, in the region $0 < |q| < 1$.

Thus, $j(\tau) = \sum_{n=-\infty}^{\infty} c_n q^n$ is a Laurent expansion i.e., all but finitely many $c_n (n < 0)$ vanish.

In fact, $j(\tau) = \frac{1}{q} + 744 + \sum_{n \geq 1} c_n q^n$ with $c_n \in \mathbf{Z}\ \forall\ n$.
$c_1 = 196884, c_2 = 21493760, c_3 = 864299970$ etc.

Complex multiplication

We defined the j-function on $\mathbf{H}$. One can think of j as a function on lattices $\mathbf{Z} + \mathbf{Z}\tau$.

Complex multiplication

We defined the j-function on $\mathbf{H}$. One can think of j as a function on lattices $\mathbf{Z} + \mathbf{Z}\tau$.

In particular, if $\mathcal{O}$ is an order in an imaginary quadratic field $\mathbf{Q}(\sqrt{-n})$, it can be viewed as a lattice in $\mathbf{C}$. In fact, any proper, fractional $\mathcal{O}$-ideal I can be 2-generated i.e, is a free $\mathbf{Z}$-module of rank 2 i.e., is a lattice in $\mathbf{C}$. Then, it makes sense to talk about $j(I)$.

Complex multiplication

We defined the j-function on $\mathbf{H}$. One can think of j as a function on lattices $\mathbf{Z} + \mathbf{Z}\tau$.

In particular, if $\mathcal{O}$ is an order in an imaginary quadratic field $\mathbf{Q}(\sqrt{-n})$, it can be viewed as a lattice in $\mathbf{C}$. In fact, any proper, fractional $\mathcal{O}$-ideal I can be 2-generated i.e, is a free $\mathbf{Z}$-module of rank 2 i.e., is a lattice in $\mathbf{C}$. Then, it makes sense to talk about $j(I)$.

$j(I)$ is an algebraic number of degree $\leq$ class number of $\mathcal{O}$.

The First main theorem of Complex multiplication :
Let $\mathcal{O}$ be an order in an imaginary quadratic field K. Let $I \subset \mathcal{O}$ be a factional $\mathcal{O}$-ideal. Then, $j(I)$ is an algebraic integer and $K(j(I))$ is the Hilbert (ring) class field of $\mathcal{O}$.

The First main theorem of Complex multiplication :
Let $\mathcal{O}$ be an order in an imaginary quadratic field K. Let $I \subset \mathcal{O}$ be a factional $\mathcal{O}$-ideal. Then, $j(I)$ is an algebraic integer and $K(j(I))$ is the Hilbert (ring) class field of $\mathcal{O}$.

For τ imaginary quadratic, it follows that $j(\tau)$ is an algebraic integer of degree $=$ class number of $\mathbf{Q}(\tau)$ i.e, $\exists$ integers $a_0, \ldots, a_{h-1}$ such that $j(\tau)^h + a_{h-1}j(\tau)^{h-1} + \ldots + a_0 = 0$.

The largest D such that $\mathbf{Q}(\sqrt{-D})$ has class number 1 is 163 (there are only finitely many such D).

The largest D such that $\mathbf{Q}(\sqrt{-D})$ has class number 1 is 163 (there are only finitely many such D).

The so-called ring of integers is $\mathbf{Z} + \mathbf{Z}(\frac{-1+i\sqrt{163}}{2})$; so $j(\frac{-1+i\sqrt{163}}{2}) \in \mathbf{Z}$.

But $j(\tau) = \frac{1}{q} + 744 + \sum\limits_{n \geq 1} c_n q^n$ with $c_n \in \mathbf{Z}$ and

$$q = e^{2\pi i(\frac{-1+i\sqrt{163}}{2})} = -e^{-\pi\sqrt{163}}.$$

But $j(\tau) = \frac{1}{q} + 744 + \sum_{n \geq 1} c_n q^n$ with $c_n \in \mathbf{Z}$ and

$$q = e^{2\pi i(\frac{-1+i\sqrt{163}}{2})} = -e^{-\pi\sqrt{163}}.$$

Thus
$-e^{\pi\sqrt{163}} + 744 - 196884\ e^{-\pi\sqrt{163}} + 21493760\ e^{-2\pi\sqrt{163}} + \ldots =$
$j(\tau) \in \mathbf{Z}$.

But $j(\tau) = \frac{1}{q} + 744 + \sum_{n \geq 1} c_n q^n$ with $c_n \in \mathbf{Z}$ and

$$q = e^{2\pi i(\frac{-1+i\sqrt{163}}{2})} = -e^{-\pi\sqrt{163}}.$$

Thus
$-e^{\pi\sqrt{163}} + 744 - 196884\ e^{-\pi\sqrt{163}} + 21493760\ e^{-2\pi\sqrt{163}} + \ldots =$
$j(\tau) \in \mathbf{Z}$.

In other words,

$$e^{\pi\sqrt{163}} - integer = 196884\ e^{-\pi\sqrt{163}} + 21493760\ e^{-2\pi\sqrt{163}} \ldots \approx 0!$$

A popular myth (with no basis whatsoever!) credits Ramanujan with the above assertion. Talking of Ramanujan, we may say:

Ramanujan did mathematics somehow;
we still can't figure out even now.
He left his mark on "p of n",
and wrote pi in series quite often.
The theta functions he called 'mock'
are subject-matter of many a talk.
He died very young - yes, he too!
He was only thirty-two!
His name prefixes the function tau.
Truly, that was his last bow!