The story of i
Transcript

Date: Wednesday, 7 February 2007 - 1:00PM
Location: Barnard's Inn Hall
7 February 2007

The Story of $i$

Professor Robin Wilson

Introduction

In my last lecture I outlined the story of $\pi$. This lecture concerns the number $i$, and next time I'll tell you about $e$, the exponential number, and bring the three numbers together in one of the most famous results in mathematics.

'The story of $i$ is not the title of my autobiography — rather, it concerns the square root of minus 1 and the so-called imaginary or complex numbers. But is there such a thing as $\sqrt{-1}$? — after all, if you square either 1 or -1 you get 1, so what can you possibly square to get -1?

The answer started to emerge in the 16th century, but even three centuries later, there was much anguish about the subject.

The Astronomer Royal, George Airy, said that he had not the smallest confidence in any result which is essentially obtained by the use of imaginary symbols, while Augustus De Morgan, Professor of Mathematics at University College, London, said We have shown the symbol $\sqrt{-1}$ to be void of meaning, or rather self-contradictory and absurd.

Furthermore, the great Euler, who made so many contributions to the development of complex numbers, said:

All such expressions as $\sqrt{-1}$, $\sqrt{-2}$, etc., are consequently impossible or imaginary numbers, since they represent roots of negative quantities; and of such numbers we may truly assert that they are neither nothing, nor greater than nothing, nor less than nothing, which necessarily constitutes them imaginary or impossible.

So what are these imaginary and impossible objects? How did they arise? And why did they cause so much confusion for several centuries?

Different types of number

In an earlier lecture I showed you how our usual number system is built up. As the 19th-century German mathematician Leopold Kronecker remarked:

'God created the natural numbers, and all the rest is the work of man.'

So, starting with the natural numbers, 1, 2, 3, ..., we then obtained all the integers — positive, negative and zero. This was not a trivial process, taking thousands of years, and negative numbers were treated with the same ridicule that the imaginary numbers would later face — after all, what is meant by 'minus 2 sheep'? These days we have no difficulty understanding negative temperatures in our weather forecasts, even if they leave us cold, and it seems hard to see why negative numbers caused so much disbelief.

The next step is to divide one integer by another, and we get fractions, or rational numbers. All we need to remember is not to divide by 0, and that different fractions can represent the same rational number: for example, $\frac{1}{2}$ is the same as $\frac{2}{4}$ or $-\frac{10}{-20}$.

But many numbers cannot be written as fractions — for example, $\pi$, $\sqrt{2}$, $3\sqrt{7}$, and the number $e$ that we'll meet next time. These are irrational numbers, which when combined with the rational numbers form the real numbers. In previous lectures I discussed such issues as how to define the real numbers, and how to show that there are far more irrational numbers than rational numbers, but today I'm heading in a different direction.

For many purposes the real numbers are all we need — but suppose that we do agree to allow this mysterious object called '$\sqrt{-1}$'. We can then form many more 'numbers' such as $3 - 4\sqrt{-1}$. Ignoring for the moment what this actually means, we can then carry out calculations with such objects.

Addition is easy:

$$(2 + 3\sqrt{-1}) + (4 + 5\sqrt{-1}) = (2 + 4) + (3 + 5)\sqrt{-1} = 6 + 8\sqrt{-1},$$

and so is multiplication (using $\sqrt{-1} \times \sqrt{-1} = -1$):

$$(2 + 3\sqrt{-1}) \times (4 + 5\sqrt{-1}) = (2 \times 4) + (3\sqrt{-1} \times 4) + (2 \times 5\sqrt{-1}) + (15 \times \sqrt{-1} \times \sqrt{-1})$$
In fact, we can carry out all the standard operations of arithmetic on these new objects. As mentioned earlier, we call the object \( a + bi \) a complex number. \( a \) is its real part, and \( b \) is its imaginary part. Later, we shall sometimes follow Leonard Euler who in 1777 introduced the letter \( i \) to mean \( \sqrt{-1} \), so that \( i^2 = -1 \).

**Solving equations**

Let's now go back to our various types of numbers and look at them from a different point of view. If we're restricted to the natural numbers, then we can solve certain equations – for example, the equation that we now write as \( x + 3 = 7 \) has the solution \( x = 4 \). But to solve the equation \( x + 7 = 3 \) we need to expand our number system to the negative integers, and the solution is \( x = -4 \). So we can now solve equations of the form \( x + a = b \), where \( a \) and \( b \) are any integers.

The next stage is to bring in fractions: using them we can now solve an equation such as \( 7x = 5 \); the solution is \( x = 5/7 \). We can also solve linear equations – those of the form \( ax = b \), where \( a \) and \( b \) are integers or rationals.

Once we have introduced real numbers, we can go beyond linear equations and find solutions for equations such as \( x^2 = 2 \) or \( x^3 = 7 \), or for the equation \( 4x - 10x^2 + 1 = 0 \), which has \( \sqrt{2} + \sqrt{3} \) as a solution. But even now we cannot solve all quadratic equations – to solve the equation \( x^2 = -1 \) we need to introduce another type of number, the square root of \( -1 \). Do we need to introduce anything else?

Let's look at three quadratic equations:

- For the quadratic equation \( x^2 - 4x + 3 = 0 \), we can factorize:
  \[ x^2 - 4x + 3 = (x - 3)(x - 1) = 0, \]
  so there are two solutions: \( x = 3 \) and \( x = 1 \);

- For the quadratic equation \( x^2 - 4x + 4 = 0 \), we can again factorize:
  \[ x^2 - 4x + 4 = (x - 2)(x - 2) = 0, \]
  so we have a repeated solution: \( x = 2 \);

- To factorize the quadratic equation \( x^2 - 4x + 5 = 0 \), we need to bring in \( \sqrt{-1} \):
  \[ x^2 - 4x + 5 = (x - 2 - \sqrt{-1})(x - 2 + \sqrt{-1}) = 0, \]
  so the two solutions are: \( x = 2 + \sqrt{-1} \) and \( x = 2 - \sqrt{-1} \), as we can easily check.

To explain the differences between these solutions let's recall the quadratic equation formula: if \( ax^2 + bx + c = 0 \), then
\[ x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}. \]

Taking \( a = 1 \) and \( b = -4 \), as in the examples above, we see that the equation \( x^2 - 4x + c = 0 \) has solutions
\[ x = \frac{1}{2}(4 \pm \sqrt{(16 - 4c)}) = 2 \pm \sqrt{4 - c}. \]
When \( c = 3 \), we have \( x = 2 \pm \sqrt{1} \), giving \( x = 3 \) or \( 1 \);
when \( c = 4 \), we have \( x = 2 \pm \sqrt{0} \), giving just \( x = 2 \);
when \( c = 5 \), we have \( x = 2 \pm \sqrt{-1} \).

If we plot the graphs of these three quadratics we find, as we would expect, that the first curve crosses the \( x \)-axis twice (at 3 and 1), the second just touches the \( x \)-axis (at 2), and the third (with complex solutions) misses the \( x \)-axis altogether.

**The Fundamental Theorem of Algebra**

What happens when we look at higher-degree equations, such as this one:
\[ x^6 - 12x^5 + 60x^4 - 160x^3 + 239x^2 - 188x + 60 = 0? \]
Can this be solved? If so, can it be done with just real and complex numbers, or do we need to introduce yet another type of number?

Around 1700 there was some discussion about what forms the solutions of these more complicated equations might take. By this time, equations of degrees 1, 2, 3 and 4 had been solved, as we’ll see, but what about equations of degree 5 or more, which no-one could solve in general? There seemed to be several scenarios:

- we can solve all equations using only real and complex numbers;
To get an idea of which of these is the case, let’s first try to take the square root of $i$. Do we need to introduce further numbers, or will our existing complex numbers be enough? If the latter, then we can write:

$$x^2 = (a + bi)^2 = i,$$ so $$a^2 - b^2 + 2abi = i.$$  
So $a^2 - b^2 = 0$ and $2ab = 1$, giving $a = b = \sqrt{1/2}$ or $a = b = -\sqrt{1/2}$, so that the solutions are $\pm \sqrt{1/2} (1 + i)$. So in this case, complex numbers are enough. 

In fact, complex numbers are always enough for any polynomial equation; for example, 

$$x^6 - 12x^5 + 60x^4 - 160x^3 + 239x^2 - 188x + 60$$
$$= (x^2 - 4x + 3) (x^2 - 4x + 4) (x^2 - 4x + 5)$$
$$= (x - 1) (x - 3) (x - 2)^2 (x - 2 + i) (x - 2 - i) = 0,$$  
so the solutions of the equation are 1, 3, 2 (twice) and 2 ± $i$.  

This is a special case of what became known as the Fundamental Theorem of Algebra. It can be stated in various ways:

* every polynomial $p(x)$ can be factorized into linear and quadratic polynomials with real coefficients; 
* every polynomial $p(x)$ can be factorized completely into linear factors with complex coefficients; 
* every polynomial equation $p(x) = 0$ of degree $n$ has at least one complex solution; 
* every polynomial equation $p(x) = 0$ of degree $n$ has exactly $n$ complex solutions (as long as we count them appropriately).

For a long time this result was folklore, but it seems to have been stated formally by Albert Girard in 1629, in the form ‘every equation of algebra has as many solutions as the exponent of the highest term indicates’, and others seemed aware of it — for example, Descartes also stated the result.  

But it was not until the eighteenth century that the question received any serious discussion — notably, by d’Alembert, Euler and Lagrange; d’Alembert simplified the problem by assuming that the solutions exist and he deduced that they must be complex — but his proof can be completed without too much difficulty. 

Carl Friedrich Gauss, dismissed these earlier efforts and gave the first ‘proof’ in his doctoral dissertation of 1799, but it too is deficient and less easy to patch up; Gauss realised this and later provided three corrected proofs — but the waters surrounding all these proofs is very murky, and it’s difficult to be sure who gave the first ‘rigorous’ proof.

**The origin of $\sqrt{-1}$**

In view of the above, it seems that $\sqrt{-1}$ should have made its first appearance in the solutions of quadratic equations; surprisingly, this is not the case. 

Let’s look briefly at some early attempts to solve equations. Here’s one in sexagesimal notation from a Mesopotamian clay tablet from around 1800 BC:  

*I have subtracted the side of my square from the area: 14,30. You write down 1, the coefficient. You break off half of 1. 0;30 and 0;30 you multiply. You add 0;15 to 14,30. Result 14,30;15. This is the square of 29;30. You add 0;30, which you multiplied, to 29;30. Result: 30, the side of the square.*

Putting this into modern algebraic notation, we obtain the quadratic equation $x^2 - x = 870$, and the above sequence of steps gives us successively: $1, \frac{1}{2}, (\frac{1}{2})^2 = \frac{1}{4}, 870\frac{1}{4}, 29\frac{1}{4}, 30$. 

This is just one of a dozen or more similar problems on the same clay tablet, and may have been used for teaching purposes. It turns out that if we carry out the same operations for the general equation $x^2 - bx = c$, we get the same result as we would get nowadays by applying the quadratic equation formula. Thus, 4000 years ago the Mesopotamians could solve particular quadratic equations by using essentially the same sequence of operations that we use today. 

However, there are differences. In particular, they seemed to be satisfied with just one solution: any suggestion that there
might be others didn’t arise — and in this case it would have been a negative number which would have been meaningless to them anyway.

A similar situation arose with the Islamic scholars of Baghdad in around 900 AD. In his book *Hitab al-jabr w’al muqabalah*, from which we derive our word *algebra*, Khwarizmi presents a lengthy account of how to solve quadratic equations. Since negative numbers were not considered meaningful, he split the equations into six types, corresponding (in modern notation) to the forms $ax^2 = bx$, $ax^2 = b$, $ax^2 + bx = c$, $ax^2 = bx + c$, where $a$, $b$, and $c$ are *positive* constants. He then proceeded to solve instances of each type, such as $x^2 + 10x = 39$, using a geometrical form of ‘completing the square’.

Again, their concern seemed to be only with whether a solution exists and how to find it. There are no discussions of any further solutions, and an equation such as $x^2 + 1 = 0$ would be considered as having no solutions, much as we now usually say that the equation $\sin x = 2$ has no solutions.

Earlier, in the first century AD, a certain amount of fudging was used when the square root of a negative quantity did turn up. In his *Stereometria*, Heron of Alexandria was trying to find the height $h$ of a frustrum of a pyramid (that is, a pyramid with its top chopped off), where the sides of the base and the top have lengths $a = 28$ and $b = 4$ units, and the slant edge-length $c$ is 15. The appropriate formula is

$$h = \sqrt{c^2 - 2 (\frac{1}{2} (a - b)^2)} = \sqrt{152 - 2 (122)} = \sqrt{225 - 288} = \sqrt{-63}.$$  

Since this was clearly far too dangerous to contemplate, it appeared in the *Stereometria* simply as $\sqrt{63}$.

As I mentioned earlier, the first discussion of square roots of negative quantities was not in connection with quadratic equations, but actually with cubic ones. I discussed these in one of my lectures last year, but let’s review them briefly.

In 1545 Gerolamo Cardano published his important book *Ars Magna* (The Great Art), in which he showed how to solve cubic and quartic equations — equations of degrees 3 and 4. Here’s his method, which works in general, for solving a particular cubic equation.

To solve $x^3 + 6x = 20$, we seek two other numbers $u$ and $v$ such that $u - v = 20$ and $uv = (1/3 \times 6)^3 = 8$. Since $v = u - 20$ we have $uv = u(u - 20) = u^2 - 20u = 8$. This is a quadratic equation which he easily solved to give $u = 10 + \sqrt{108}$, so that $v = -10 + \sqrt{108}$. The solution for $x$ then has the form $3\sqrt{u} - 3\sqrt{v}$, which is

$$x = 3\sqrt{(10 + \sqrt{108})} - 3\sqrt{(-10 + \sqrt{108})}.$$  

If you calculate this on a calculator, you’ll get the much simpler answer 2, which clearly satisfies the original equation. But Cardano seemed unable to perform the necessary simplifications, which are certainly unpleasant to carry out.

Cardano also asked how one can divide 10 into two parts whose product is 40. If the parts are taken to be $x$ and $10 - x$, then $x(10 - x) = 40$, and Cardano obtained the solutions 5 + $\sqrt{-15}$ and 5 – $\sqrt{-15}$. He could see no meaning to these, but observed ‘*Nevertheless we will operate, putting aside the mental tortures involved*’, and found that everything works out correctly:

$$(5 + \sqrt{-15}) \times (5 - \sqrt{-15}) = 52 - (\sqrt{-15})^2 = 25 - (-15) = 40.$$  

The situation was clarified by Rafael Bombelli, an engineer who was an expert in draining swampy marshes. In his *Algebra* of 1572 he applied Cardano’s method to the equation $x^3 = 15x + 4$ and found that

$$x = 3\sqrt{(2 + \sqrt{-121})} + 3\sqrt{(2 - \sqrt{-121}),}$$  

which involves these imaginary numbers — but this equation has three real roots, 4, –2 + $\sqrt{3}$ and –2 – $\sqrt{3}$, with no imaginary number in sight. Cardano himself was aware that this sort of thing could happen, and complained: ‘*So progresses arithmetic subtlety the end of which is as refined as it is useless.*’

Bombelli simplified the expression by looking for real numbers $a$ and $b$ such that

$$(a + bi)^3 = 2 + \sqrt{-121} \quad \text{and} \quad (a - bi)^3 = 2 - \sqrt{-121},$$  

so that he could take the cube roots. After some rearrangement he quickly found that $a = 2$ and $b = 1$:

$$2 + \sqrt{-121} = 2 + \sqrt{-121} \quad \text{and} \quad 2 - \sqrt{-121} = 2 - \sqrt{-121},$$  

so $x = (2 + \sqrt{-1}) + (2 - \sqrt{-1}) = 4$, as expected.

**Representing complex numbers geometrically**

How can we represent complex numbers geometrically? Descartes had employed various devices to draw a number of curves,
and had given geometrical constructions for the square root of a positive quantity and for the positive root of the quadratic equation \( x^2 = ax + b^2 \). Influenced by him, several other seventeenth-century mathematicians attempted to represent algebraic ideas geometrically.

In particular, John Wallis, Savilian Professor of Geometry in Oxford, constructed the square root of the product \( bc \), where \( b \) and \( c \) are both positive, by drawing a circle with diameter \( AC \) of length \( b + c \), and constructing a perpendicular from the point at distance \( b \) from \( A \); the length of this perpendicular is the required square root. Wallis then attempted to modify the process in order to construct the square root of \( bc \), when \( b \) is negative and \( c \) is positive. He also gave a construction that hinted at the idea of an imaginary number being at right angles to a real one, but didn’t quite get there.

### The complex plane

The answer was given by the Norwegian surveyor Caspar Wessel, but unfortunately his article (in Danish) was overlooked for a hundred years and his ideas had no influence on the development of the subject. In 1797 he presented a paper ‘On the analytic representation of direction: an attempt’ to the Royal Danish Academy of Sciences, in which he outlined the idea of what we now call the complex plane.

The idea is to regard each complex number \( a + bi \) as the point \((a, b)\) in the plane, or as the vector from the origin to this point. To add two complex numbers, we add their vectors using the parallelogram law. This corresponds to the addition law \((a + bi) + (c + di) = (a + c) + (b + d)i\).

We can also write each complex number as \((r, \theta)\) or as \(r(\cos \theta + i \sin \theta)\), in terms of the length \(r\) of the vector and the angle \(\theta\) that it makes with the \(x\)-axis. We can then multiply two complex numbers by multiplying the corresponding lengths and adding the corresponding angles. This corresponds to the multiplication law \((r, \theta) \times (s, \phi) = (rs, \theta + \phi)\).

It follows from this that multiplying by \(i\) corresponds to rotating anti-clockwise through 90° and that multiplying by \(i\) twice gives a rotation through 180°, corresponding to the rule \(i \times i = -1\).

Using these ideas, Wessel was able to calculate the powers and roots of complex numbers. For example, on taking \(r = 1\), he was able to obtain such results as

\[
\cos \theta + i \sin \theta)^3 = (\cos 3\theta + i \sin 3\theta)\]

These are special cases of a result known as De Moivre’s theorem.

Note that by Pythagoras’s Theorem, we can represent the unit circle with centre 0 and radius 1 by the simple equation \(|z| = 1\). We can then plot the various roots of 1:

- the square roots satisfy \( z^2 - 1 = (z - 1)(z + 1) = 0 \), and are \( z = 1 \) and \(-1\);
- the cubic roots satisfy \( z^3 - 1 = (z - 1)(z^2 + z + 1) = 0 \), and are \( z = 1 \) and \( \frac{1}{2}(-1 \pm \sqrt{3}i) \);
- the fourth roots satisfy \( z^4 - 1 = (z - 1)(z^3 + z^2 + z + 1) = 0 \), and are \( z = 1, -1, i \) and \(-i\);
- the sixth roots satisfy \( z^6 - 1 = (z - 1)(z^5 + z^4 + z^3 + z^2 + z + 1) = 0 \), and are \( 1, -1, \frac{1}{2}(1 \pm \sqrt{3}) \) and \( \frac{1}{2}(-1 \pm \sqrt{3}) \).

If we plot these, we find that they always lie on the circle at the corners of a regular polygon.

### Argand and Gauss

The complex plane is often called the Argand diagram, but this is not historically accurate, as we have seen. But the results were ‘in the air’ and were rediscovered more than once, in particular by the Swiss-born Jean-Robert Argand.

In 1806 Argand wrote an Essay on the geometrical interpretation of imaginary quantities, which he had privately printed for his friends without his name on the title page. He sent a copy to the famous mathematician Legendre who in turn sent it to a mathematician called François Français, who then died shortly after. Fortunately his brother Jacques, also a mathematician, was looking through François’s papers and, intrigued by these results, published his own paper on the subject, mentioning Legendre’s letter and inviting the originator of the ideas to make himself known. Again fortunately, Argand learned about this and did so.

In Germany the complex plane or Argand diagram is often called the Gaussian plane, and it was indeed Carl Friedrich Gauss who...
first gave the name of complex numbers. He had already been working on all these ideas for some years but never told anyone, claiming in 1812 that ‘I have in my papers many things for which I could perhaps lose the priority of publication, but you know, I prefer to, let things ripen.’ In 1831, however, he did commit himself on the subject and such was his reputation that the subject was given a great boost.  

In particular, he studied the number-theoretic properties of the so-called Gaussian integers $a + bi$, where $a$ and $b$ are integers. These objects behave surprisingly like the ordinary integers — you can factorize them into ‘primes’ in only one way, just like factorizing an ordinary whole number into primes.

### Hamilton

The idea of representing each complex numbers $x + iy$ as a point $(x, y)$ in the plane was developed by Sir William Rowan Hamilton, the Astronomer Royal of Ireland.

Even as late as the 1830s there was still a great deal of suspicion about complex numbers, and about so-called ‘imaginary’ numbers that don’t seem to exist. Hamilton diffused much of this suspicion by saying that the complex numbers $a + bi$ should be defined as pairs $(a, b)$ of real numbers, which we combine by using the following rules: $(a, b) + (c, d) = (a + c, b + d)$ and $(a, b) \times (c, d) = (ac - bd, ad + bc)$

[corresponding to the equations

$$(a + bi) + (c + di) = (a + c) + (b + d)i$$

and

$$(a + bi) \times (c + di) = (ac - bd) + (ad + bc)i.$$]

The pair $(a, 0)$ corresponds to the real number $a$, the pair $(0, 1)$ corresponds to the number $i$, and we have the equation $(0, 1) \times (0, 1) = (-1, 0)$ [corresponding to the equation $i \times i = -1$].

Hamilton then tried to extend his ideas to three dimensions. If the plane of points $(a, b)$ corresponds to complex numbers of the form $a + bi$, where $i^2 = -1$, then three-dimensional space ought to correspond to objects of the form $a + bi + cj$, where $i^2 = -1$. Certainly, addition works well:

$$(a + bi + cj) + (d + ei + fj) = (a + d) + (b + e)i + (c + f)j.$$  

But he couldn’t make multiplication work:

$$(a + bi + cj) \times (d + ei + fj) = (ad - be - cf) + (ae + bd)i + (af + cd)j + (bf + ce)k.$$  

This gives four terms, rather than three. How can we get rid of the last term?

We can’t let $ij = 0$, because then $0 = (ij)2 = ij( -1) = 1$.  

Hamilton tried everything, such as writing $ij = 1$ or $ij = -1$, but nothing seemed to work: in a letter to one of his sons he later wrote:

_Every morning, on my coming down to breakfast, your little brother William Edwin and yourself used to ask me, ‘Well Papa, can you multiply triples? Wherefore I was obliged to reply, with a shake of the head: ‘No, I can only add and subtract them’._

Hamilton struggled with the problem for fifteen years, until one day he took a walk along the canal:

_As I was walking with Lady Hamilton to Dublin, and came up to Brougham Bridge, I then and there felt the galvanic circuit of thought close; and the sparks which fell from it were the fundamental equations exactly as I have used them ever since._

I pulled out on the spot a pocket book and made an entry… it is fair to say that this was because I felt a problem to have been at that moment solved — an intellectual want relieved which had haunted me for at least fifteen years since.

What he had come up with was his quaternions: objects of the form $a + bi + cj + dk$, where $i^2 = j^2 = k^2 = ij = -1$.

In order to make multiplication work, he had to abandon the commutative law, in which we can multiply numbers either way round $(3 \times 4 = 4 \times 3)$. The rules that made these quaternions work are

$ij = k$, but $ji = -k$; $jk = i$, but $kj = -i$, $ki = j$, but $ik = -j$  
—or, more concisely, $i^2 = j^2 = k^2 = ijk = -1$.

Hamilton was so excited that he carved these basic rules on the bridge, and there is now a plaque to commemorate their discovery, and over the years the Irish Post Office has issued several stamps featuring Hamilton and his discovery.
Conclusion

To conclude, the complex number system in general, and quaternions in particular, have proved to be of enormous importance, both for their theoretical properties and also in their applications to physics and engineering, topics which I am not qualified to talk about.

Those of you in the know will also realize that I failed to talk about the work of Euler, and in particular one of the most important results of all, Euler's theorem. Since this result involves the number $e$, I decided not to mention it today, but it will feature prominently in my next lecture.