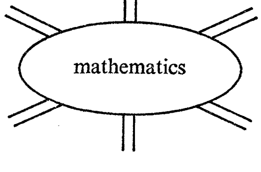
Proc. R. Soc. Lond. A. **347**, 291–299 (1976)
Printed in Great Britain

Bakerian Lecture, 1975
Global geometry

By M. F. Atiyah, F.R.S.
Mathematical Institute, Oxford

(*Lecture delivered* 12 *June* 1975 – *MS. received* 14 *July* 1975).

It must be many years since a pure mathematician was asked to address this society on such an occasion. Certainly none appears in the records going back to 1940. This is of course no reflection on the impartiality of the Council of the Royal Society, but simply an acknowledgement of the wide gap that separates pure mathematicians from other scientists and of the serious difficulties of communicating across that gap. Fortunately we have our intermediaries – the applied mathematicians – who extract from the body of mathematical knowledge the most useful parts and bring them to bear on recognizable scientific problems in a wide variety of fields, all the way from the traditional areas of physical science right through to the biological and social sciences. Many distinguished applied mathematicians have indeed spoken to the society on problems such as aerodynamic noise or stellar evolution, which are heavily dependent on mathematical analysis, but which can be explained in physical terms readily understood by a wide audience. On such occasions the mathematical techniques involved will quite properly have been relegated to decent obscurity. As a result scientists at large probably have only the vaguest ideas about mathematical research *per se*. They will understand mathematical work related to their particular field, frequently, I may add, better than the mathematicians themselves, but they must find it hard to visualize mathematics in the abstract. It is therefore perhaps worthwhile for a pure mathematician to attempt to explain how we view our subject and what motivates our research in the absence of any particular scientific interpretation or application. If I might summarize the situation diagrammatically, consider mathematics as some kind of giant computer with a large number of terminals on its periphery, representing fields of application.



A practising scientist is like the terminal user. He is primarily interested in the output and will know something about what the computer can do for him, but he is not involved in what goes on inside the heart of the computer. In the early days of computers, users and designers were frequently the same people, but with their rapid growth and sophistication this is now the exception rather than the rule. Similarly it is the increasing sophistication of mathematics which has led to the large gap between 'users' and 'designers'.

The immediate difficulty, when we separate mathematics from its physical interpretations, is that we are left in a sort of vacuum. Bertrand Russell, the only mathematician ever to win the Nobel Prize for Literature, made this point in his usual vivid and provocative style when he said that 'Mathematics may be defined as the subject in which we never know what we are talking about, nor whether what we are saying is true'. He went on to say that he hoped people would find comfort in this definition and would probably agree that it is accurate. I myself find it somewhat disturbing, and I would prefer to define mathematics rather more constructively. We all recognize the important rôle that *analogy* plays in scientific thought. When we think of molecules in a gas as small billiard balls, or light as made up of waves we are relating the unfamiliar to the familiar as an aid to conceptual understanding. When we take these analogies seriously and pursue their consequences we are developing mathematical models, in which the similarities are emphasized and the differences are ignored. Thus mathematics can I think be viewed as the *science* of *analogy* and the widespread applicability of mathematics in the natural sciences, which has intrigued all mathematicians of a philosophical bent, arises from the fundamental rôle which comparisons play in the mental process we refer to as 'understanding'.

But let us come down from these lofty philosophical heights and ask a more pragmatic question. If a biologist is someone who studies plants and animals, what does a mathematician study? The answer should surprise no-one – he studies equations; first, at the lowest level, algebraic equations and then, at a higher level, differential equations. This oversimplification does at least have the merit of being generally understood, since I imagine that an equation of the form

$$ax^3+bx^2+cx+d = 0$$

will scare few people in this audience. Let me take this as my starting point and compress into a few minutes, what it took mankind several centuries to master.

The most basic question to ask about this equation is: how many solutions (for the unknown x) does it have? This depends on the values of the coefficients a, b, c, d and a complete enumeration of all possible cases is lengthy and tedious. In fact a famous mathematician who would undoubtedly have earned a Nobel Prize for Literature, had he not lived in the twelfth century, wrote a whole treatise on this topic enumerating, I believe, 26 different cases. I refer to the Persian mathematician, astronomer, scholar and poet Omar Khayyám. It took several hundred years before some order was imposed on this chaos by mathematicians adopting a sufficiently enlarged point of view, so that the equation always has three solutions (unless all coefficients are zero). This requires three conventions

(i) a solution may be 'repeated', so as to count more than once,

(ii) a solution may be infinite (if $a = 0$),

(iii) a solution may not be real, i.e. it may involve the imaginary number $\mathrm{i} = \sqrt{-1}$.

Of these by far the most significant is number (iii), the introduction of *complex numbers*. Familiarity, in mathematics as elsewhere, breeds contempt, and $\sqrt{-1}$ is now employed with gay abandon by physicists, engineers and schoolboys. Nevertheless, it must rank as one of the remarkable creations of the human mind and as Gauss, the greatest mathematician since Newton, said 'the true metaphysics of $\sqrt{-1}$ is elusive'.

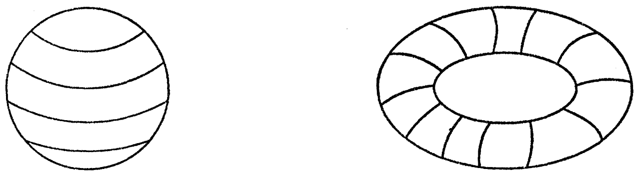
In this sense therefore, an equation of degree 3 has three solutions and an equation of degree n (i.e. involving x^n) has n solutions. Note that this is a basic qualitative statement, only the number of solutions is considered, not their magnitude, reality or other properties, let alone any discussion of how to find them.

Suppose now we move on to an equation with two unknowns, for example

$$x^2+y^2 = 1, \tag{1}$$

$$x^3+y^2 = 1. \tag{2}$$

Every schoolboy now recognizes the first as the equation for a circle in the (x, y) plane, but this geometrical way of representing algebraic solutions was the great contribution of Descartes. Considered as the graph of the function $y = \pm\sqrt{(1-x^2)}$ this is the forerunner of graphical representations of functions so familiar to all scientists. It becomes, however, a little less familiar, and more impressive, when as before we allow complex solutions (as well as infinite and multiple solutions). The (x, y) plane must now be thought of as having two complex or four real dimensions ($x = x_1+\mathrm{i}x_2$, $y = y_1+\mathrm{i}y_2$) and the set of solutions (the graph or curve) has one complex dimension or two real dimensions. It turns out that, in this sense, the points satisfying equation (1) form the surface of a sphere, while those satisfying equation (2) form the surface of a torus (bicycle tire), i.e. a sphere with a hole bored through it.



This is a fundamental qualitative difference between the two equations. The *shape* of this surface is the analogue of the *number* of solutions for equations in one variable. Again we ignore size and other features. Thus faced with equations in two variables we classify them like a good biologist, not according to superficial differences, but according to basic structural features. Of course, as in biology, one expects such structural features to reflect themselves functionally in important ways. This certainly happens with our equations because the integrals

$$\int \frac{\mathrm{d}x}{\sqrt{(1-x^2)}} \quad \text{and} \quad \int \frac{\mathrm{d}x}{\sqrt{(1-x^3)}}$$

turn out to differ sharply in their behaviour; the first involves trigonometric functions and the second involves elliptic functions, whose study was a major preoccupation of the nineteenth century.

More generally any polynomial equation in two variables gives rise to a surface with a number of holes called its *genus*, and this can take any value. This is the reason why surfaces with holes are not just an amusing pastime but of quite fundamental importance for the theory of equations. So much was known by the middle of the last century. Since then great efforts have been devoted, notably by Sir William Hodge, to extending this classification to equations in more variables, and also to simultaneous equations. The solutions must now be represented by points of a k-dimensional complex 'surface' or a $2k$-dimensional real 'surface' – usually called *manifolds*.

As might be expected the possibilities are now much more varied and numerous. To start with we have 'holes' of different types or dimensions. For example the inside of a circle in the plane is a one dimensional hole like that of a torus because we can tie a piece of string round it, whereas the inside of a sphere is a two dimensional hole because we need a bag to enclose it. Moreover, the various holes are not independent of each other, they can interact in quite complicated ways. For example on a manifold of dimension 4 there may be holes of dimension 1, 2, 3 and 4. The number of 1-holes turns out to be equal to the number of 3-holes and there is just one 4-hole, the 'inside'.

There are great conceptual difficulties in visualizing these higher-dimensional manifolds, but they are just as 'real' and structurally important as the ones we can represent in familiar three space, since the graph of a function is not really related to physical space. To deal with these difficulties a whole new language with new techniques has had to be developed and this is called *topology*.

When we considered equations in two variables we ended up with surfaces, and these we are familiar with and know how to recognize: we can 'see the holes'. Before dealing with equations in many variables we must first understand how manifolds can be constructed, how they can be built around their holes. At this stage we are not restricted to looking at manifolds of solutions of equations (which turn out now to be a small sub-class) and in particular we can consider manifolds of any real dimension, not necessarily even.

A relevant question which one may now ask is how far will our physical three dimensional intuition take us in higher dimensions. Besides certain obvious extensions, such as allowing bigger dimensional holes, are there entirely new phenomena having no counterpart in our own experience? To illustrate the point imagine we lived entirely in a two-dimensional world. It would then be a great surprise when the first venturesome mathematician went into three-dimensional space and discovered the existence there of knots.

Well, one entirely new phenomenon in higher dimensions was discovered about twenty years ago by a young American mathematician named John Milnor. To explain the nature of his discovery let me remind you of an important distinction in mathematics between a continuous curve, i.e. one without jumps (figure 1)

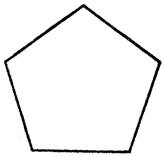


FIGURE 1

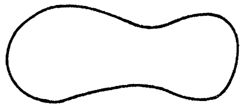


FIGURE 2

and a smooth curve, i.e. one with no corners (no jump in the tangent directions) (figure 2). To a topologist these curves are equivalent and he is quite happy to 'round off' corners. Milnor's astonishing discovery was that in higher dimensions this rounding off of corners is not such an easy process – in fact it may be impossible. More specifically Milnor exhibited (mathematically speaking!) two seven dimensional manifolds, one the surface of the standard sphere S^7 in eight space given by $\sum_{i=1}^{8} x_i^2 = 1$ and the other an 'exotic' sphere M^7 such that M^7 could be deformed into S^7 if corners are allowed but not smoothly. In other words there are two different schemes of classification, one (topology) using continuity and the other (differential topology) requiring also continuity of tangents (derivatives). Moreover Milnor showed that there are precisely 28 different seven dimensional 'spheres'.

Without any doubt this was a landmark in twentieth-century mathematics and it opened up an enormous new field of active exploration. However, for some time people felt that these exotic manifolds were perhaps artificial constructs not likely to appear in any natural setting, particularly in relation to algebraic equations. Of course the Milnor sphere has seven dimensions so it cannot directly represent the solution manifold of algebraic equations. However, it turned up a few years later in a closely related way as I shall now explain.

Let me first refer to the words in my title. Global geometry is, as the word suggests, the study of the total manifold including, for example, the number of holes. Local geometry is the study of small pieces of a manifold. Now from the point of view of topology such small pieces look the same. Interesting questions only arise either if we study more refined questions, involving distance, curvature, etc. (local differential geometry, which I shall not pursue) or if our manifold has what are called *singular points*. I have so far ignored these but they may arise already for curves (figure 3),

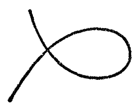


FIGURE 3

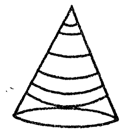


FIGURE 4

If we think of a cone (figure 4), then the local behaviour near the singular point (vertex) clearly reflects the global behaviour of the base of the cone (which has one lower dimension). Roughly speaking we might say that on close inspection under a microscope the geometry near the singular point will get enlarged and show up as the global geometry of the base of the cone. In particular if our cone represented the solution of an algebraic equation its base would have an odd dimension. It turns out that Milnor's exotic sphere arises in this way as describing the behaviour at the origin of the deceptively simple equation (Brieskorn 1966)

$$x_1^3+x_2^5+x_3^2+x_4^2+x_5^2 = 0.$$

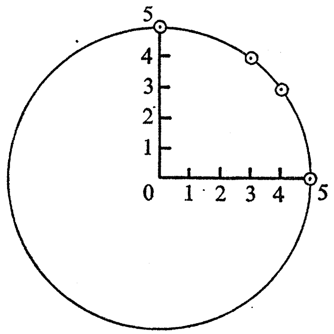
Not unnaturally this discovery has stimulated a tremendous amount of work on the local structure of singularities.

It should be clear from this example that the geometrical structure of the solutions of an equation can be remarkably intricate, even for fairly simple equations. It is by no means easy to decide, from the algebraic form of an equation what the topological properties will be. In fact it is not *a priori* clear that any definite formula can be found for the number of holes: it might be that we simply have to program a computer for each equation and wait until it has computed sufficiently many solutions so that we can, with confidence, plot its graph. Fortunately the situation is not as bad as all that. Quite simple formulae sometimes exist and in general there is an algebraic procedure which, starting from the equations, will determine the number of holes. This has, however, required the concerted efforts of the world's best mathematicians and has only been solved in recent years. In fact the main impetus for this algebraic solution came from the Theory of Numbers and takes us back several thousand years.

In the early days an equation (for example)

$$x^2+y^2 = 25$$

was one in which the coefficients were integers and x, y stood for unknown *integers*, only much later were fractions, real numbers and finally complex numbers admitted. Solving such a Diophantine equation is quite a different story. In geometrical terms we are looking for points of a curve which happen to be points of a lattice in the plane



In this special case the solutions are easy to find but in general the difficulties are enormous, and after 2000 years the subject is still in its infancy. As a first step, therefore, number theorists look at congruences *modulo a* prime number p, that is

$$f(x, y) \equiv 0 \mod p,$$

which means that p divides the integer $f(x, y)$. Clearly an integer solution gives a congruence solution for all primes p. The converse is not usually true but it is a solid step in the right direction.

It is easy to see that if (x, y) is any solution of our congruence we can get other solutions by adding multiples of p. We can therefore insist that our integers x, y take one of the values $0, 1, 2, \ldots, p-1$ and in all calculations we throw away multiples of p. In fact the set $\{0, 1, \ldots, p-1\}$ treated in this way behaves just like the set of all fractions: we can add, multiply and divide (by anything $\neq 0$). Thus if $p = 5$ we have $3.2 \equiv 1 \mod 5$ or $\frac{1}{2} \equiv 3 \mod 5$. For this new algebraic system (which is finite) we now want to know how many solutions our equation has. One of the major achievements of recent times has been the essentially complete solution of this problem for any number of equations, any number of variables and all primes. The remarkable result is that the number of solutions can be expressed in terms of the holes of the manifold of complex solutions of the original equation. Thus our belief that the structure of the manifold of solutions should manifest itself in other properties of the equation is triumphantly justified.

Let me try to show how one makes this miraculous link between the holes in the complex solutions and the numbers of mod p solutions. For this I need first two elementary congruences

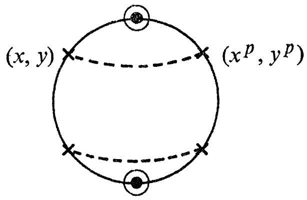
$$(x+y)^p \equiv x^p + y^p \mod p, \tag{1}$$

$$a^p \equiv a \mod p. \tag{2}$$

The first follows from the binomial theorem, since all intermediate coefficients are divisible by p, and the second follows by repeated application of the first (put $y = 1$, $x = 1, 2, \ldots$). The two congruences have, therefore, a slightly different character, the first is true for formal reasons and remains true even if we enlarge our system $\{0, 1, \ldots, p-1\}$ in the same way as the rationals are enlarged to the complex numbers, while the second is true only for $a = 0, 1, \ldots, p-1$. In fact $x^p - x \equiv 0 \bmod p$ is an equation for x which has $0, 1, \ldots, p-1$ as *all its solutions*, even when our system is enlarged: in other words it factorizes in the form

$$x^p - x \equiv x(x-1)(x-2)\ldots(x-(p-1)) \quad \bmod p.$$

Now suppose we have a congruence $f(x, y) \equiv 0 \bmod p$. We first look at all its solutions in an extended number system: the analogue of our complex solution manifold. If we replace (x, y) by (x^p, y^p), the congruence (1) above (together with the more obvious one $(xy)^p = x^p y^p$) shows that we get a new solution. Thus we get a 'motion' of our 'manifold' of solutions



and the solutions which we really want are those for which $x^p \equiv x$ and $y^p \equiv y$, namely the *fixed points* of our 'motion'. What we need therefore is a formula for calculating the number of fixed points of a 'motion'.

Let us now revert to our 'real life' manifold of complex solutions. On these replacing (x, y) by (x^p, y^p) is not permissible (it does not preserve the equation $f(x, y) = 0$), but there are other more geometrical motions. For instance on the ordinary sphere we might have a rotation with two fixed points (the poles) or, reflexion in the centre, taking every point to its antipode which has no fixed points. The difference between these two cases is that reflexion turns the sphere inside out: it affects the two dimensional hole. This situation turns out to be typical in the following sense. Any motion of a manifold has different effects on the different holes and there is the famous formula, called the *Lefschetz fixed point formula* which expresses the number of fixed points quite simply in terms of the effect of the motion on the various holes. The French mathematician André Weil conceived the audacious idea of applying this Lefschetz formula not to honest geometrical motions of real manifolds but to the algebraic motion of replacing (x, y) by (x^p, y^p) in the solution manifold of the mod p congruence. For this idea to make any sense we have to know how to identify holes algebraically and the search for such an algebraic identification has been going on for about twenty years, culminating in a resounding success just two years ago with the work of the young Belgian mathematician Pierre Deligne. I am happy to report that this great enterprise did not pass unnoticed by the Royal Society and that both Lefschetz and Weil figure among our Foreign Members.

Perhaps I could try to sum up with a few crisp phrases. Whereas nineteenth-century mathematics was primarily concerned with functions of one variable, the dominant theme in the twentieth century has been the problem of many variables. Great emphasis has therefore been put on basic structural features and these have in turn led to spectacular links between the discrete and continuous aspects of algebraic equations. In some ways this is reminiscent of the wave-particle duality of matter, whose mathematical formulation lies in the field of differential equations.

References

Brieskorn, E. 1966 Examples of singular normal complex spaces which are topological manifolds. *Proc. natn. Acad. Sci. U.S.A.* **55**, 1395–1397.

Deligne, P. 1974 La conjecture de Weil I. Publ. Math. Institut des Hautes Etudes Scientifiques No. 43, 273–307.

Milnor, J. 1956 On manifolds homeomorphic to the 7-sphere. *Ann. Math.* **64**, 399–405.

Weil, A. 1949 Number of solutions of equations in finite fields. *Bull. Am. math. Soc.* **55**, 497–508.