THE UNITY OF MATHEMATICS

M. F. ATIYAH†

1. *Introduction*

There are several views on the function of a presidential address. There is the view, put forward so ably by my predecessor David Kendall, that the president has to fill the awkward gap between the AGM and the annual dinner and so he should be brief and entertaining. Another rather lofty view is that a presidential address is a unique event (for the speaker if not the audience) and that it should be used to present a grand survey of mathematics, or of some major branch of it. I confess that I was tempted by this second view, particularly after re-reading Hermann Weyl's article in the AMS monthly entitled "A half-century of mathematics". Bearing in mind the increasing pace of our subject and the fact that no one can hope to emulate Weyl either in mathematical scope or in literary style, I would have been content with a more modest quarter-century. Even this I found daunting and there was in addition the danger of flying at such high altitude that nothing would be visible except cloud. Because of this and bearing in mind Kendall's dictum, I have decided to stay closer to earth.

I want therefore to use this occasion to express my personal attitude to mathematics, but to do this by way of simple example rather than by philosophical generalities. The aspect of mathematics which fascinates me most is the rich interaction between its different branches, the unexpected links, the surprises, and my aim will be to illustrate this by considering some simple problems.

The lecture will be divided into two halves, like some examinations, with an easy compulsory first half followed by an optional second half for more advanced candidates. In the first part I will list three simple items from three different branches of mathematics and proceed to show how they are related to one another. Although simple they contain germs of ideas which have been extensively developed over the past twenty years, and in the second part I will mention some of the striking results that have emerged as the end-product of this development. So if the first part appears too easy just wait for the end, while if the second part is too difficult, remember that the essential ideas are all contained in the simple examples!

2. *Three examples*

I begin with the well-known fact of *Number Theory* that unique factorization fails for the ring $Z[\sqrt{-5}]$ consisting of elements $a+b\sqrt{-5}$ with a, b ordinary integers. Specifically we have two different factorizations:

$$9 = 3^2 = (2-\sqrt{-5})(2+\sqrt{-5}). \tag{2.1}$$

Unique factorization is restored if we introduce the *ideal* elements

$$p = (3, 2-\sqrt{-5})$$
$$q = (3, 2+\sqrt{-5})$$

†Presidential address delivered to the London Mathematical Society on 19 November 1976.

[Bull. London Math. Soc., 10 (1978), 69–76]

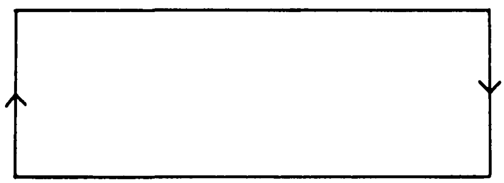
whose products are given by

$$pq = (3), p^2 = (2-\sqrt{-5}), q^2 = (2+\sqrt{-5}).$$

The unpleasant property of 9 is then explained by the identity:

$$(pq)^2 = p^2q^2.$$

Next I remind you of a famous object from *Geometry*, namely the Möbius band. This is most conveniently described by identifying one pair of opposite sides of a rectangle, but putting in a twist:



Without a twist we would get a cylinder, and the interest of the Möbius band is precisely that it is qualitatively quite different from the cylinder.

Finally I will take an equation from *Analysis* of the form:

$$f'(x) + \int a(x, y) f(y) dy = 0.$$

This is a linear integro-differential equation, which depends of course on the kernel function $a(x, y)$, whose precise nature will be stipulated later.

In the succeeding sections I will show how these three examples from Number Theory, Geometry and Analysis all link up quite naturally.

3. *The circle*

In order to make the connection from our Number Theory example to the Möbius band it is natural to consider, as an intermediate object, the circle (the axis of the Möbius band) given by the usual equation:

$$x^2 + y^2 = 1. \tag{3.1}$$

We consider the ring $R[x]$ of real polynomials in x as analogous to the ring Z of ordinary integers. This is reasonable since unique factorization holds in both. The irrationality $y = \sqrt{(1-x^2)}$ we regard as analogous to $\sqrt{-5}$. Then in the ring $R[x, y]$, modulo the relation (3.1), we have

$$x^2 = (1-y)(1+y), \tag{3.2}$$

which like (2.1) says that unique factorization fails.

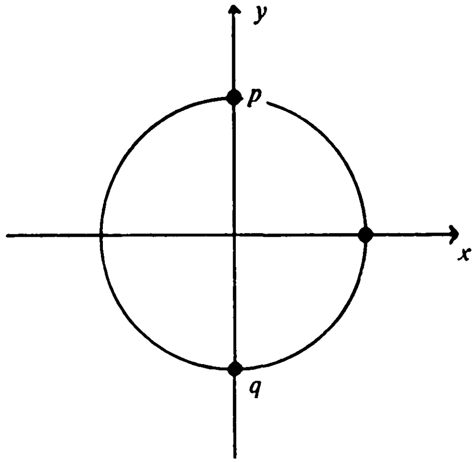
Proceeding as in §2 we introduce ideal elements

$$p = (x, 1-y)$$
$$q = (x, 1+y)$$

whose products are given by

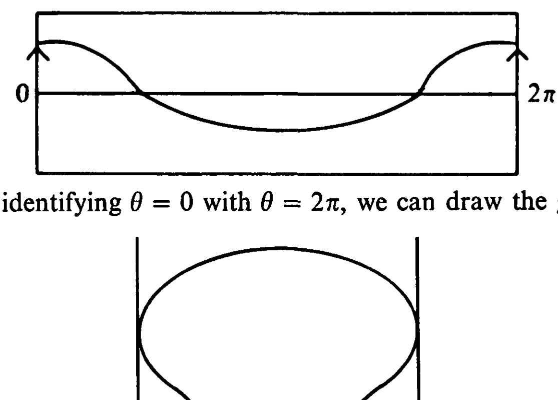
$$pq = (x),\ p^2 = (1-y),\ q^2 = (1+y). \tag{3.3}$$

The advantage of the ring $R[x, \sqrt{(1-x^2)}]$ over $Z[\sqrt{-5}]$ is that we have the geometry of the circle at our disposal. The ideal elements p and q are represented by the points p and q in the figure

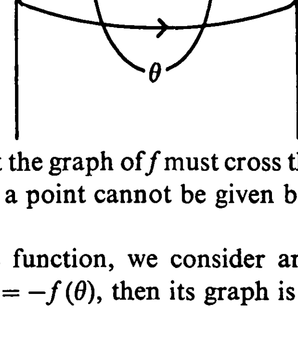


In fact p is the unique point on the circle satisfying both equations $x = 0$ and $1-y = 0$, while q is similarly given by $x = 0$ and $1+y = 0$. The identities (3.3) can then be interpreted geometrically as saying that $x = 0$ cuts the circle in the points p and q, while $1-y = 0$ is the tangent at p and $1+y = 0$ is the tangent at q. The failure of unique factorization in $R[x, \sqrt{(1-x^2)}]$ is thus tied to the fact that a single point on the circle cannot be given by a single extra polynomial equation $f(x, y) = 0$.

If we put $x = \cos\theta$, $y = \sin\theta$ then any polynomial $f(x, y)$ becomes in particular a continuous function $f(\theta) = f(\cos\theta, \sin\theta)$ which is periodic: $f(\theta+2\pi) = f(\theta)$. The graph of f can be drawn in the usual way in the plane

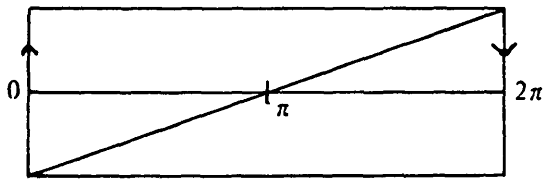


or better still, by identifying $\theta = 0$ with $\theta = 2\pi$, we can draw the graph of f on the cylinder



It is now intuitively clear that the graph of f must cross the $\theta = 0$ circle an even number of times. Thus the fact that a point cannot be given by one equation is essentially a topological fact.

If, instead of a periodic function, we consider an antiperiodic function, i.e. a function $f(\theta)$ with $f(\theta+2\pi) = -f(\theta)$, then its graph is naturally drawn on a Möbius band. Moreover such an f can very well have a single zero in the interval $[0, 2\pi]$: take for example $f(\theta) = \theta - \pi$ with graph



To sum up we see that the existence of the Möbius band is intimately related to the non-uniqueness of factorization in $R[x, \surd(1-x^2)]$ which is formally similar to that in $Z[\surd -5]$.

4. *Parity*

I shall now connect the Möbius band (via anti-periodic functions) with our integro-differential equation. To make this precise we define the linear operator A by

$$(Af)(x) = f'(x) + \int_0^{2\pi} a(x, y) f(y) dy,$$

where we now assume that a is real, continuous and skew, i.e. $a(x, y) = -a(y, x)$. We will further assume either,

(i) $a(x, y)$ is *periodic* in each variable (with period 2π) or

(ii) $a(x, y)$ is *anti-periodic* in each variable.

In case (i) A acts on periodic functions f, while in case (ii) it acts on anti-periodic functions. In both cases it is a skew-adjoint operator, i.e. $\int (Af)g = -\int f(Ag)$.

Consider first the trivial case $a = 0$ (which is common to both (i) and (ii)). Then $Af = 0$ implies that f is a constant. Hence in case (i) $Af = 0$ has a 1-dimensional solution space while in case (ii) it consists only of 0. In general we have the following

THEOREM. *The dimension of the space of solutions of $Af = 0$ is odd in the periodic case and even in the anti-periodic case.*

Proof. Since A is skew-adjoint its eigenvalues are purely imaginary. Since A is also real the non-zero eigenvalues occur in complex conjugate pairs. Hence the multiplicity of the 0-eigenvalue of A, taken modulo 2, is invariant under continuous change of a. Replacing a by ta and making $t \to 0$ reduces us to the trivial case of $a = 0$ which we noted above.

Remarks. (1) This proof uses basic continuity properties of the eigenvalues of A which are easy to establish.

(2) Taking

$$a(x, y) = \sum_{n=1}^{N} \frac{n \sin n(x-y)}{\pi} \quad \text{(periodic case)}$$

$$= \sum_{n=1}^{N} \frac{(n-\frac{1}{2})\sin[(n-\frac{1}{2})(x-y)]}{\pi} \quad \text{(anti-periodic case)}$$

gives examples in which the dimension of the null space of A is $2N+1$ (or $2N$), showing that all possibilities actually occur.

The theorem shows that the topological difference between the Möbius band and the cylinder is reflected in an analytical parity for our operator A.

5. *Modules and Bundles*

From a modern point of view the lack of unique factorization in our examples is expressed by saying that the ideals are not all principal. An alternative statement can be made in terms of modules which is somewhat more geometrical. I shall illustrate this by our examples, beginning with the Möbius band M.

Let us consider M as a band of infinite width (analogue of an infinite cylinder). Then M can be described as a family of lines M_θ parametrized by a point θ on the circle. Each line is a real vector space of dimension one but there is no preferred basis. The normals to M form another such "line bundle" $M^\perp$ over the circle. The direct sum $M \oplus M^\perp$ is a 2-dimensional vector bundle over the circle (fibre at θ being $M_\theta \oplus M_\theta{}^\perp$). We can regard this as the normal bundle to the central axis of the Möbius band in R^3. Since this axis is a standard flat circle its normal bundle is trivial, i.e. it has a global basis. Thus the line-bundle M (which is not trivial) appears as a direct summand of a trivial bundle.

Now let us consider the ring $Z[\surd -5]$ and the ideal $p = (3, 2-\surd -5)$ as before. Computing its inverse we find

$$p^{-1} = p^{-2}.p = \frac{1}{2-\surd -5}(3, 2-\surd -5) = \left(1, \frac{1-\surd -5}{3}\right).$$

Since the determinant of the matrix

$$\begin{pmatrix} 1 & \dfrac{1-\surd -5}{3} \\ 3 & 2-\surd -5 \end{pmatrix}$$

is equal to 1 it follows that the module $p \oplus p^{-1}$ is free of rank 2 (over $Z[\surd -5]$), while p is not free of rank 1 (not being principal).

In matrix terms, having a non-trivial summand of a free module of rank 2 is equivalent to having a 2×2 matrix T (with coefficients in the given ring) such that $T^2 = T$ but with T not conjugate to the standard idempotent matrix, i.e. T cannot be put in the form

$$T = Q\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}Q^{-1}$$

with Q and Q^{-1} being matrices with coefficients in our ring.

The two rings we have considered explicitly are $Z[\surd -5]$ and $R[x, \surd(1-x^2)]$. The latter relates to the circle as the real algebraic curve $x^2+y^2 = 1$, but we can also consider the circle simply as a topological space in which case the appropriate ring $\mathscr{R}$ (circle) consists of all continuous real-valued functions on the circle. For all three rings we have an "interesting" 2×2 matrix T as above.

The analogy between vector bundles and projective modules (direct summands of free modules) which I have been exemplifying has proved very fruitful. The basic ideas are due to Serre [6] who in particular was led to conjecture that projective modules over a polynomial ring (over a field) are necessarily free. Since a polynomial ring corresponds geometrically to a linear space, which is contractible when the field is R, this conjecture is what one would expect from the topological analogue. Serre's conjecture was a major outstanding problem for many years but it has recently been settled affirmatively in a brilliant short proof by Quillen [7].

Vector bundles in topology, differential and algebraic geometry have been extensively studied in recent years with a wealth of interesting results and applications. Perhaps it is sufficient to recall the solution by Adams [1] of the famous vector-field problem on spheres, which asks for the maximum number of linearly independent tangent vector fields on the n-sphere (as a function of n).

6. *Bundles and Operators*

The Möbius band owes its existence to the fact that R^*, the multiplication group of the real field, is not connected. More generally vector bundles of dimension n over general topological spaces depend on the topological properties of $GL(n, R)$. For various purposes it is convenient to stabilize by increasing n, using the natural embedding $GL(n, R) \subset GL(n+1, R)$ given by $A \mapsto \begin{pmatrix} A & 0 \\ 0 & 1 \end{pmatrix}$. We put $GL_\infty = \bigcup_n GL(n, R)$.

This space figures in the famous Bott periodicity theorems [4] which are the basis of K-theory [2]. On the other hand it is essentially equivalent to a certain space of operators which turns up naturally in functional analysis, and this equivalence lies behind the parity discussion of §3.

To explain this let us consider a real Hilbert space H (for example the space of real L^2 functions on the circle). We consider all bounded linear operators $A : H \to H$ with the usual operator norm $\|A\| = \sup_{|x|=1} |Ax|$. An operator A is skew-adjoint if $\langle Ax, y\rangle = -\langle x, Ay\rangle$ for all $x, y \in H$. It then has a spectrum on the imaginary axis. We now assume in addition that $\lambda = 0$ is an eigenvalue of finite multiplicity and is isolated in the spectrum of A. Equivalently A is invertible on the orthogonal complement of the (finite-dimensional) null-space. Let $\mathscr{A}$ denote the space of all such A: it has a metric space topology induced by the norm. Then we have the following [3]:

THEOREM. *GL_∞ is homotopically equivalent to $\mathscr{A}$.*

In particular both spaces have the same number (namely 2) of components. In GL_∞ the components correspond to the sign of the determinant, while in $\mathscr{A}$ they are determined by the parity of the dimension of the null-space.

7. *Functional Analysis*

The Theorem in the previous section has an analogue when the real field is replaced by the complex field and this in turn has had a remarkable application to a theorem of pure functional analysis which I shall now describe.

We consider now a complex Hilbert space H and recall that a bounded linear operator A on H is said to be *Fredholm* if its range is closed and both A and its adjoint A^* have finite-dimensional null spaces $\mathcal{N}(A)$ and $\mathcal{N}(A^*)$. We then define

$$\text{index } A = \dim \mathcal{N}(A) - \dim \mathcal{N}(A^*).$$

Note that an operator with $A^* = \pm A$ has index zero. In §6 our operators (besides being real) were skew-adjoint Fredholm operators and so had index zero: however $\dim \mathcal{N}(A) \bmod 2$ was then a mod 2 analogue.

The standard example of a Fredholm operator with non-zero index is the k-shift A, defined in terms of an orthonormal base $\{e_i\}$ by $Ae_i = e_{i+k}$ for an integer $k \geqslant 0$. Clearly $A^*e_i = e_{i-k}$, where $e_i = 0$ if $i \leqslant 0$, and $\mathcal{N}(A^*)$ is spanned by $e_1, \ldots, e_k$ while $\mathcal{N}(A) = 0$. Thus index $A = -k$, and this result holds also for negative k if we again interpret e_i as zero for negative i. Note that

$$A^*A = 1 + (\text{finite rank operator})$$

and so in particular

$$A^*A = 1 + \text{compact operator}.$$

If we now perturb A by a compact operator K, putting $B = A + K$, then we find

(1) index $B = -k$

$$(2)\begin{cases} B^*B = 1 + \text{compact operator} \\ B\ B^* = 1 + \text{compact operator}. \end{cases}$$

Property (1) is a consequence of the general fact that the index of any Fredholm operator is invariant under compact perturbation, while (2) follows from the fact that the compact operators form a 2-sided ideal in the algebra of all bounded operators. Note also that property (2) implies that B is Fredholm.

We can now ask the converse question, namely is every solution of (1) and (2) a compact perturbation of a k-shift? This is answered by the following:

THEOREM. *If B is a bounded operator satisfying* (1) *and* (2) *with $k \neq 0$ then, with respect to a suitable orthonormal basis of H, $B = A + K$ where K is compact and A is the k-shift.*

This theorem is a very special case of the deep results of Brown, Douglas and Fillmore [5]. It is clearly a theorem of pure Hilbert space analysis and it answers a simple natural question. However its proof rests on the important link with topology on the lines of §6. The excluded case $k = 0$ is also covered by the results of [5], but the conclusion is then different. The point is that the essential spectrum of B is the whole unit circle if $k \neq 0$, but can be any closed subset Σ if $k = 0$. Clearly Σ is an additional invariant for B and the general result is that two B with $k = 0$ and the same Σ are unitarily equivalent modulo compact operators.

8. *Concluding Remarks*

The main theme of my lecture has been to illustrate the unity of mathematics by discussing a few examples that range from Number Theory through Algebra, Geometry, Topology and Analysis. This interaction is, in my view, not simply an occasional interesting accident, but rather it is of the essence of mathematics. Finding analogies between different phenomena and developing techniques to exploit these analogies is the basic mathematical approach to the physical world. It is therefore hardly surprising that it should also figure prominently internally within mathematics itself. I feel that this needs to be emphasized because the axiomatic era has tended to divide mathematics into specialist branches, each restricted to developing the consequences of a given set of axioms. Now I am not entirely against the axiomatic approach so long as it is regarded as a convenient temporary device to concentrate the mind, but it should not be given too high a status.

A secondary theme implicit in my lecture has been the importance of simplicity in mathematics. The most useful piece of advice I would give to a mathematics student is always to suspect an impressive sounding Theorem if it does not have a special case which is *both* simple *and* non-trivial. I have tried to select examples which satisfy these conditions.

Both unity and simplicity are essential, since the aim of mathematics is to explain as much as possible in simple basic terms. Mathematics is still after all a human activity, not a computer programme, and if our accumulated experience is to be passed on from generation to generation we must continually strive to simplify and unify.

References

1. J. F. Adams, " Vector fields on spheres ", *Ann. of Math.*, 75 (1962), 603–632.
2. M. F. Atiyah, *K-Theory* (Benjamin, 1967).
3. M. F. Atiyah and I. M. Singer, " Index Theory for Skew-adjoint Fredholm Operators ", *Publ. Math. Inst. Hautes Etudes Sci. Paris* (1969), No. 37.
4. R. Bott, " The stable homotopy of the classical groups ", *Proc. Nat. Acad. Sci. U.S.A.*, 43 (1957), 933–935.
5. L. G. Brown, R. G. Douglas and P. A. Fillmore, " Extensions of C^*-algebras, operators with compact self-commutators, and K-homology ", *Bull. Amer. Math. Soc.*, 79 (1973), 973–978.
6. J. -P. Serre, " Faisceaux algébriques cohérents ", *Ann. of Math.*, 61 (1955), 197–278.
7. D. Quillen, " Projective modules over polynomial rings ", *Inventiones Math.*, 36 (1976), 167–172.

Mathematical Institute
University of Oxford
Oxford, OX1 3LB.