DUALITY IN MATHEMATICS AND PHYSICS*

SIR MICHAEL F. ATIYAH

Abstract. Duality is one of the oldest and most fruitful ideas in Mathematics. I will survey its history, showing how it has constantly been generalized and has guided the development of Mathematics. I will bring it up to date by discussing some of the most recent ideas and conjectures in both Mathematics and Physics.

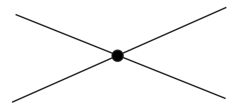
Introductory remarks

Duality in mathematics is not a theorem, but a "principle". It has a simple origin, it is very powerful and useful, and has a long history going back hundreds of years. Over time it has been adapted and modified and so we can still use it in novel situations. It appears in many subjects in mathematics (geometry, algebra, analysis) and in physics. Fundamentally, duality gives *two different points of view of looking at the same object*. There are many things that have two different points of view and in principle they are all dualities.

Linear duality in the plane. It starts off classically in geometry with linear duality in the plane.

In the plane we have points and lines. Two different points can be joined by a unique line. Two different lines meet in one point unless

* Editorial note. On December 18, 2007, Atiyah delivered the lecture *Riemann's Influence in Geometry, Analysis and Number Theory* at the Facultat de Matemàtiques i Estadística of the Universitat Politècnica de Catalunya and the lecture *Duality in Mathematics and Physics*, this one sponsored by the Centre de Recerca Matemàtica (CRM), at the Institut de Matemàtica de la Universitat de Barcelona (IMUB). The present text has been produced from a transcription of the second lecture and can be downloaded from the web pages www.imub.ub.es or www-fme.upc.edu. The text of the first lecture can also be downloaded from the latter web page.

they are parallel. People did not like this exception and so they worked hard and realized that if they added some points at infinity then they got what is called the *projective plane* in which the duality is perfect: the relationship between points and lines is perfectly symmetrical. This led to the classical principle of projective duality, which says that the "dual statement" of a theorem is also a theorem, so that we can talk about the dual theorem.

Linear algebra. In linear algebra duality involves the pairing

$$\langle \boldsymbol{\xi}, \boldsymbol{x} \rangle = \sum_i \xi_i x_i$$

of *dual vector spaces.*

For example, let $n = 3$. If we fix a $\boldsymbol{\xi}$, then the equation $\langle \boldsymbol{\xi}, \boldsymbol{x} \rangle = 0$ gives a linear condition on $\boldsymbol{x}$ which determines a plane in 3-dimensional vector space. On the other hand if we fix $\boldsymbol{x}$ and let $\boldsymbol{\xi}$ vary, then we get a plane in the dual space. These are vector spaces and if we factor out by homogeneous coordinates (this leads to one dimension lower) we get the lines of the projective plane. So the relationship between projective geometry and linear algebra is very simple: we write equations and we identify those that differ by a non-zero scalar factor.

Linear analysis

When we go from linear algebra to linear analysis, in principle we make $n = \infty$ (infinite dimension linear spaces), and we have to be careful about questions of convergence. Formal infinite series do not make sense and so we have to have suitable continuity conditions.

Example. If we have functions of one variable x, then we can define the $\langle f, g \rangle$ of two functions f and g by

$$\langle f, g \rangle = \int f(x) g(x) dx.$$

Assuming that the integral converges, this defines a pairing which is an ∞-dimensional version of the finite pairing of linear algebra. So immediately the ideas of duality go (provided we are careful about convergence) into linear analysis.

Fourier transform. Now in linear analysis one of the most powerful tools is that of Fourier theory. Let us recall what is the *Fourier transform*. Let $\boldsymbol{x} = (x_1, \ldots, x_n)$ be n variables and $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)$ corresponding 'dual' variables. Then given a function $f(\boldsymbol{x})$, its *Fourier transform* $\hat{f}(\boldsymbol{\xi})$ is defined by

$$\hat{f}(\boldsymbol{\xi}) = \int f(\boldsymbol{x}) \exp(2\pi i \langle \boldsymbol{x}, \boldsymbol{\xi} \rangle) d\boldsymbol{x}.$$

Fourier series. We can also have Fourier series, which in fact are more elementary. If we have a function $f(x)$ of the variable x which is *periodic* (it is a function defined on the circle $\cong \mathbb{R}/\mathbb{Z}$), then we have Fourier *coefficients* a_n:

$$a_n = \int f(x) \exp(2\pi i n x) d\boldsymbol{x}.$$

This transforms functions on the circle $\mathbb{R}/\mathbb{Z}$ to functions on the integers $\mathbb{Z}$. The *inversion formula* says how to go backwards.[1] In this example we see that the two things that are dual need not be of the same kind. In linear duality usually the two sides are the same sort of thing, but here one side is a circle and the other side are the integers, but it is still an example of duality.

Poisson summation formula. This is a very useful formula in connection with the theory of Fourier series. It uses Fourier theory on $\mathbb{R}$ and on $\mathbb{R}/\mathbb{Z}$ and it says, given a function f, that

$$\sum f(n) = c \sum \hat{f}(n),$$

c a suitable constant. In the particular case when the function is the Gaussian $f(x) = \exp(-x^2)$, and taking into account that its Fourier

[1] $f(x) = \sum_n a_n \exp(-2\pi i n x)$

transform $\hat{f}$ is also Gaussian, we get the famous transformation law[2]

$$\Theta(-1/z) = (-iz)^{1/2}\Theta(z)$$

of the *theta function*

$$\Theta(z) = \sum_n \exp(\pi i n^2 z).$$

Besides this formula we have another obvious relation:

$$\Theta(z+2) = \Theta(z).$$

The two operations $z \mapsto z+2$ and $z \mapsto -1/z$ generate a subgroup of index two the modular group $SL(2,\mathbb{Z})$ (we would get the whole group $SL(2,\mathbb{Z})$ if we had the condition $\Theta(z+1) = \Theta(z)$). A function which transforms according to such laws under this group is called a *modular form*. They are very important things in arithmetic and the Θ *modular form* is one of the most classical examples. As we see, modular forms turn up in connection with dualities in Fourier theory. This is of course all very classical, as it was known to Riemann and to people before him.

Duality for Abelian groups

The real line, the circle and the integers are different-looking things. But in fact they are all abelian groups, and for these groups there is a duality theory. Given an abelian group G, the *dual group* $\hat{G}$ is the group of characters of G (homomorphisms of G to the circle group $U(1)$):

$$\hat{G} = \operatorname{Hom}(G, U(1)).$$

Examples. 1) If $G = \mathbb{Z}$, $\hat{G} = S^1$ (circle of radius 1), where $x \in S^1 \cong \mathbb{R}/\mathbb{Z}$ corresponds to the homomorphism $n \mapsto \exp(2\pi i n x)$.

2) If $G = \mathbb{R}$, $\hat{G} = \mathbb{R}$.

3) If G is discrete, then $\hat{G}$ is a compact group (like $S^1 = \hat{G}$ for $G = \mathbb{Z}$).

The general class of abelian groups for which duality theory works are the *locally compact topological abelian groups*. This class includes the groups $\mathbb{R}$, S^1 and $\mathbb{Z}$, and also others. If G is finite (hence it is both

[2] This relation gives one of the ways in which the functional equation for the Riemann ζ function can be proved.

discrete and compact), then $\hat{G}$ is finite. Besides the examples given, there are also important examples in number theory, as the group of p-adic integers:

$$\mathbb{Z}_p = \varprojlim \mathbb{Z}/p^n\mathbb{Z}.$$

Fourier theory works also in this case and is important in applications in number theory.

Non-Abelian groups

Abelian groups give a unifying framework for these classical dualities. We have a good theorem, but it is natural to try to go beyond that. In that regard a reasonable aim is to look at non-abelian groups.

(1) Let us start with a (non-abelian) finite group G. Introduce the *set* $\hat{G}$ of (isomorphism classes) of irreducible *unitary representations* $\mathrm{Hom}(G, U(n))$, which is is called the *dual* of G. It is not a group anymore, it is only a (finite) set.

(2) Consider now a compact Lie group G, like for example the orthogonal group $O(n)$. Then the dual $\hat{G}$, defined as the set of (isomorphism classes of) irreducible unitary representations, is an *infinite discrete set*.

Fourier theory. For any class function f on G, there is a Fourier expansion

$$f(g) = \sum_{\chi \in \hat{G}} a_\chi \chi(g), \quad a_\chi = \int_G f(g)\overline{\chi}(g)dg.$$

The notation $\chi \in \hat{G}$ indicates that the sum is extended over all irreducible characters χ of G.[3] It may be finite or an infinite series depending on the group G.

[3] The character $\chi = \chi_\rho$ of a unitary representation $\rho : G \to U(n)$ of G is the function $\chi : G \to \mathbb{C}$ defined by $\chi(g) = \mathrm{trace}(\rho(g))$. It only depends on the isomorphism class of the representation. The character is said to be *irreducible* if the representation is irreducible. The mapping of isomorphism classes of irreducible representations to irreducible characters is one-to-one.

(3) Non-compact Lie groups. In this case the theory gets more difficult. Let G be a non-compact Lie group, as for example the general linear group $GL(n)$. Then $\hat{G}$ should be the set of isomorphism classes of irreducible representations, including those that are infinite dimensional (i.e., representations in Hilbert space). If G satisfies some conditions, $\hat{G}$ is a set with a measure on it (the Plancherel measure $d\mu$) and for any class function f on G the Fourier inversion formula is

$$f(g) = \int_{\hat{G}} \hat{f}(\chi)\chi(g)d\mu, \quad \hat{f}(g) = \int_G f(g)\overline{\chi}(g)dg.$$

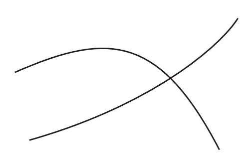
This theory, valid for semisimple Lie groups, is due to Harish-Chandra. It involves infinite-dimensional spaces and continuous parameters, but it is very satisfactory as it generalizes all the other cases.

Non-linear geometry

After the preceding excursions to algebra and to groups, let us go back to geometry.

Suppose we have an n-dimensional manifold which is compact and oriented. Then we can define the *homology groups*, which are given by cycles of various dimensions up to the maximum dimension n. If q is an integer, $0 \leq q \leq n$, we can define the homology group in dimension q, H_q. It is a finite dimensional vector space if we use real coefficients. Then we also have cycles in the complementary dimension, $n-q$, which give the homology group (space) H_{n-q}.

Intersection pairing. If we have a cycle of dimension q and a cycle of dimension $n-q$ in general position, then they intersect in finitely many points and so we get an intersection number.



From this it is possible to construct a pairing of real vector spaces

$$H_q \otimes H_{n-q} \to \mathbb{R}$$

(the proof uses the local Euclidean nature of the manifold).

Poincaré duality. The pairing above is non-degenerate. Therefore each of the homology spaces H_q and H_{n-q} becomes the dual of the other. Thus we obtain a duality between finite-dimensional vector spaces from a curved manifold. In particular, $\dim H_{n-q} = \dim H_q$.

Poincaré duality actually works at the global level of topology and it is a very important tool in for the study of the topology of higher dimensional manifolds.

Hodge theory

Another way of examining this, which is a combination of geometry and analysis, is to use Hodge theory. Given an n-dimensional manifold M, let us consider, instead of cycles, differential forms. The differential forms of degree q, which form a space that we will denote Ω^q, are the natural integrands of integrals on cycles of dimension q.

We will also use the *derivative operator*

$$d : \Omega^q \to \Omega^{q+1}.$$

The $\alpha \in \Omega^q$ such that $d\alpha = 0$ are said to be *closed forms*.

Now if we have $\alpha \in \Omega^q$ and $\beta \in \Omega^{n-q}$, then the wedge product $\alpha \wedge \beta$ is an n-form (also called a *volume form*). This form can be integrated over M,

$$\int_M \alpha \wedge \beta,$$

and this gives a bilinear pairing

$$\Omega^q \times \Omega^{n-q} \to \mathbb{R}$$

under which Ω^q and Ω^{n-q} are dual ∞-dimensional spaces.

Given a Riemannian metric on M, we also have the Hodge duality operator $* : \Omega^q \to \Omega^{n-q}$ that is defined by the relation $(\alpha, \beta \in \Omega^q)$

$$\int_M \alpha \wedge *\beta = (\alpha, \beta),$$

where (α, β) denotes the pairing of α and β by the natural extension of the Riemannian metric to Ω^q.

Hodge theory is concerned with the solutions (called *harmonic forms*) of the two equations

$$d\omega = 0, \quad d * \omega = 0.$$

Thus ω is harmonic if and only if ω and $*\omega$ are closed.

Hodge theorem. The space H^q of harmonic forms of degree q has finite dimension and is isomorphic to the cohomology space of dimension q. Moreover, $* : H^q \to H^{n-q}$ is an isomorphism. In particular, H^q and H^{n-q} are dual finite dimensional spaces (Poincaré duality) and $\dim H^q = \dim H^{n-q}$.

This theorem provides a reinterpretation of the geometrical duality using harmonic forms.

Physics

Now we go to (classical) physics.

1. Position-momentum. Here is an elementary situation. Let x be a variable in ordinary space (it may describe the position of a particle) and a dual variable ξ (it may represent momentum of the same particle). The the Fourier transform gives rise to "dual pictures", which is what physicists call *spectral analysis*.

2. Quantum mechanics. In quantum mechanics, there is the famous particle-wave duality. Duality appears here as two ways of looking at the same thing: either a particle behaves like a point going around or as (quantum) wave. These two aspects are part of the mystery of quantum mechanics.

3. Electromagnetism. In ordinary Minkowski space $\mathbb{R}^{3,1}$, the electromagnetic force is described by a 2-form (skew-symmetric 2-tensor) ω. In this notation, Maxwell's equations *in vacuo* are

$$d\omega = 0, \quad d * \omega = 0,$$

where now $*$ is defined using the Lorentz metric in $\mathbb{R}^{3,1}$. Formally they are the same as Hodge equations for forms of degree 2 on a 4-dimensional Riemannian manifold, but here the space is ordinary Minkowski space, not the 4-dimensional Euclidean space. From this it appears that Maxwell's equations, which unified electricity and magnetism, also encode a duality between electricity and magnetism in the sense that the $*$ operator interchanges both aspects. Physically, this is a very fundamental fact of the universe.

Maxwell's equations actually motivated Hodge for his work on harmonic forms in general. As indicated, Maxwell's equations are about forms of degree 2 in 4 dimensions and Hodge went to forms of any degree q in any dimension n. He also worked in Riemannian geometry, not in Minkowski space.

Modern Physics: Gauge-Theory

A lot of modern physics is concerned about *gauge theory*. This is what physicists use to describe elementary particles. In some naive sense it is just a (non-abelian, non-linear) matrix generalization of Maxwell's theory.

A *Yang-Mills field* F is the curvature of a connection[4] on a fibre bundle (F may be thought as a as a 2-form with matrix coefficients, or as a matrix of 2-forms). Then the Yang-Mills equations take the same form again:

$$dF = 0, \quad d * F = 0.$$

These are now matrix equations. When defined on Lorentz manifolds, they are the fundamental equations that physicists use for elementary particle physics. They can also be defined on a Riemannian 4-manifold, and this leads to geometry, as in Donaldson's celebrated theory, one of the most exciting developments in the last quarter of the 20th century. In particular it produced new invariants of 4-manifolds. The physics inspired the equations and then Donaldson used the equations, in the context of Riemannian geometry, to develop his theory.

This was just mathematics. We might say that mathematicians took an equation that the physicists had written down and studied it in the Riemannian context, just as Hodge did with Maxwell's equations. But subsequently Witten showed ([W94]) that what Donaldson did could be understood in the language of physics. In fact, he interpreted

[4] In gauge theory, connections play the role of potentials.

Donaldson's theory as a Topological Quantum Field Theory, which can be thought of as a "non-abelian Hodge theory".

Going beyond ordinary quantum mechanics, which describes the quantum behaviour of particles, quantum field theory (QFT) describes the quantum behaviour of fields, like the electromagnetic field. Usually quantum field theory supplies information about real particles and their interactions.

In the special case of a topological quantum field theories, the only things that can be extracted are topological invariants. These theories supply discrete quantities that have a topological meaning, not real numbers that can vary continuously. Donaldson's invariants, for example, turned out to be, as shown by Witten, instances of such topological invariants of a QFT. This physical interpretation of Donaldson's theory made the link between physics and mathematics much easier to understand.

One of the most exciting things that physicists have discovered is that in QFT there are many dualities which are not at all easy to understand geometrically. These theories are non-linear, and so they are not trivial. An important observation here is that if we have a given classical geometrical picture, then there is some kind of procedure, called 'quantization', by which we replace classical variables by operators to produce, with a bit of guesswork, a 'quantization' of the original theory. In this quantized theory there is a parameter that plays the role of Planck's constant and which allows us to recover the classical theory by letting it go to 0. Given a quantum theory, however, it may turn out that there are several ways in which it can be realized as the quantization of a classical geometry picture. For example, for a single classical particle we can use the position observable or the momentum observable (they are actually symmetrical), and so there are two ways of reaching the same theory from two geometrical points of view. Then these two points of view are called a duality. This is, of course, a very simple example, but it turns out that physicists have found many much deeper dualities which exist in complicated QFTs which link two very different-looking geometric pictures.

Example: Donaldson's theory. A first example of such a duality is that Donaldson's theory, which has been very roughly described above, is "dual" to Seiberg–Witten theory. Physicists predicted that these two theories should be equivalent, in the sense that one quantum theory has two different descriptions: one in terms of Donaldson's geometry and the other in terms of Seiberg–Witten geometry.

In Donaldson's theory we just take 2×2 matrices and write down equations analogous to Mawwell's equations. In Seiberg–Witten theory, instead of 2×2 matrices, the ordinary circle group is used, and Maxwell's equations for the (classical) electro-magnetic field, but coupled to a spinor field.

Spinors are very important things in physics (they describe particles like the electron) and they turn out to be mysteriously important in geometry. Their geometrical meaning is much more difficult to understand than differential forms, but they exist and they are beautiful geometrical objects. They satisfy the famous equation introduced by Dirac (the Dirac equation), which is a linear equation for spinors like Maxwell's equations are for the electro-magnetic field, but now the spinors and the electro-magnetic field can be coupled together. The reason that they can be coupled together is fairly simple algebra: spinors can be multiplied together and the results are differential forms (not a single degree, but all degrees simultaneously). So spinors are like square roots of differential forms, and that explains why they are so mysterious.[5]

So Spinors$\times$Spinors=Forms. If the form obtained as the product of two spinors is related to Maxwell's equations, we get equations that are quadratic, and therefore non-linear. So there is a non-linear coupling between spinor fields and the electro-magnetic field and this coupling gives rise to a system of equations that are called the *Seiberg–Witten equations*. They involve on one hand spinors, which satisfy Dirac's

[5] One of the most fantastic discoveries in mathematics was the invention of the square root of -1. It took 200 years to understand, and even when it was understood Gauss said that "the real metaphysics of the square root of -1 is not simple". In a similar way, spinors are something like trying to understand the square root of 'area', or the square root of 'volume'. What does that mean? Very mysterious notions, but nevertheless spinors exist in a formal way and can be used in formulas just as i is used all the time in mathematics and physics.

equation, and involve also the electromagnetic field, which satisfies not just the pure Maxwell equations, but the Maxwell equation with a right hand side which is constructed from a spinor roughly speaking by squaring.

The Seiberg–Witten equations are, in some ways, very elementary equations, more elementary than the matrices that Donaldson used, but they are not linear—otherwise they would be rather trivial. What was predicted, on the grounds of physics intuition, was that Donaldson theory and the Seiberg-Witten theory should be equivalent: they should give two different ways of realizing the same QFT. So the invariants that Donaldson had written down and the similar invariants that we can get by solving the Seiberg-Witten equations, are really equivalent information, and what this means is that we have some kind of Poisson summation formula to relate the two. This does not at all say that the invariants on one side are equal to the invariants on the other side. What it says is that the sum of all the Donaldson invariants (they depend on an integer degree) is equal to the sum of the Seiberg–Witten invariants (they also depend on an integer degree), or that the two series give the same information. The equality (Poisson summation formula) is in the sense of generating functions. Symbolically:

$$\sum_{SW} = \sum_{D} .$$

It is a very deep kind of relationship, a non-linear ∞-dimensional version of a Poisson summation formula—that's why this formula was mentioned at the beginning. It is a very important way of getting interesting formulae in ordinary classical mathematics. But it also points to much deeper things that link non-linear dualities in geometry and in physics. This is very much part of modern day physics, for it was physicists who discovered it, not mathematicians.

Of course, we might say that we ought to prove theorems, in this case this prediction of the physicists. It is a very hard question and *no direct mathematical proof is known yet*. Ideas for a proof could come from the fact that the same geometric results are obtained by applying Donaldson's theory or Seiberg–Witten's theory, so that in practical terms they are equally useful. But mathematically, strictly speaking, we don't have a proof because we do not yet have a framework in which this new kind of duality involving infinite dimensional spaces has been formally set up. We are still in the early stages of a theory which will probably be developed in the future. In any case the question is obviously very important.

Example: Mirror symmetry. Another example, also coming out of physics, was equally spectacular. It was called *mirror symmetry*. It does not have the name 'duality', but 'symmetry' here means essentially the same thing. Mirror symmetry can be described in many different ways. Here I will consider one of them.

In geometry there are different kind of geometries:

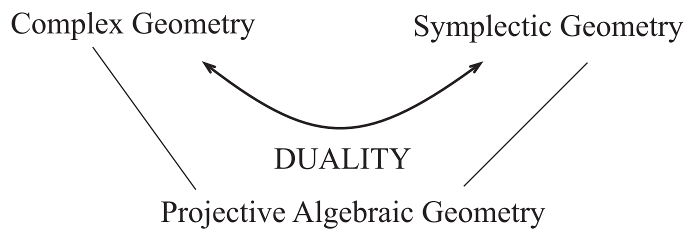
- *Complex geometry*. The coordinates are complex and the admissible changes of variables are holomorphic. Examples: Riemann surfaces and algebraic varieties over the complex numbers.
- *Symplectic geometry*. Real manifolds with a non degenerate skew-symmetric 2-form. They generalize the phase space of classical Hamiltonian mechanics. Algebraic varieties in complex projective space are also symplectic, because the latter has a natural symplectic structure (given by the Kähler metric) and this structure induces a symplectic structure on subvarieties.
- *Riemannian geometry*. Real manifolds with a positive definite metric.

These geometries may be thought as the non-linear versions of the three classical (lie) groups:

- Complex general linear group, $GL(n, \mathbb{C})$.
- Symplectic group, $Sp(n)$ (preserves a skew-symmetric form).
- Orthogonal group, $O(n)$ (preserves a metric).

Seen the other way around, each of these linear theories has a non-linear generalization to a geometry, and this is how complex geometry, symplectic geometry and Riemannian geometry arise.

What the physicists discovered is that there is a certain duality, or symmetry, between the complex geometry and the symplectic geometry. This duality relates in a rather precise way the information on one side to the information on the other side. In particular, *when this*



symmetry is applied to complex projective algebraic varieties we obtain again complex projective algebraic varieties.

There are many concrete examples of this, which were first discovered by physicists. In one class of examples, the algebraic varieties M and M' have (complex) dimension 3 (so 6 real dimensions) and the duality exchanges information on one side with information on the other (very subtle information, as we will see below). Moreover,

$$\dim H^{\mathrm{ev}}(M) = \dim H^{\mathrm{odd}}(M')$$

The dimensions of the homology are the same, but even homology on one side corresponds to the odd homology on the other, which is very hard to understand. It means, for example, that the Euler characteristic (the dimension of the even part minus the dimension of the odd part) changes sign when we go from one to the other. It is a very precise relationship, but very mysterious. Moreover, the details of the geometry give the following: On the M side the information is about *algebraic curves*, for example enumerative questions about rational curves of any degree, and on the M' side it is related to *periodic matrices*.[6]

So this marvellous theorem tells us that easy information on one side (periodic matrices, that can be calculated by classical means) is equivalent to difficult information on the other side (algebraic curves, for whose determinations there is very little information). In physics language, the easy information is what is called classical and the difficult one is what is called quantum. We are thus getting information of a quantum character on one side out of a classical calculations on the other side.

[6] Period matrices started with Riemann and its entries are integrals of forms on cyles.

Using this fantastic duality the physicists were able to calculate the numbers of rational curves of any degree on very simple examples of algebraic varieties of dimension 3. The formulae they got were so spectacular to algebraic geometers that at first they did not believe them, but eventually they were converted. Then they began a big industry that has produced many books on mirror symmetry. It is a whole new area in algebraic geometry that arises out of this particular simple example, just one example of duality in quantum theory.

The applications of physics to mathematics have been numerous over the years. Usually the mathematics is quite close to the physics, but complex algebraic varieties seem as far away removed from physics as one can imagine. Why should physics have anything to do with these concepts in pure geometry?

[So its a spectacular coup: physicists go up into the sky, they land by parachute in the middle of algebraic geometers and they capture immediately the whole city.]

The discovery of mirror symmetry is certainly one of the most remarkable developments of the last part of the 20th century. It provided an example of two different classical theories, two different algebraic varieties, giving rise to the same quantum theory, and with spectacular applications. The mirror symmetry and Donaldson's theory, two of the most interesting developments in recent times, had their origin in physics and in terms of the dualities we have been considering. We see, therefore, that the payoffs of these dualities is very great.

It should be emphasized that the physics, even if it gives the exact answers, does not necessarily give proofs. It gives answers we can believe if we believe that the physics intuition is correct. But mathematicians like to have a proof, and then they have to work very hard to prove the theorems by other methods. Now lots of this has been done, and when they have succeeded, essentially the physicists have been correct. This is an interesting interaction: the physicists generate results that the mathematicians can prove by other means and then they can exchange information.

Number Theory

The preceding two examples were from geometry. Now we will move to number theory and to begin with we will consider a nice beautiful simple example about prime numbers.

Let p be and odd prime. Then we can ask whether it can be represented as the sum of two squares of two integers. The answer is that it can be done if and only if $p \equiv 1 \mod 4$. In other words, if $p \equiv 3 \mod 4$, the answer is no, and if $p \equiv 1 \mod 4$, the answer is yes.

Writing $p = x^2 + y^2$ (x, y integers), which is equivalent to

$$p = (x + iy)(x - iy),$$

we see that we are factorizing a prime number, not in the rational numbers, but in the field we get by adjoining i to the rational numbers.[7] This is an example of algebraic extension of the field of rational numbers.

This is the beginning of what is called *class field theory*, which is trying to give information which relates algebraic extensions of a number field to properties of primes in that field (in the example, information on $\mathbb{Q}(i)$ relates to properties of primes in $\mathbb{Q}$). The theorem above is one of the facts at the origin of class field theory, which essentially worked for all abelian field extensions.

Abelian class field theory, a very fine theory finished in the 20th century, was founded by Gauss, Kronecker and others, and gives a complete answer to the story about abelian extensions of the rational numbers, and also about abelian extensions of number fields. The important thing that comes in is the Galois group of the extension. If the Galois group is assumed to be abelian, the answer is that we can relate this to constructions that can be performed inside the field concerning primes, whereas the extensions are outside the field.

This was a great triumph of classical number theory and ended the questions that Gauss and Hilbert posed of developing a theory in general for all number fields (the original works were about quadratic fields). But in the 20th century people became more ambitious and

[7] If we do this with the integers, we obtain the *Gaussian integers*.

asked about what could be done for non-abelian extensions. Would there still be some kind of duality?

Taking into account the role of representations of groups in linear duality, it is natural to ask whether this idea will also apply to number theory. In other words, is there a way of using representations of non-abelian groups to extend the classical class field theory?

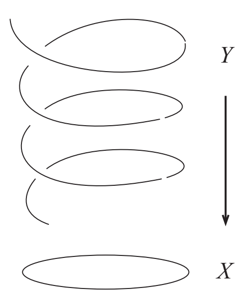
The search for non-abelian class field theory has been going on for over forty years and nowadays it goes under the name of the Langlands programme, because Langlands was the person who put this in a rather precise form—he made many conjectures and proved many special cases.

Langlands programme. The Langlands programme is a large programme which, if eventually completed, will give a very satisfactory answer to the question of how to extend abelian class field theory to the general case of non-abelian extensions. It is a very big programme, that in many ways has just started, and in it a fundamental role is played by complex linear representations of the Galois groups of algebraic extensions.

Simple geometric analogue. Let us consider a simple example or analogue between number theory and ordinary geometry.

In ordinary geometry, or topology, if X is a space, for example the circle S^1, there are two things that we can do.

One is that we can look at the covering spaces that project down onto it as the picture indicates.



For example, if we take X to be a circle, then Y can be the real line, which covers the circle infinitely many times (that is why we take the real numbers modulo 1). When we go to a covering space that cannot go any further, that is called the *universal covering space*.

Consider the fundamental group of the space X, $\Gamma = \pi_1(X)$, It is constructed inside the space, by geometry.[8] Then the fact is that this is related to the covering space situation: if we have the universal covering space Y, then the fundamental group Γ acts freely on Y and the quotient of Y by this group action is X. This is an example of the sort of duality we are talking about, the fundamental group and the universal covering being the 'inside' and 'outside' aspects of the situation.

For example, given a representation of Γ into a matrix group (a homomorphism $\Gamma \to GL(n, \mathbb{C})$), then this gives rise over the space X to what is called a *vector bundle* (the quotient of $Y \times \mathbb{C}^n$ by the action of Γ induced by the representation). This bundle is *flat*.[9] In the case where $n = 1$, we get a *flat line bundle* (a line bundle is a vector bundle of dimension 1). Note that in this case $GL(1, \mathbb{C}) = \mathbb{C}^*$ is abelian.

If X is an algebraic curve or, in other words, a Riemann surface, then it is well known that a Riemann surface is a close analogue of a number field. Not the same thing, of course, but only an analogue. Points of the Riemann surface are the analogues of prime numbers and the number field itself is the analogue of the rational functions (meromorphic functions) on the curve (Riemann surface).

$$\text{Rational (meromorphic) functions on } X \leftrightarrow \text{Number field}$$

The analogy is stressed by the fact that many theorems just look the same on both sides. Even more, people can guess what a theorem

[8] This group is obtained by taken closed paths in X and identifying two paths if they are deformable one into the other.

[9] Any closed loop yields a linear endomorphism of the vector bundle, which only depends on the homotopy class of the loop. The bundle is said to be flat because loops that can be deformed to a point produce the identity endomorphism. If we had a non-trivial endomorphisms even for loops in small regions of X, then the bundle is said to have 'curvature'. Flat vector bundles are vector bundles with trivial curvature. For flat bundles, therefore, the only interesting endomorphisms come from going around big paths.

should be on the number field side by using the corresponding piece of geometry and making the translation of languages.[10] There is a whole dictionary. For example, the fundamental group of the Riemann surface corresponds to the Galois group of the number field and coverings correspond to field extensions.

We also remark that there is some analogy between these pictures and what we have been saying about class field theory. On both sides we have the abelian case and the non-abelian case, depending to which we want to take.

Geometric Langlands programme

The Langlands programme has, along these lines, something called the geometric Langlands programme, which replaces the number fields by Riemann surfaces. It is a very interesting theory which is much easier than the number field case, but not trivial and still quite big. It is developed by using the theory of vector bundles on Riemann surfaces. This theory has been going on for quite a long time, has nice results and in it there are geometrical analogues of the Langlands conjectures. There is also the intermediate case of the Langlands program on algebraic curves over finite fields from which we can transfer either to Riemann surfaces or to a number fields.

Very recently, in the last two years, Witten and his collaborators (mainly Kapustin and Gukov, two young Russians) have managed to deduce what is required for the geometric Langlands program from non-abelian dualities in physics. The kind of dualities they use are close to the dualities in Donaldson's theory and to the dualities in the mirror symmetry, and are based (at least) on the electric-magnetic duality.

The original dualities in physics are those between electricity and magnetism. They are very well understood and known, but it is a linear theory. When people started to play with non-linear theories, and tried to look for something analogous to duality, they had vague ideas that perhaps is would be possible to have some generalization of the electric-magnetic duality to the non-abelian case.

[10] The Weil conjectures are of this kind.

In the dualities it often happens that there is a pair of corresponding parameters, one on each side, that are inverse.[11] That means that if the parameter is large on one side, then it is small on the other side. In physics there are many theories that have a parameter called the *coupling constant*. If this parameter is small, then we can expand in a power series, and one gets a good approximation of the theory (this approach is called *perturbation theory*). If the parameter is large, this is hopeless. But if the theory can somehow be related to a dual theory with a small coupling constant, then of course we are in good business. We expand on the other side, then translate backwards and we get information about the theory with a large coupling constant, which is very difficult to get by other means. This was the dream that physicists had.

Could it be that the strong force, with a large coupling constant, and the weak force, with a weak coupling constant, are in some sense dual? There was lot of speculation along these lines and eventually, for example, the theories that Seiberg and Witten produced were of that kind.

Now the question of what Witten and his collaborators showed is that the geometric Langlands program for a Riemann surface can be reconstructed from one of the non-abelian dualities that involve basically electric-magnetic duality. They take a duality in 4 dimensions, where electricity and magnetism are given by a 2-form on each side, then they develop a non-abelian theory which is applied in the special case where the 4 dimensions is the product of the (curved) Riemann surface by 2 flat dimensions. In this situation they reconstruct interesting results about the Riemann surface and out of it they get the geometric Langlands program. Mirror symmetry of particular algebraic varieties plays a key role in this programme. The particular algebraic varieties they use are (Hitchin) moduli spaces of certain kinds of bundles on curves (bundles with some extra information called a Higgs field).

Altogether it is a great tour de force (see [K-W, G-W, F-W]) and one of the most exciting things happening in the last few years. It is a non trivial, very deep operation, and it is quite spectacular that the

[11] This phenomenon appear in ordinary Fourier theory, where the Fourier transform of the Gaussian e^{-a^2} is the Gaussian e^{-1/a^2}.

electric-magnetic duality should be behind the duality which is used in this Langlands programme, itself born out of number theory.

I must admit that many years ago, when I knew just a little bit of these ingredients, I made some speculations that perhaps the Langlands programme could be understood in terms of duality of electricity and magnetism. It was a good guess: 20 years later we are there.[12]

Modular forms

To finish with, let us go back to modular forms (at the beginning we mentioned the modular form of the Θ function, which corresponds to $SL(2,\mathbb{Z})$). Modular forms arise all over the place in mathematics, and in physics, and basically they are objects which transform nicely under certain arithmetic groups (there are large numbers of these groups; they are like groups of integer matrices).

Modular forms are the things which number theorists really play with. Much of the Langlands programme is a reinterpretation of questions about modular forms, which are a very key part of that theory. The framework of representations and so on that Langlands developed is a framework in order to handle modular forms. Results about modular forms have immediate applications and so they are really at the heart of the matter.

Interestingly enough, modular forms also frequently turn up in quantum field theories, usually in the form of what is called a *partition*

[12] Editorial note. "For a gauge group G, the GNO dual group is actually the same as the Langlands dual group ${}^LG$, which plays an important role in formulating the Langlands conjectures. [...] This was observed by Atiyah, who suggested to the second author [Witten] at the end of 1977 that the Langlands program is related to quantum field theory and recommended the two papers [10,12]." (Quoted from [K-W], p. 3). The two papers cited in this quote are: [10] P. Goddard, J. Nuyts, and D. I. Olive, *Gauge Theories And Magnetic Charge*, Nucl. Phys. B125 (1977) 1-28; and [12] C. Montonen and D. I. Olive, *Magnetic Monopoles As Gauge Particles?*, Phys. Lett. B72 (1977) 117-120.

function.[13] There are many examples of modular forms in the literature, but the reason why they exist, where they come from, is often not at all clear.

One of the big unsolved problems will be to understand the connection between number theory and physics which involves modular forms. That will be related to what I have been talking about, maybe much beyond that, and might need new ideas, but it is very concrete: we see modular forms created on both sides, one asks why, and the answer is that we don't know. But the evidence is there, it is very spectacular, and large parts of mathematics have been swept up by this kind of band-wagon of duality.

Duality is an old topic, but it is still very much alive and kicking. We have seen how it relates to many things that everyone is familiar with in mathematics (group theory, topology, analysis, Fourier theory), and so it is not surprising that it also arises in physics, where one can use the same sort of ideas. It is a very exciting story that hopefully the younger people here will continue by proving some theorems on modular forms for the future.

References

[FGV] E. Frenkel, D. Gaitsgory, and K. Vilonen: *On the geometric Langlands conjecture*. J. AMS, **15** (2001), 367-417.

[F-W] E. Frenkel, E. Witten: *Geometric endoscopy and mirror symmetry.* Preprint `arXiv:0710.5939v2 [Math AG] 14 Mar 2008`. 130 pp.

[G-W] S. Gukov, W. Witten: *Gauge theory, ramification, and the geometric Langlands program* Preprint `arXiv:hep-th/0612173v2 8 Oct 2007`. 160 pp.

[K-W] A. Kapustin and E. Witten: *Electric-Magnetic duality and the geometric Langlands program.* Preprint `arXiv:hep-th/0604151 10 Jan 2007`. 231 pp.

[Mor] John w. Morgan: *The Seiberg–Witten equations and applications to the topology of smooth four-manifolds.* Mathematical Notes 44, Princeton University Press, 1996.

[13] The notion of partition function comes from statistical mechanics, but it is basically the simplest invariant quantity that one can construct from quantum field theory. In particular, topological quantum field theories supply topological invariants.

[S-W] N. Seiberg and E. Witten: *Monopoles, duality and chiral symmetry breaking in $N = 2$ supersymmetric QCD*. Nuclear Phys. B.

[W94] E. Witten: *Monopoles and four-manifolds*. Math. Res. Lett. **1** (1994), 769-796.

[W08] E. Witten: *Mirror symmetry, Hitchin's equations, and Langlands duality*. Preprint `arXiv:0802.0999v1 [Math.RT] 7 Feb 2008`.