THE MATHEMATICAL STYLE OF MODERN PHYSICS*

N. MUKUNDA
Centre for Theoretical Studies, Indian Institute of Science, Bangalore 560 012, India.

ABSTRACT

Two important ingredients in the mathematical style of modern physics – the many roles of symmetry and the uses of unobservable quantities – are recounted and reviewed. Examples are taken from pre-relativistic and relativistic physics, particle and field mechanics, classical and quantum theory to illustrate these ideas. The words of many Masters illuminate our understanding of these concepts.

IT is a special privilege to speak to such a distinguished audience and I thank the Academy for giving me the chance on this occasion. Special because one finds representatives from all areas of science gathered at one place, to whom one has the opportunity to convey something significant from a particular area. For the same reason what I wish to present under the title 'The Mathematical Style of Modern Physics' will not be the latest technical advances in this field, but instead some characteristic features it has acquired over the past few decades and which are of course shared by recent developments. Naturally what I say must be taken as coming from one who constantly struggles to comprehend the Masters and who wishes to communicate his understanding to a wider audience.

A certain well-known book on mechanics describes physics as the science of measurement and change. In physics, as in other natural sciences, particular phenomena are isolated far enough to make precise observations and measurements, then models and theories are constructed in our minds to explain them and predict new phenomena. This involves relying on refined instruments of observation to aid our limited human senses, especially as we explore phenomena far removed from the human scale. Such instruments are of course based on previously understood phenomena and can be regarded as extensions of ourselves. The important point is that as we look at processes taking place at the microscopic or the macroscopic level, far smaller or far larger than ourselves, intuition gathered from everyday experience often fails as a guide to understanding. In its place we have to develop and rely on mathematics as our guide and make it into a sixth sense.

Mathematics is of course used and most effectively, also to describe phenomena on our own scale and it is easy to underestimate the difficulties faced in the past in the creation of new concepts. Be that as it may, it is generally agreed that with the developments of relativity and quantum theory the texture of theoretical physics has become much more subtle and abstract than might have been anticipated. This situation was described by Dirac in 1931 in these words:

"The steady progress of physics requires for its mathematical formulation a mathematics that gets continually more advanced. This is only natural and to be expected. What, however, was not expected by the scientific workers of the last century was the particular form that the line of advancement of the mathematics would take, namely it was expected that the mathematics would get more and more complicated, but would rest on a permanent basis of axioms and definitions, while actually the modern physical developments have required a mathematics that continually shifts its foundations and gets more abstract. Non-euclidean geometry and non-commutative algebra, which were at one time considered to be purely fictions of the mind and pastimes for logical thinkers, have now been found to be very necessary for the description of general facts of the physical world. It seems likely that this process of increasing abstraction will continue in the future and that advance in physics is to be associated with a continual modification and generalization of the axioms at the base of the mathematics rather than with a logical development of any one mathematical scheme on a fixed foundation".

This passage conveys most eloquently the changing relationship between mathematics and physics at the fundamental level. It can well be contrasted with, say, the situation in fluid dynamics where the basic

---

* Text of invited lecture at the 52nd Annual Meeting of Indian Academy of Sciences, Varanasi, November 8, 1986. Some of these ideas have been presented earlier to: Isituto Italiano per gli Studi Filosofici, Napoli, Italy, September 1983; Mathematical Association of India (Delhi Chapter), Delhi, April 1986; National Symposium on Theoretical Physics, C.M.S. College, Kottayam, Kerala, September 1986.

equations of Navier and Stokes have been known for a very long time and the problem lies in solving them under various conditions.

As parts of this changing style in which mathematical structures are used in physical theories, I would like to describe two sets of ideas today. One is the increasing importance of the ideas of symmetry and invariance; the other is the often unavoidable use of unobservable quantities in physical theories.

On the eve of his retirement from the Institute for Advanced Study, Hermann Weyl gave a set of lectures on Symmetry which have since become a classic. In it he says: 'Symmetry, as wide or as narrow as you may define its meaning, is one idea by which man through the ages has tried to comprehend and create order, beauty and perfection'. The subject of Weyl's discourse was symmetry in the static sense, the most immediate sense in which we all at first appreciate this notion. To say that an object is symmetric—such as a beautiful building or a well-grown crystal—is to say that it presents the same appearance before and after the application of certain transformations to it. These transformations are geometrical in character, being made up of rotations, reflections and translations; and the symmetry of an object is conveyed by the set of all transformations that leave it unchanged. The mathematical language to handle such static symmetry—static because time is not involved—is developed in Weyl's book and is the theory of finite and of discrete groups. But the focus of the present discussion is not the static symmetries of objects in space; rather it is the symmetries of physical laws describing processes taking place in space and time and to appreciate this requires some amount of abstraction. In Bargmann's words, ". . . . . those laws of physics which express a basic 'invariance' or 'symmetry' of physical phenomena seem to be our most fundamental ones".

Symmetry in this more fundamental sense operates at three levels which may be called the descriptive, the restrictive and the creative. To see this let us first recall with Wigner that there are three ideas of equal importance when discussing any set of physical laws; these are the laws themselves, then the allowed choices of initial conditions and finally the symmetries of the laws. Again as Wigner says, "The purpose . . . . . of all equations of physics is to calculate, from the knowledge of the present, the state of affairs that will prevail in the future". To begin with, let us consider such deterministic laws of motion alone. So they tell us, given some observed initial condition of a physical system, how the system evolves and what its condition is at all later times. Thus each solution of the equations determines one possible sequence of states in time, one history, corresponding to one choice of initial condition. In this context, a symmetry is an operation that leads us from one solution of the equations of motion to another generally different one. Such a symmetry is not a property of the condition of a physical system at an initial or any other time; rather it consists in the unchanging relationship at each time between the physical conditions on two different histories or solutions of the equations of motion. As opposed to static symmetry, this is a dynamical concept describing a property of the concerned physical laws and not of this or that state or condition. It is the equations that are preserved under the symmetry operation; this makes it somewhat abstract since the symmetry "cannot be seen by the eye but only by the mind".

In this sense one says that the equations of mechanics of Galileo-Newton are symmetric or invariant under the transformations of the Galilei group. Similarly the Maxwell equations of the Faraday-Maxwell theory of electromagnetism are symmetric under the Lorentz—or better Poincare—transformations. And these are the two prime instances of the descriptive role of symmetry, since it happened in both cases that the relevant equations were discovered well before the complete understanding of their respective symmetries (chart 1).

Descriptive Role of Symmetry

Galilean-Newtonian Mechanics: Galilei Group and Transformations

Faraday-Maxwell Electromagnetism: Poincare Group and Lorentz Transformations

---

(Chart 1)

However from the early years of this century came a shift of emphasis and a change to a new point of view, due principally to Poincare and Einstein. It arose from the realization that the Lorentz transformations and Lorentz invariance, though first seen in the context of Maxwell's equations, actually described general properties of space, time and measurement and so had a much wider significance. This led to the use of symmetry as a restrictive principle in the construction of new theories. In the words of Bargmann again, speaking of special relativity which governs space-time in the absence of gravitation: ". . . . . . every physical theory is supposed to conform to the basic relativistic principles and any concrete physical problem involves a synthesis of relativity and some specific physical theory".

Many striking examples of this restrictive role of symmetry are concerned with special relativity, some are in the framework of classical physics, others in connection with quantum theory and quantum mechanics and yet others with quantum field theory. It is well worth devoting a few minutes to quickly recounting them (chart 2)

Restrictive Role of Symmetry

Mass Energy Equivalence $E=mc^2$
Ten Conservation Laws
Dirac-Lorentz Equation

Sommerfeld Fine Structure Formula
Photon Momentum $P=E/c$
Planck's $E=h\nu$ to De Broglie's $P=hk$

Dirac Electron Equation
Weyl Neutrino Equation
Wigner Analysis of Elementary Systems

Fermi Weak Interaction Theory
Pauli Spin Statistics Theorem
Tomonaga Feynman Schwinger Renormalization Theory

(Chart 2)

The most famous classical result is perhaps the equivalence of mass and energy, $E=mc^2$; this came from amending the Galilean-Newtonian mechanics of material particles so that it too would share the Lorentz invariance of electromagnetism. Thus the two separate prerelativistic conservation laws of mass and energy were combined into one. More generally, special relativity or Lorentz invariance of a theory (almost) automatically ensures the ten basic conservation laws of energy, momentum, angular momentum and moment of energy. One of the most impressive uses of this was Dirac's 1938 treatment of the classical relativistic point electron; using essentially only the energy-momentum conservation laws he was able to obtain equations of motion, now called the Lorentz-Dirac equations, including the radiation-reaction terms. In the period of the old quantum theory, one can recall the use of special relativity by Sommerfeld in deriving the fine structure of the hydrogen spectrum To that same period also belongs the association of a momentum to a light quantum with the energy-momentum relation $E=pc$, which requires and can only be understood on the basis of special relativity. Slightly later, special relativity showed de Broglie the way to extend Planck's energy frequency relation $E=h\nu$ to his own momentum wave number relation $P=hk$ for material particles: thus he associated a relativistic wave with a moving particle, the particle properties of energy-momentum being proportional to the wave properties of frequency and wave number through Planck's constant. Turning to quantum mechanics, one has first the amazing discovery of the relativistic wave equation for the electron by Dirac in 1928. It came about by combining three elements—the general structure of quantum mechanics, the requirement of symmetry with respect to special relativity and the genius of Dirac—and it ended up explaining more things than its discoverer could have hoped for. the spin of the electron, its magnetic moment, the hydrogen fine structure, and the existence of the positron and antimatter. This last was of course a prediction and not an explanation. After this inauguration of relativistic quantum mechanics, one can mention Weyl's discovery of the wave equation for the massless neutrino; and somewhat later the analysis by Wigner of the quantum mechanical representations of the symmetry group of special relativity, which gave a systematic classification of all possible free relativistic systems. Finally in this recounting of the restrictive role of symmetry we have some instances from quantum field theory and elementary particle physics. Soon after Fermi constructed a theory of the weak interactions in 1934, it was seen that on the basis of special relativity there were five independent forms for this interaction. This was based on the assumption that space reflection was a symmetry of nature. After it was shown by Lee and Yang in 1956 that this was not a valid symmetry for weak processes, the number of forms of interaction allowed by relativity jumped to ten; but it was quickly reduced to one by the discovery in 1957 of the universal V-A interaction by Sudarshan and Marshak. This incidentally then led to a new symmetry called Chirality. In quantum field theory itself the remarkable connection between spin and statistics—the fact for instance that photons obey Bose statistics while electrons obey Fermi statistics—was shown by Pauli to be a consequence of relativity. In fact he concludes his paper on the subject with the words: "... we wish to state, that according to our opinion the connection between spin and statistics is one of the most important applications of the special relativity theory". Later in the 1940's relativistic invariance was one of the crucial guiding principles that enabled Tomonaga, Feynman and Schwinger to develop a consistent way to handle divergences and infinities in quantum field theory calculations, the renormalization theory, and thus to make meaningful predictions that could be compared with experiment.

These illustrative examples of the restrictive function of symmetry show the power and the fruitfulness of the point of view introduced by Poincare and Einstein in the early 1900's. It is by carrying these ideas to one higher level of sophistication—so to speak by pursuing them to their logical conclusion in various contexts—that one arrives at the creative role of symmetry (chart 3).

Creative Role of Symmetry

Abelian Gauge Invariance	→	Electrodynamics
General coordinate Transformation Invariance	→	General Relativity
Non Abelian Gauge Invariance	→	Yang Mills Theory

(Chart 3)

This is however quite a subtle step which has delicate connections with the second main idea I wish to present, namely the use of unobservable quantities in physical theories. Maxwell's electromagnetism is a relatively simple instance, while the general theory of relativity and the more recent nonabelian gauge theory are quite intricate instances, of this situation. Before going on to a description of these interrelationships, it may be well to recall the words of Dirac which motivate so beautifully the transition from the restrictive to the creative role of symmetry: "The growth of the use of transformation theory, as applied first to relativity and later to the quantum theory, is the essence of the new method in theoretical physics. Further progress lies in the direction of making our equations invariant under wider and still wider transformations".

Let me begin to describe the uses of unobservable quantities in physical theories, which occur at several levels, so that at a suitable level the interface with the creative function of symmetry can be brought in. The ideas are best conveyed through examples, the first of which is from the field of classical optics. If one takes a black and white photograph, say, one is making a record of the variation of the total intensity of light over the photographic film at a certain time. A colour photograph records the intensities of light at various frequencies. Now the fundamental theory of light at the classical level is given by the electromagnetic field equations of Maxwell. They tell us how from given initial conditions the electric and magnetic fields develop in the course of time. However the intensity of light involves essentially the sum of the squares of the electric and magnetic fields; and it is *not true* that if we knew the initial distribution of light intensity, say in some region of space, we could predict it elsewhere or at a later time. If we had provisionally defined the intensity of light to be the only observable quantity in optics, then in order to see how intensity changes with space and time; we would have been forced to introduce something called the two-point correlation function—an unobservable quantity at this level—and express the laws of evolution in terms of it. The two-point function is a measure of the correlation between the electromagnetic field at one point of space at one time and at another point of space at a possibly different time. It is of the same mathematical nature as, but physically distinct from, the light intensity. The Maxwell equations for the electric and magnetic fields lead to definite laws of propagation for the two-point function, but the intensity being a particular case of the two-point function does not obey any propagation law on its own. Once one admits that the Maxwell fields are observable, then so is the correlation function. This example is in a sense rather elementary since what is initially regarded as unobservable becomes, in a wider framework and with better understanding, an observable quantity.

Our next and less trivial, example concerns electromagnetism again but now assuming that the electric and magnetic fields are—at least classically—observable. In the presence of classical charged particles, the combined system of Maxwell's equations for the field strengths and Lorentz's equations for the particles involve observable quantities only—field strengths on the one hand, particle positions on the other. The system is deterministic in the sense assumed earlier and is also local. In practical calculations one finds it convenient to express the field strengths in terms of an auxiliary quantity called the vector potential. However the potential is in principle unobservable because there are transformations or changes in the potential—gauge transformations as they were called by Weyl—which do not change the observable field strengths at all. Quite generally, even in other contexts, gauge transformations are transformations which vary continuously but arbitrarily from point to point in space time, staying of course within a given class; and those quantities which do change under a gauge transformation are unobservable. As a result, the equations for the potential cannot determine it completely since they must allow for an arbitrary gauge transformation; but this causes no problem since the potential was introduced for convenience only and can be dispensed with. But the situation changes when the charged particles are subject to the laws of quantum mechanics, assuming for the moment that the field is classical and externally given. The quantum equation of motion for the particles, the Schrodinger equation, uses the vector potential in an essential way. In quantum theory it is much more awkward to eliminate the unobservable vector potential than in classical theory. One can do so and it has been done not only for the case considered but also for the complete system of quantized matter and Maxwell fields, using a method due to Dirac and Mandelstam. But one then has to work with nonlocal quantities and equations—quantities depending not just on a point in space-time but on an arbitrary path leading up to that point. If one is prejudiced in favour of locally defined quantities and equations, one has to use the unobservable vector potential with the associated freedom of gauge transformations.

The third example concerns general relativity. The original way in which the equations of this theory were derived and presented depended very heavily on the invariance requirements placed upon them. These requirements were strong enough to almost determine the equations—the creative role of symmetry. One considers events taking place in space and time and describes them with the help of space and time coordinates. The essential point now is that one allows a great deal of freedom in the assignment of coordinates to events and demands that the equations of the theory must retain their form under any changes of coordinates. This requirement of symmetry makes the coordinates really unobservable. In the words of Wigner: "The basic premise of this theory is that coordinates are only auxiliary quantities which can be given arbitrary values for every event.... coordinates are only labels to specify space-time points. Their values have no particular significance unless the coordinate system is somehow anchored to events in space-time". Now-a-days relativists use the term "coordinate markers" to convey this quality of coordinates and compare the situation to a telephone directory; indeed one of the best known books on the subject *is* a telephone book. As long as one retains the freedom to make arbitrary changes of coordinates, they cannot be anchored to space-time events in any way and so remain unobservable. Of course in recent times more refined mathematical methods have been brought in to formulate the laws of general relativity in what is called an intrinsic coordinate free description, thus eliminating the unobservable coordinates altogether. Nevertheless the problem of deciding what mathematical quantities qualify as observables remains tricky and has no easy answers.

The nonabelian gauge theories discovered by Yang and Mills in 1954—and which are basic to the unification of electromagnetism and the weak interactions and also to the currently accepted theory of nuclear forces—stand midway between electromagnetism and general relativity in complexity. The arbitrary space-time dependent transformations now do not act on the space-time coordinates but in an internal space describing properties which are a generalization of electric charge. Once again there is a vector potential which changes under these transformations, but it is more intricate than in the case of electromagnetism since now even the analogues of electric and magnetic fields change when the potential changes. This makes both the potentials and the field strengths unobservable. Here again the increased demands of symmetry are powerful enough to almost determine the basic equations; the difference is that now the analogues of the Maxwell equations involve the potential in an essential way. The problem of constructing observables is somewhat more easily solved here than in the case of relativity, while the nonlocality involved in trying to express everything in terms of them is more severe than in the electromagnetic case.

At this stage some general comments connecting the creative function of symmetry to the use of unobservable quantities can be made.

	. Nonobservables:
Gauge Symmetry →	. Nondeterministic Equations.
	. Restriction on initial Conditions

At any rate at a classical level one can say that in a theory without any symmetry of the gauge type, such as Galilean Newtonian mechanics or Maxwell-Lorentz electrodynamics not using the potential, all quantities in the theory are in principle observable and the basic laws can be expected to be deterministic. However in the presence of a gauge type symmetry, three related things happen: those quantities which change under the transformations must be regarded as unobservable; because of the arbitrary elements in these transformations the equations of motion cannot be fully deterministic; and on the technical side restrictions emerge on the allowed initial conditions. In terms of the three components involved in the discussion of any set of physical laws—the laws themselves, the possible initial data and the symmetries—it means that an increase of the third component to gauge type symmetries has important repercussions on the first two components. If the freedom to perform gauge transformations is maintained, one has local quantities obeying local but not completely deterministic equations; if one wants to work with observable quantities alone, some degree of nonlocality is unavoidable. Conversely the fully local description will involve some unobservable quantities.

The creative uses of symmetry in both general relativity and nonabelian gauge theory give to these theories a strongly geometric flavour. One is reminded of Klein's well-known Erlangen Program and gets the feeling that physics is being geometrized or becoming geometry. What saves the situation is that, as Regge said, physics is not geometry but geometry plus an action principle. Hence the statement made more than once earlier that gauge type symmetry almost completely determines the form of the basic equations, but not quite.

While unobservable quantities seem to be closely related to local symmetries at the classical level, this connection is weakened in quantum theory, which is the fourth and last of our examples. In some respects the situation is similar to that of classical optics except that it is not provisional. According to quantum mechanics not all the physical quantities associated with an atomic system can be simultaneously measured or specified as numbers. There are definite limitations on the amount of "information" we can have about an atomic system at one time. If by means of a measurement one has obtained maximal permitted information at a certain time, that can be mathematically represented by something called a wave function. The basic laws of quantum mechanics then determine how the wave function varies with time and at that level things are deterministic. However the wave function itself is unobservable. At each time the wave function determines the probabilities for various outcomes of various experiments that may be performed at that time and these probabilities are essentially quadratic in the wave function. Thus the observable quantities are essentially these probabilities, but there is no way to directly calculate how they change and evolve in time. There is no way of avoiding the use of the unobservable wave function or something essentially like it so as to be able to express all the features of quantum phenomena.

This discussion of the uses of unobservable quantities in physical theories shows that the rule of three operates here just as in so many other contexts.

The Rule of Three

Physical Laws	→	. Fundamental Equations . Initial Conditions . Symmetries
Symmetries	→	. Descriptive . Restrictive . Creative
Nonobservables	→	. Provisional, Temporary . Convenience, Locality; Avoidable with effort . Essential, Unavoidable

Thus such quantities may appear in a provisional and temporary sense alone; or they may be used as a matter of convenience, it being a matter of lesser or greater difficulty to dispense with them; or finally they may be essential and unavoidable. If one has not come across any of these possibilities, one may feel that there is something strange or even alarming in nonobservable quantities playing such an important role in physical theory. But one can take comfort in the words of Max Planck: "It is absolutely untrue, although it is often asserted, that the world picture of physics contains, or may contain, directly observable magnitudes only"; and in Richard Feynman's reassurance: "It is not true that we can pursue science completely by using only those concepts which are directly subject to experiment. In quantum mechanics itself there is a probability amplitude, there is a potential and there are many constructs that we cannot measure directly......, It is absolutely necessary to make constructs". This suggests that these ideas have a wider range of relevance than just physics and one also recalls Einstein's advice to Heisenberg: "It is never possible to introduce only observable quantities in a theory. It is the theory which decides what can be observed".

It has been said that each generation of physicists feels that the next generation is too mathematical. Why is this so and why does physical theory get more and more abstract as it develops? One can do no better than quote Dirac in answer: "The methods of progress in theoretical physics have undergone a vast change during the present century. The classical tradition has been to consider the world to be an association of observable objects (particles, fluids, fields, etc) moving about according to definite laws of force, so that one could form a mental picture in space and time of the whole scheme. This led to a physics whose aim was to make assumptions about the mechanism and forces connecting these observable objects, to account for their behaviour in the simplest possible way. It has become increasingly evident in recent times, however, that nature works on a different plan. Her fundamental laws do not govern the world as it appears in our mental picture in any very direct way, but instead they control a substratum of which we cannot form a mental picture without introducing irrelevancies". What a contrast to Lord Kelvin's statement from the last century that "It seems to me that the test of 'Do we or do we not understand a particular point in physics?' is 'Can we make a mechanical model of it?'". Far from this, it has become increasingly necessary to rely on our feeling for the abstract and on our mathematical sensibilities in trying to comprehend the developing physical picture of nature. And though I have quoted from many Masters, it seems that more than anyone else the writings of Dirac express beautifully the style of, and his works have contributed a great part of the content of, the changing mathematics that underlies modern theoretical physics.

1. Bargmann, V., *Rev. Mod. Phys.*, 1957, **29**, 161
2. Dirac, P. A. M., *The principles of quantum mechanics*, 1958, Clarendon Press, Oxford; *Proc. R. Soc. (London)*, 1931, **A133**, 50; 1938, **A167**, 148.
3. Feynman, R. P., *Lectures on physics*, Addison-Wesley, New York, 1963, Vol. 1, Ch. 38.
4. Planck, M., *Scientific autobiography and other papers*, Greenwood Press, Westport, Connecticut, 1971.
5. Weyl, H., *Symmetry*, Princeton University Press, Princeton, 1952.
6. Wigner, E. P., *Symmetries and reflections*, Indiana University Press, Bloomington, 1967.
7. Yang, C. N., 'Geometry and Physics', in 'To fulfill a vision: Jerusalam Einstein Centennial Symposium (ed.) Yuval Ne'eman, Addison-Wesley, New York, 1981.

ANNOUNCEMENT

THIRD NATIONAL SYMPOSIUM ON BIO-ORGANIC CHEMISTRY

The Third National Symposium on Bio-organic Chemistry will be held during July 9–11 1987, under the joint auspices of the Centre for Cellular and Molecular Biology, Regional Research Laboratory, and the Department of Biochemistry, Osmania University. This symposium is held once in two years. Topics to be covered in the symposium are: i) Lipids, Biomembranes, Chemical messengers and Receptors, ii) Biotechnology, Immunochemistry Enzymes and related topics, iii) Biotransformations, Biosynthesis and Biommetic Chemistry, iv) Synthetic methods in Peptides, Oligonucleotides and Carbohydrates, v) Macromolecular interactions and Structural methods in Bioorganic Chemistry.

The scientific programme will consist of plenary lectures and half-an hour oral presentations by invited speakers. The Proceedings of the Symposium will appear as a special issue of the Journal of Biosciences, published by the Indian Academy of Sciences, Bangalore.

One of the objectives of the symposium is to generate and promote interest in areas related to bio-organic chemistry among young scientists. It is planned to select twenty-five young and promising research scholars and teachers in national laboratories and for participation in the symposium. The partial or complete expenses on participants under this category will be borne by the organizers, depending on the availability of funds.

Further particulars may be had from: Dr K. N. Ganesh, Convenor, Third National Symposium on Bio-organic Chemistry, Centre for Cellular and Molecular Biology, Hyderabad 500 007.