

## I

*Introduction*

Not enough has been written about the philosophical problems involved in the application of mathematics, and particularly of group theory, to physics. On the one hand, mathematics is created to solve specific problems arising in physics, and, on the other hand, it provides the very language in which the laws of physics are formulated. One need only think of calculus or of Fourier analysis as examples of this dual relationship.

We are all familiar with the exploitation of symmetry in the solution of a mathematical problem. On the other hand, the very assertion of symmetry is often the most profound formulation of a physical law or the key step in the development of a new theory.

Another example is provided by Hamiltonian mechanics. The mathematical theory underlying Hamiltonian mechanics is currently called symplectic geometry. We briefly recall the basic definitions and early history. A symplectic vector space is a real vector space equipped with an antisymmetric, nondegenerate bilinear form. For example, on  $\mathbb{R}^2$  we can define the form  $(, )$  by  $(u_1, u_2) = q_1 p_2 - q_2 p_1$ , where  $u_1 = \begin{pmatrix} q_1 \\ p_1 \end{pmatrix}$  and  $u_2 = \begin{pmatrix} q_2 \\ p_2 \end{pmatrix}$ . It is not hard to prove that such a vector space must be even-dimensional (if finite-dimensional) (see Section 21 below). A linear isomorphism of a symplectic vector space  $V$  (or more generally of  $V$  onto  $W$ ) is called symplectic if it preserves the bilinear form. (In two dimensions, a linear transformation is symplectic if and only if its determinant is 1. In higher dimensions, the condition is more restrictive.) A differentiable map of an open subset of a symplectic vector space  $V$  into  $W$  is called symplectic if its (Jacobian) matrix of partial derivatives is symplectic at every point. A symplectic manifold  $M$  is an even-dimensional manifold that locally has the

structure of a symplectic vector space. This means that one has local charts  $\psi_i$  mapping open sets  $U_i$  of  $M$  onto open subsets of some fixed symplectic vector space  $V$  and such that the change of coordinates maps  $\psi_i \circ \psi_j^{-1}$  defined on  $\psi_j(U_i \cap U_j)$  are symplectic. (Alternatively, thanks to a theorem of Darboux (see Chapter II, Section 22), a symplectic manifold is a manifold together with a closed 2-form of maximal rank.) One has the obvious definition of symplectic diffeomorphism (i.e., one-to-one smooth transformations with smooth inverses, which are locally symplectic in the above sense). In the older literature, symplectic diffeomorphisms were called canonical transformations. Symplectic geometry is the study of symplectic manifolds and diffeomorphisms. The relation with mechanics is usually expressed by saying that the *phase space* of a mechanical system is a symplectic manifold and that the time evolution of a (conservative) dynamical system is a one-parameter family of symplectic diffeomorphisms. The role of the symplectic structure had first appeared, at least implicitly, in Lagrange's work on the variation of the orbital parameters of the planets in celestial mechanics. Its central importance, emerged, however, from the work of Hamilton.

At the age of eighteen, Hamilton submitted a paper entitled "Caustics" to Dr. John Brinkley, then the first royal astronomer for Ireland, who, as a result, is said to have remarked "This young man, I do not say *will be* but *is* the first mathematician of his age." Brinkley presented the paper to the Royal Irish Academy. It was referred, as usual, to a committee, whose report, while acknowledging the novelty and value of its contents recommended that it should be further developed and simplified before publication. Five years later, in greatly expanded form, the paper finally appeared, entitled "Theory of systems of rays," published in 1828 in the Transactions of the Royal Irish Academy.

The gist of Hamiltonian optics, in modern language is as follows: One is interested in studying the geometry of rays of light as they pass through some optical system. Suppose our system is aligned along some axis, and we study rays that enter the system at the left and emerge from the right. The portion of the rays to the left of the system are straight line segments. One needs four variables, locally, to specify such a line – two variables to specify the point of intersection of the line with a plane perpendicular to the optical axis and two additional angular variables giving the inclination of the line to this plane. The problem is to relate the incoming line segments, to the left of the system, to the outgoing line segments to the right. The first basic assertion is that if we use the proper coordinates (which involve the index of refraction of the ambient space), the transformation from the incoming to the outgoing coordinates is a symplectic diffeomorphism. Thus geometrical

optics is reduced to symplectic geometry. Hamilton showed that if the graph of a symplectic transformation satisfies an appropriate “transversality” condition, then the transformation determines and is determined by a function of one-half the number of incoming and outgoing variables – the so-called generating function of the symplectic transformation. As this function is determined solely by the physical properties of the optical system, Hamilton called it a characteristic function. Depending on the transversality assumptions made, it can be a function of the points of intersection of the incoming and outgoing rays with the transversal planes – the point characteristic, the incoming points of the intersection and the outgoing angles – the mixed characteristic, or the incoming and outgoing angles – the angle characteristic. These functions are of use in combining optical systems, that is, in composing the corresponding symplectic transformations. They are also extremely useful in describing the deviation of the symplectic transformations from linearity – the *geometric aberrations* of the optical system. Finally, they are closely related to the *optical length* of the light rays themselves, and these light rays can be characterized as being extremals for optical length (Fermat’s principle).

Some years later, Hamilton realized that this same method applies unchanged to mechanics: Replace the optical axis by the time axis, the light rays by the trajectories of the system, and the four incoming and four outgoing variables by the  $2n$  incoming and outgoing variables of the phase space of the mechanical system. Hamilton’s methods, as developed by Jacobi and other great nineteenth-century mathematicians, became a powerful tool in the solution or analysis of mechanical problems. Hamilton’s analogy between optics and mechanics served as a guiding beacon to the development of quantum mechanics some one hundred years later.

The main purpose of this chapter is to discuss the relation between linear optics, geometric optics, and wave optics, stressing Hamilton’s point of view and the corresponding relations between classical and quantum mechanics.

We first make a few comments on the relation between various theories of optics. In the history of physics it is often the case that when an older theory is superseded by a newer one, the older theory still retains its validity – either as an approximation to the newer theory, an approximation that is valid for an interesting range of circumstances, or as a special case of the newer theory. Thus Newtonian mechanics can be regarded as approximation to relativistic mechanics, valid when the velocities that arise are very small in comparison to the velocity of light. Similarly, Newtonian mechanics can be regarded as an approximation to quantum mechanics,

valid when the bodies in question are sufficiently large. Kepler's laws of planetary motion are a special case of Newton's laws, valid for the inverse-square law of force between two bodies. Kepler's laws can also be regarded as an approximation to the laws of motion derived from Newtonian mechanics when we ignore the effects of the planets on one another's motion.

The currently held theory of light is known as quantum electrodynamics. It describes very successfully and very accurately the interaction of light with charged particles, explaining both the discrete character of light, as evinced in the photoelectric effect, and the wavelike character of electromagnetic radiation. The triumph of nineteenth-century physics was Maxwell's electromagnetic theory, which was a self-contained theory explaining electricity, magnetism, and electromagnetic radiation. Maxwell's theory can be regarded as an approximation to quantum electrodynamics, valid in that range where it is safe to ignore quantum effects. Maxwell's theory fails to explain a whole range of phenomena that occur at the atomic or subatomic level.

One of Maxwell's remarkable discoveries was that visible light is a form of electromagnetic radiation, as is "radiant heat." In fact, since Maxwell, optics is a special chapter of the theory of electricity and magnetism that treats electromagnetic vibrations of all wavelengths: from the shortest  $\gamma$  rays of radioactive substances (having a wavelength of one hundred millionth of a millimeter) up through x rays, ultraviolet, visible light, and infrared to the longest radio waves (having a wavelength of many kilometers).

Maxwell's theory dealt with the source of electromagnetic radiation as well as its propagation. Before Maxwell, there was a fairly well developed wave theory of light, due mainly to Fresnel, which dealt rather successfully with the propagation of light in various media, but had nothing to say about the production of light. Fresnel's theory did account for three physical effects that could not be explained by earlier theories – diffraction, interference, and polarization. Diffraction has to do with the behavior of light in the immediate vicinity of surfaces through which it is transmitted or reflected. A typical diffraction effect is the fact that we cannot produce an absolutely straight, arbitrarily narrow beam of light. For example, we might try to produce such a beam by lining up two opaque screens with holes in them to collimate light arriving from the left, as shown in Figure 1.1. When the holes are made very small (of the order of the wavelength of the light) we find that the region to the right of the second screen is suffused with light, instead of there being a narrow beam. Interference refers to those phenomena where the wave character of light

Cambridge University Press

978-0-521-38990-7 - Symplectic Techniques in Physics

Victor Guillemin and Shlomo Sternberg

Excerpt

[More information](#)

## I. Introduction

5

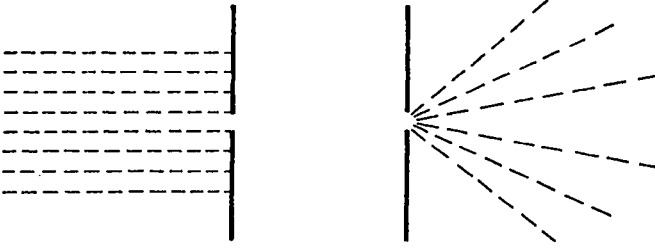


Figure 1.1.

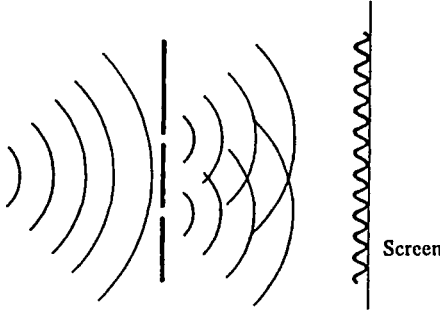


Figure 1.2.

manifests itself by the constructive or destructive superposition of light traveling different paths. Typical is the famous Young interference experiment illustrated in Figure 1.2. Polarization refers to the fact that when light passes through certain materials it appears to acquire a preferred direction in the plane perpendicular to the ray; such effects can be observed, for example, by using Polaroid filters.

Geometrical optics is an approximation to wave optics in which the wave character of light is ignored. It is valid whenever the dimensions of the various apertures are very large when compared to the wavelength of the light and when we do not examine too closely what is happening in the neighborhood of shadows or foci. It does not account for diffraction, interference, or polarization.

Linear optics is an approximation to geometrical optics which is valid when the various angles that enter into consideration are small. In linear optics one makes the approximation  $\sin \theta \doteq \theta$ ,  $\tan \theta \doteq \theta$ ,  $\cos \theta \doteq 1$ , etc.; that is, all quadratic (or higher-order) expressions in the angles are ignored. For example, in geometrical optics, *Snell's law* says that if light passes from a region whose index of refraction is  $n$  into a region whose index of refraction is  $n'$ , then  $n \sin i = n' \sin i'$ , where  $i$  and  $i'$  are the angles that the

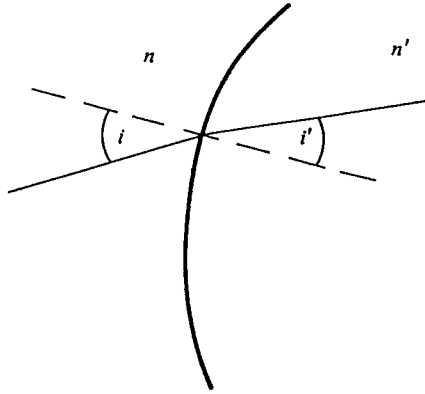


Figure 1.3.

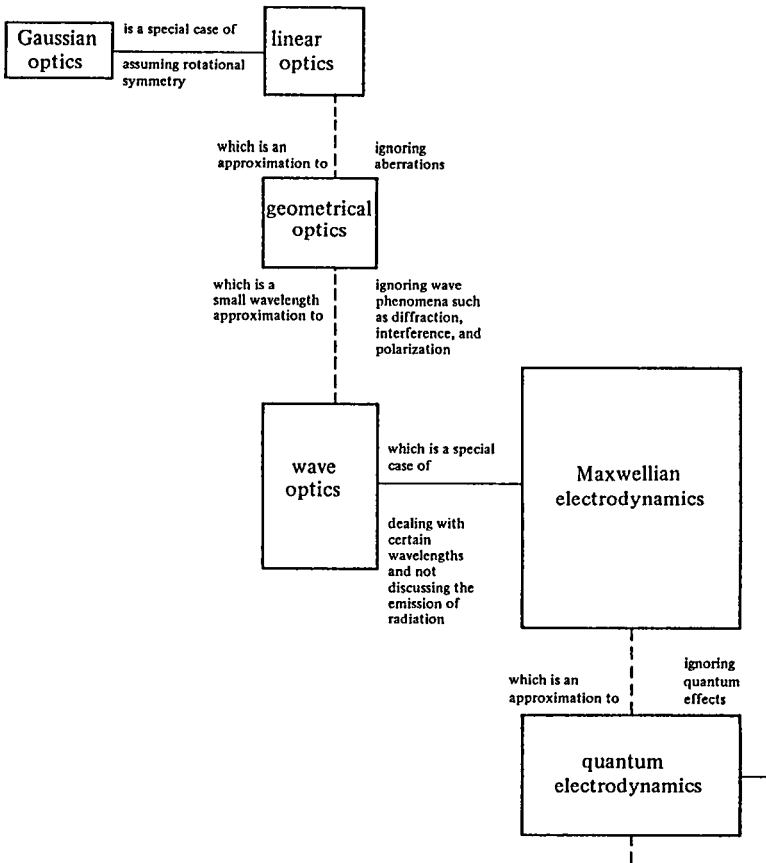


Figure 1.4.

light ray makes with the normal to the surface separating the regions (Figure 1.3). In linear optics, we replace this law by the simpler law  $ni = n'i'$ , which is a good approximation if  $i$  and  $i'$  are small. (This approximate law was known to Ptolemy.) The deviations of geometrical optics from the linear-optics approximation are known as (geometrical) aberrations.

Gaussian optics is a special case of linear optics in which it is assumed that all the surfaces that enter are rotationally symmetric about a central axis. This is a very important special case, since all ground lenses and most polished mirrors have this property. We summarize our discussion in Figure 1.4.

### 1. Gaussian optics

In linear optics we will make the approximation that the index of refraction is constant between refracting surfaces. (In our study of geometrical optics we will make the opposite approximation: that the index of refraction is a smoothly (possibly rapidly) varying function.) In Gaussian optics we are interested in tracing the trajectory of a light ray as it passes through the various refracting surfaces of the optical system (or is reflected by reflecting surfaces). We introduce a coordinate system so that the  $z$  axis (pointing from left to right in our diagram) coincides with the optical axis (i.e., the axis of symmetry of our system).

We shall study rays which are coplanar – ones that lie in one plane with the  $z$  axis; we shall prove later on that, in this approximation, the study of rays that do not lie in a single plane with the optical axis can be reduced to the study of coplanar rays.

By rotational symmetry, it is clearly sufficient to restrict attention to rays lying in one fixed plane. The trajectory of a ray as it passes through apertures which are large compared to the wave length consists of pieces of straight lines. Our problem is to relate the straight line of the ray after it emerges from the system to the straight line that entered. For this we need to have a way of specifying straight lines. We do so as follows: We choose some fixed  $z$  value. (This amounts to choosing a plane perpendicular to the optical axis, called the *reference plane*.) Then a straight line is specified by two numbers, its height  $q$  above the axis at  $z$ , and the angle  $\theta$  that the line makes with the optical axis. The angle will be measured in radians and will be considered positive if a counterclockwise rotation carries the positive  $z$  direction into the direction of the ray along the straight line (see Figure 1.5).

It is convenient to choose new reference planes, suitably adjusted to each stage in the calculation. Thus, for example, if light enters our optical system

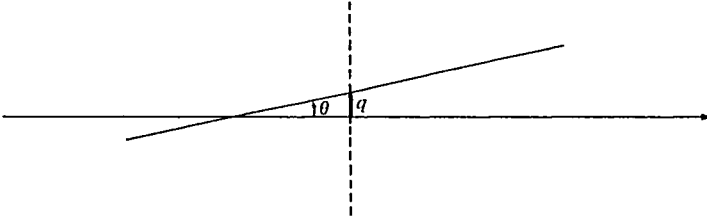


Figure 1.5.

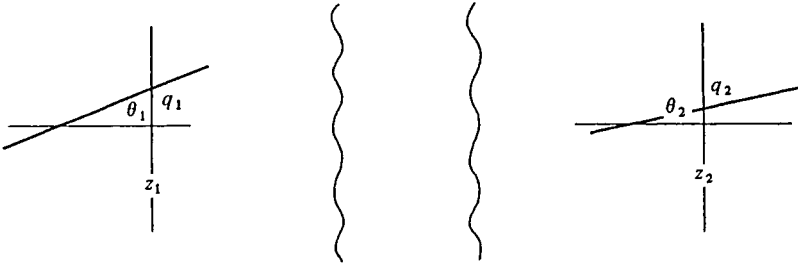


Figure 1.6.

from the left and emerges from the right, we would choose one reference plane  $z_1$  to the left of the system of lenses and a second reference plane  $z_2$  to the right. A ray enters the system as a straight line specified by  $q_1$  and  $\theta_1$  at  $z_1$  and emerges as a straight line specified by  $q_2$  and  $\theta_2$  at  $z_2$  (Figure 1.6). Our problem, for any system of lenses, is to find  $(q_2, \theta_2)$  as a function of  $(q_1, \theta_1)$ .

Now comes a simple but crucial step of far-reaching significance that is basic to the geometry of optics and mechanics: Replace the variable  $\theta$  by  $p = n\theta$ , where  $n$  is the index of refraction of the medium at the reference plane. (In mechanics, the corresponding step is to replace velocity by momentum.)

We thus describe a light ray by the vector  $\begin{pmatrix} q \\ p \end{pmatrix}$ , and our problem is to find  $\begin{pmatrix} q_2 \\ p_2 \end{pmatrix}$  as a function of  $\begin{pmatrix} q_1 \\ p_1 \end{pmatrix}$ . Since we are ignoring all terms quadratic or higher, it follows from our approximation that  $\begin{pmatrix} q_2 \\ p_2 \end{pmatrix}$  is a linear function of  $\begin{pmatrix} q_1 \\ p_1 \end{pmatrix}$ ; that is, that

$$\begin{pmatrix} q_2 \\ p_2 \end{pmatrix} = M_{21} \begin{pmatrix} q_1 \\ p_1 \end{pmatrix}$$

1. Gaussian optics

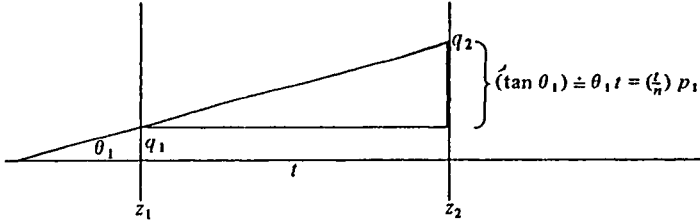


Figure 1.7.

for some matrix  $M_{21}$ . The key effect of our choice of  $p$  instead of  $\theta$  as variable, is the assertion that

$$\det M_{21} = 1;$$

in other words, that the study of Gaussian optics is equivalent to the study of the group  $Sl(2, \mathbb{R})$  of  $2 \times 2$  real matrices of determinant 1. To prove this, observe that if we have three reference planes,  $z_1$ ,  $z_2$ , and  $z_3$ , situated so that a light ray going from  $z_1$  to  $z_3$  passes through  $z_2$ , then by definition

$$M_{31} = M_{32}M_{21}.$$

Thus, if our optical system is built out of two components, we need only verify  $\det M = 1$  for each component separately. To simplify the exposition we assume that our system does not contain mirrors. Then any refracting lens system can be considered as the composite of several systems of two basic types:

(a) A translation, in which the ray continues to travel in a straight line between two reference planes lying in the same medium. To describe such a system we must specify the gap  $t$  between the planes and the refractive index  $n$  of the medium (Figure 1.7). It is clear for such a system that  $\theta$  and hence  $p$  does not change, and that  $q_2 = q_1 + (t/n)p_1$ . We write  $T = t/n$  (called the *reduced distance*) and see that

$$\begin{pmatrix} q_2 \\ p_2 \end{pmatrix} = \begin{pmatrix} 1 & T \\ 0 & 1 \end{pmatrix} \begin{pmatrix} q_1 \\ p_1 \end{pmatrix},$$

$$\det \begin{pmatrix} 1 & T \\ 0 & 1 \end{pmatrix} = 1.$$

(b) Refraction at the boundary surface between two regions of differing refractive index. We must specify the curvature of the surface and the two indices of refraction,  $n_1$  and  $n_2$ . The two reference planes will be taken immediately to the left and immediately to the right of the surface, respectively.

Cambridge University Press

978-0-521-38990-7 - Symplectic Techniques in Physics

Victor Guillemin and Shlomo Sternberg

Excerpt

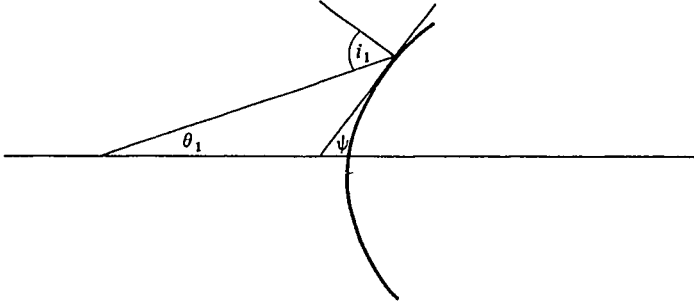
[More information](#)

Figure 1.8.

At such a surface of refraction the  $q$  value does not change. The angle and hence the value of  $p$  change according to (the linearized version of) Snell's law. Now Snell's law involves the slope of the tangent to the surface at the point of refraction. In our approximation we are ignoring quadratic terms in this slope and hence terms of degree 3 or higher in the surface. We may thus assume that the curve giving the intersection of this surface with our plane is a parabola:

$$z - z_1 = \frac{1}{2}kq^2.$$

Then the derivative of  $z$  with respect to  $q$  is  $z'(q) = kq$ , which is  $\tan(\pi/2 - \psi)$ , where  $\psi$  is the angle in the figure. For small angles  $\theta$ , that is, for small values of  $q$ ,  $\psi$  will be close to  $\pi/2$ , and hence we may replace  $\tan(\pi/2 - \psi)$  by  $\pi/2 - \psi$ , if we are willing to drop higher-order terms in  $q$  or  $p$ . Thus  $\pi/2 - \psi = kq$  in our Gaussian approximation. On the other hand, if  $i_1$  denotes the angle that the incident ray makes with this tangent line (Figure 1.8), then the fact that the sum of the interior angles of a triangle adds up to  $\pi$  shows that  $(\pi - \psi) + \theta_1 + (\pi/2 - i_1) = \pi$  or

$$i_1 = \theta_1 + kq,$$

and similarly,

$$i_2 = \theta_2 + kq,$$

where  $q = q_1 = q_2$  is the point where the rays hit the refracting surface. Multiplying the first equation by  $n_1$  and the second equation by  $n_2$  and using Snell's law in the approximate form  $n_1 i_1 = n_2 i_2$  gives

$$\begin{pmatrix} q_2 \\ p_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -P & 1 \end{pmatrix} \begin{pmatrix} q_1 \\ p_1 \end{pmatrix},$$

where  $P = (n_2 - n_1)k$  is called the *power* of the refracting surface.