Geometry and dynamics in refracting systems

This article has been downloaded from IOPscience. Please scroll down to see the full text article.
1995 Eur. J. Phys. 16 14
(http://iopscience.iop.org/0143-0807/16/1/003)

View the table of contents for this issue, or go to the journal homepage for more

Download details:
IP Address: 132.248.33.126
The article was downloaded on 27/06/2011 at 21:00

Please note that terms and conditions apply.
Geometry and dynamics in refracting systems*

Kurt Bernardo Wolf and Guillermo Krötzsch

Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas-Cuernavaca, Universidad Nacional Autónoma de México, Apartado Postal 139-B, 62191 Cuernavaca, Morelos, Mexico

Received 14 March 1994, in final form 27 May 1994

Abstract. The geometric and dynamic postulates for rays in inhomogeneous optical media lead succinctly to the two Hamilton equations in regions where the inhomogeneity is smooth; at a surface of discontinuity between two smooth media, they lead to two conservation laws. One of these is the Ibn Sahl (=Snell-Descartes) law of finite refraction. The transformation due to finite refraction can in general be factorized into two simpler root transformations. These conclusions apply for mechanical as well as optical systems.

1. Introduction and purpose

Modern mathematical physics has taught us to distinguish between the geometric and dynamic foundations of evolving systems. In this vein, we propose two fundamental postulates for the lines that represent light rays in classical optics. In regions where the medium is smoothly inhomogeneous, the two Hamilton equations result. On surfaces where the medium has a finite discontinuity, conservation laws apply, of which one has been known as Snell's law [1]. Recent historical research [2], however, shows that the law of refraction was clearly known to Abū Saʿād al-ʿAlāʾ Ibn Sahl before the year 984. Apart from reflection—known to the Greeks—refraction is probably the oldest dynamic law formulated for Nature.

In section 2 we present a pair of geometric and dynamic postulates for the geometric optics model of light phenomena. Theoretical models of physical phenomena are based on 'self-evident' mathematical statements, as mathematical theories are based on axioms. Geometric optics—like point-particle mechanics—is one of several models of optical phenomena, particularly useful for instrument design. It renounces ab initio to explain polarization, interference, or quantum coherence effects, in the same way that Newtonian mechanics explains a world with no relativity or quantum problems.

We propose as self-evident two postulates for the lines in space that model the classical optic rays, and write the formulae in incremental form, for small, finite distances on the lines. In this way, letting the increment vanish, we obtain the two Hamilton evolution equations for general inhomogeneous, but smooth optical media. If the medium has a finite discontinuity across a (smooth) surface, the incremental form leads to two conservation laws in refraction, one being the 'Snell' law of sines—or 'Snell-Descartes' as it is sometimes known. Ibn Sahl's paternity of the law is briefly discussed.

Section 3 distinguishes optical rays from point-particle trajectories by an additional postulate regarding the physical interpretation of the time parameter. The Hamilton equations establish a correspondence between optical media with smooth refractive indices and smooth mechanical potentials. Section 4 follows this correspondence for finite refraction in optics and point-particle fall over potential steps in mechanics. It has been proven...
in [3] that the transformation of rays due to a refracting surface between two homogeneous media can be decomposed into a product of two root transformations, each factor containing the parameters of a single medium. The derivation of this property, as it appears in [4], is generalized for an arbitrary finite discontinuity between smooth inhomogeneous media. The root transformation has been particularly useful for symbolic computer aberration calculations, as we remark in the concluding section.

2. Postulates and equations

Geometric optics treats light rays as lines in three space dimensions. We describe a line by a 3-vector of functions \( q(s) = (q_x(s), q_y(s), q_z(s))^T \), of the length along the ray \( s \in \mathbb{R} \). (To save text space, we indicate column vectors as the transpose (T) of row vectors.)

**Geometric postulate.** The rays \( q(s) \) are continuous and piecewise differentiable.

That is, the line \( q(s) \) is connected and has a tangent vector \( p(s) = (p_x(s), p_y(s), p_z(s))^T \) for \( s \in \mathbb{R} \), except possibly at isolated points where \( p(s) \) may be discontinuous. We state this postulate in the incremental form shown in figure 1,

\[
\Delta q = \frac{p}{|p|} \Delta s. \tag{2.1a}
\]

We do not (yet) restrict the length of the tangent vector.

The **dynamic postulate** is a statement on how light rays bend in inhomogeneous media described by a function of space \( n(q) \). We consider \( n(q) \) to be region-wise smooth, with gradient vector (that is zero where \( n \) is constant), except possibly at surfaces (with well-defined normal vector) where \( n \) can have a finite discontinuity.

**Dynamic postulate.** The tangent vector \( p(s) \) responds to the space gradient of the function \( n(q) \).

![Figure 1. A light ray is a piecewise differential line \( q(s) \), \( s \in \mathbb{R} \), with tangent in the direction \( p(s) \).](image)

![Figure 2. The space gradient of the refractive index changes the ray tangent as it passes through an inhomogeneous region.](image)

We write this postulate in incremental form with the gradient vector \( \nabla n(q) \) as

\[
\Delta p = \nabla n(q) \Delta s \tag{2.1b}
\]

see figure 2. Addition of constants to \( n \) does not change the effect of the medium on the ray, but the relation above establishes a scale between the change of \( p \) and the change of \( n \) over the trajectory interval \( \Delta s \). This equation recalls a "force" acting along the gradient of a 'potential'; both words are in quotes because they are borrowed from mechanics. The intuition they convey is right, however, except that we have not been speaking of time but of length \( s \), and not of a mechanical potential but of the refractive index \( n(q) \) of an optical medium. In the next section we shall see that these postulates describe both optical and mechanical trajectories.

2.1. Hamilton equations

In regions where the refractive index \( n(q) \) is smooth, in the limit \( \Delta s \to 0 \) the two basic incremental equations (2.1) imply the vector equations

\[
\frac{dq}{ds} = \frac{p}{|p|}, \tag{2.2a}
\]

\[
\frac{dp}{ds} = \frac{\partial n(q)}{\partial q}. \tag{2.2b}
\]

These are the first and second Hamilton equations, respectively (in the parameter \( s \). Their structure determines that there will be an invariant \( \mathcal{H} \) (the Hamiltonian function [5]) such that \( \partial \mathcal{H}/\partial p \) equals (2.2a) and \( -\partial \mathcal{H}/\partial q \) equals (2.2b). Indeed, this is

\[
\mathcal{H} = |p| - n(q). \tag{2.3}
\]
It is constant along each ray because it is independent of $s$:

$$\frac{\delta \mathcal{H}}{\delta s} = \frac{\partial \mathcal{H}}{\partial \varphi} \cdot \frac{dp}{ds} + \frac{\partial \mathcal{H}}{\partial \psi} \cdot \frac{dq}{ds} = -\frac{\partial n}{\partial \varphi} \cdot \frac{dq}{ds} + \frac{p}{|p|} \cdot \frac{dp}{ds} = 0 \quad (2.4)$$

We choose henceforth the value $\mathcal{H} = 0$ because it leads to simplified formulae and a particularly transparent restriction for the norm of the tangent vector $p$, that is then constrained to lie on a sphere

$$|p| = n(q). \quad (2.5)$$

This has been called the Descartes sphere, by inspiration from the figures he drew in [6].

The case of free flight in a homogeneous medium, $n =$ constant, is found from (2.2b) and (2.2a) as the basic solution

$$q(s) = q(0) + sp/n \quad (n \text{ constant}) \quad (2.6)$$

$$p(s) = p(0).$$

When the medium is smoothly inhomogeneous, the Hamilton equations yield solutions of the general form and properties

$$q(s) = Q(q_0, p_0, s)$$

$$Q(q_0, p_0, 0) = q_0 = Q(q(s), p(s), -s),$$

$$p(s) = P(q_0, p_0)$$

$$P(q_0, p_0, 0) = p_0 = P(q(s), p(s), -s).$$

2.2. Finite refraction

When the refractive index $n(q)$ has a finite discontinuity at a surface $S(q) = 0$ between the ray points $q(s)$ and $q(s + \Delta s)$, where the tangents are $p(s)$ and $p(s + \Delta s)$, we may let $\Delta s \to 0$ approaching the point at $S$ from both sides. We indicate the variables at $s + \Delta s$ by primes, so in (3.4a), $\Delta q = q' - q \to 0$. Meanwhile, $\Delta p = p' - p$ remains finite in media $n = n(q)$ and $n' = n(q')$. The gradient of $n(q)$ defined through the incremental limit $\Delta q \to 0$ yields the vector normal to the surface $\Sigma = \nabla n(q)|_{S} \Delta s$, which remains finite as $\Delta s \to 0$; thus in (2.1b) we conclude that $\Delta p$ is parallel to $\Sigma$ at the point of incidence. Hence, in the limit, the two equations can be written as conservation laws for the ray at the surface $S$,

$$q|_{S} = q'|_{S} \quad (2.8a)$$

$$\Sigma \times p|_{S} = \Sigma \times p'|_{S}. \quad (2.8b)$$

The first equation is again a statement of continuity for the line at the refracting surface. The

**Figure 3.** Ibn Sahl's diagram for refraction. A ray incident on a surface with normal $\Sigma$ has direction $p$; the increment $\Delta p$ is parallel to $\Sigma$. The refracted ray has direction $p'$ such that the ratio of vector lengths $|p|$ and $|p'|$ is the ratio of the refractive indices, $n/n'$. The common right side of the two triangles is the conserved quantity.

The second equation is dynamic: if $\theta$ and $\theta'$ are the angles between $\Sigma$ and the vectors $p$ and $p'$, recalling from (2.5) that $|p| = n$ and $|p'| = n'$, the norm of (2.8b) is the well known sine law of Ibn Sahl, Snell and Descartes

$$n \sin \theta = n' \sin \theta'. \quad (2.9)$$

The three vectors are coplanar. Note that, had we chosen the Hamiltonian (2.3) to be $\mathcal{H} = E$, a non-zero constant, the coefficients in front of the two sides of (2.9) would be $|p| = n + E$ and $|p'| = n' + E$, and this would redefine the physical refractive index by that additive constant.

Figure 3 shows the beautiful and simple diagram used by Ibn Sahl in his treatise On Burning Instruments to find the optical properties of plano-convex or biconvex hyperboloidal lenses that can concentrate the light of the Sun or of near light sources, respectively. The manuscript book of Ibn Sahl was known to his disciples and followers [2], but its pages were eventually disassembled, part of them ending in Damascus, part in Tehran, and part were lost. Their reconstruction earned Professor Roshdi Rashed the 1991 Third World Academy of Sciences prize. We feel obliged to put a new name to the sine law of refraction that was written some 650 years before Wildebrod Snell found it experimentally [1, 7], and René Descartes by contrived (and incorrect) theoretical reasoning. Ibn Sahl’s diagram is drawn with the surface normal $\Sigma$ and the incident ray tangent vector $p$ at an angle $\theta$; this defines a right triangle with a hypotenuse whose length one sets to $n$. Then one builds a second right triangle with the side other than $\Sigma$ in common with the first triangle, and whose hypotenuse is of length $n'$. The law of sines (2.9) writes the length of the common side in two ways. We should underline that Ibn Sahl’s diagram does not draw the rays themselves, but their directions at the point of incidence.
3. Time and mechanics

The incremental forms of the geometric and dynamic postulates were presented in general terms, drawing on the intuitive meaning they have in classical optics; in fact, they apply equally well to mechanics as they do to optics. The distinction between the two physical systems stems from a third, physical postulate that states the relation between length intervals $\Delta s$ and time intervals $\Delta t$.

The Newtonian velocity of a light point in a medium follows the optical postulate:

$$\Delta s = \frac{c}{n(q)} \Delta t.$$  \hfill (3.1a)

where $c$ is the reference velocity of light in vacuum, and the refractive index $n$ ranges from 1 (for vacuum) up to 2.4 (for diamond). The velocity has thus finite upper and lower bounds.

On the other hand, a mechanical point of mass $m$ obeys the mechanical postulate:

$$\Delta s = \frac{|p|}{m} \Delta t.$$  \hfill (3.1b)

In classical mechanics the velocity of the particle, $dq/dr$, keeps the fixed proportion $m$ with $|p|$; it may be zero and is unbounded. The vector $p = m \times q$ is called the momentum of the object. We can use the same name for it also in optics.

For mechanical systems, the incremental equations (2.1), under the substitution (3.1b), become

$$\Delta q = \frac{p}{m} \Delta t,$$  \hfill (3.2a)

$$\Delta p = \frac{1}{2m} \frac{\partial^2 V(q)}{\partial q^2} \Delta t.$$  \hfill (3.2b)

The Hamilton equations for time evolution are obtained upon letting $\Delta t \to 0$. They have the familiar form

$$\frac{dq}{dt} = \frac{p}{m} = \frac{\partial H}{\partial p},$$  \hfill (3.3a)

$$\frac{dp}{dt} = \frac{1}{2m} \frac{\partial^2 V(q)}{\partial q^2} = -\frac{\partial H}{\partial q}$$  \hfill (3.3b)

with the well known mechanical Hamiltonian invariant

$$H = \frac{1}{2m} |p|^2 + V(q).$$  \hfill (3.4)

obtained from integration of the last equality in (3.3b) with respect to $q$, recalling (2.5), and provided

$$n^2(q) = 2m(E - V(q)).$$  \hfill (3.5)

The number $E$ is constant in time and known as the energy of the trajectory $q(t)$. The function $V(q)$ is the mechanical potential function. For fixed $E$, the mechanical momentum $p$ is constrained to the sphere

$$|p| = \sqrt{2m(E - V(q))}.$$  \hfill (3.6)

This we may call the Newton sphere, vis-à-vis the optical Descartes sphere.

In this way, the refractive index of an optical system is related to the potential of a corresponding mechanical system at a fixed energy, and vice versa (provided that $1 \leq n \leq 2.4$). A dense optical region (for example, a ‘Gaussian’ refractive index $n = 1 + \exp(-|q|^2/2\omega)$) will bend light rays in the same way as free particle trajectories of a given energy are bent by a smooth potential trough $V = E - n^2/2m$.

Common optical media such as lens systems have surfaces of discontinuity in the refractive index. The related mechanical systems are square-well potential walls along the same surfaces. Finite mechanical refraction then occurs with the conservation of two quantities: of position $q$ in (2.8a), stating that the particle trajectory is continuous in time, and of momentum $p$ in (2.8b), in obedience to the Ibn Sahl law. Time is clearly the appropriate independent (evolution) parameter for systems that follow the mechanical postulate (3.1b); if the particle stops somewhere along its trajectory, the Hamiltonian equations (3.3)-(3.4) still hold, while (2.2) would have a discontinuity.

4. Factorization of refraction

The presence of a surface of discontinuity $S(q) = 0$ between two media of refractive indices $n(q)$ and $n'(q) = n(q')$ produces a finite transformation $G_{n,n';S}$ of the ray trajectories (2.7), that we write as

$$G_{n,n';S} = \begin{cases} q_0 & q_0 \\
0 & p_0 \end{cases} \begin{cases} q_0 \\
0 \end{cases}.$$  \hfill (4.1a)

for ‘initial points’ $s = 0$ of both rays. See figure 4.

We shall now show that this transformation factorizes into two root factors $R$ in the following way:

$$G_{n,n';S} = R_{n,S}(R_{n';S})^{-1}.$$  \hfill (4.1b)

Each root factor depends only on one of the media and the shape of the surface $S$. This will be done first in full generality and then particularized for two homogeneous media, simplifying the results reported in [3], [4] and [8].

Let the point of impact of a ray $q(s)$ on the surface $S(q) = 0$ be at the value $s$ of the ray length parameter
of the existence of the inverse of a root transformation because it is canonical \([3, 9]\), the
general factorization result (4.1b) follows. This result is general; it holds for optics and for
mechanics.

To show the practical advantage of the root transformation as an intermediate step in computing
the effect of a refracting surface, we consider the case of a surface separating two homogeneous
media \(n\) and \(n'\). We introduce Cartesian coordinates in space. We

**Figure 4.** Refraction is a factorizable transformation between the initial points of the incident \((q, p)\) and
refracted rays \((q', p')\) lying on a reference plane.

\(s\). Then, using the notation of equations (2.7) in the
medium \(n(q)\), assume that

\[
S(Q(q_0, p_0, z)) = 0 \quad \text{solves for} \quad z = \tilde{z}(q_0, p_0; S).
\]

(4.2a)

There may be rays whose initial conditions \(q_0, p_0\) are such that they miss the surface, or cross it at
more than one point; we disregard these cases assuming the solution is unique. Similarly, in the
second medium \(n'(q)\) we indicate functions and quantities by primes, so that

\[
S(Q'(q_0', p_0', z)) = 0 \quad \text{solves for} \quad z = \tilde{z}(q_0', p_0'; S)
\]

(4.2b)

with the same surface \(S\) and value of \(\tilde{z}\). Then, we write the conservation laws (2.8) in the form

\[
Q(q_0, p_0, z) = Q'(q_0', p_0', \tilde{z})
\]

\[
\Sigma l_0 \times P(q_0, p_0, z) = P' = \Sigma l_0' \times P'(q_0', p_0', \tilde{z})
\]

(4.3a) (4.3b)

where \(\tilde{Q}\) is the point of impact, \(\Sigma l_0\) is the vector normal to the surface \(S\) at that point, and \(P\) is a
fixed vector.

The structure of these two conservation laws interprets the barred point-of-impact quantities in
the middle terms as the result of transformation of the initial values \(q_0, p_0\) (by propagation through \(\tilde{z}\)) in the left-hand side, and sets them equal to a similar transformation (propagation by the same \(\tilde{z}\)
in the second medium) of the values \(q_0', p_0'\).

That is,

\[
\mathcal{R}_{n, s} : \begin{cases}
q_0 = Q_0 \\
\Sigma l_0 \times p_0 = P = \mathcal{R}_{n, s} : \begin{cases}
q_0 \Sigma l_0 \times p_0
\end{cases}
\end{cases}
\]

(4.4a)

Comparing this with (4.1a) (and being assured

\[
q + \tilde{q} \bar{p}/n = \tilde{q}, \quad q_z + \tilde{q}_z p_z/n = \tilde{q}_z.
\]

(4.7a) (4.7b)

From the last equation,

\[
\tilde{z} = (\bar{q}_z - q_z) n/p_z = \zeta(q)n/p_z
\]

(4.8)

is the \(s\)-coordinate of the point of impact (i.e., the
solution of the general equations (4.2)). This is now
replaced into the first two component equations (cf (4.3a)), to yield

\[
\mathcal{R}_{n, s} : q = \bar{q} = q + \zeta(q) \frac{\bar{q}}{p_z}.
\]

(4.9a)

This is the first root equation to solve for \(\bar{q}(q, p)\).

The second root equation to be solved second stems
from (4.3b), and will be shown to be

\[
\mathcal{R}_{n, s} : p = \bar{p} = p + \Sigma(q)p_z.
\]

(4.9b)
Indeed, this is an explicit equation that depends on the result \( \bar{q}(q; \zeta(q)) \) of the first. To prove (4.9b) for homogeneous media, the quantity indicated \( \bar{P} \) for the general case (4.3b) is put in the form \( \bar{P} = \Sigma \times \bar{p} + \delta \Sigma \), for some fixed vector \( \Sigma \) and constant \( \delta \). For simplicity we choose \( \Sigma \) to be the unit normal to a \( z = \) constant plane. Then, equation (4.3b) becomes, in components,

\[
\begin{pmatrix}
\Sigma \\
-1
\end{pmatrix} \times \begin{pmatrix}
p \\
p_z
\end{pmatrix} = \begin{pmatrix}
0 & 1 \\
-1 & 0
\end{pmatrix} \begin{pmatrix}
p + p_z \Sigma \\
\Sigma \times p
\end{pmatrix}
\]

\[
= \begin{pmatrix}
0 \\
-1
\end{pmatrix} \times \begin{pmatrix}
p \\
p_z
\end{pmatrix} + \begin{pmatrix}
0 \\
\delta
\end{pmatrix}
\]

\[
= \begin{pmatrix}
0 & 1 \\
-1 & 0
\end{pmatrix} \begin{pmatrix}
p \\
p_z
\end{pmatrix} + \begin{pmatrix}
\delta
\end{pmatrix} \tag{4.10}
\]

From the first two components follows (4.9b). The third component is not independent (cf (4.6b)).

The first root equation (4.9a) may be the most difficult part of the task of finding the effect of refraction: only for planes \( S(q) \sim k \cdot q \), paraboloids \( \sim |q|^2 \), cubic and quartic surfaces does this equation solve in closed form; in general it is an implicit equation that depends on the shape of the surface \( S \) through \( \zeta(q) \). Yet it is a simple kind of implicit equation, of the form \( \bar{q} = q + f_\zeta(q,p) \), rather than the more difficult coupled pair of equations of the form \( g_e(q,p) = g_e(q',p') \), that we would encounter if we attempted to solve refraction without factorization. Indeed, the simpler implicit form of the root equations is well suited for aberration expansions, i.e., Taylor expansions of \( q'(q,p;\zeta) \), \( p'(q,p;\zeta) \), and \( \zeta(q) \) in powers of the components of \( q \) and \( p \) [3, 9]. Note that when the surface \( S \) is a \( z = \) constant plane, the root transformation is nothing but free flight (2.6). The second root equation (4.9b) is also well suited for aberration expansions of \( \bar{p}(q,p;\zeta) \). It can be handled by relatively straightforward symbolic computation algorithms [8, 9].

The next step is to find the inverse root transformation. This is

\[
\mathcal{R}^{-1}_{1;S} : q' = \bar{q} = \zeta(q)p'/p; \\
\mathcal{R}^{-1}_{2;S} : p' = \bar{p} = \Sigma(q)p'. \tag{4.11a, b}
\]

Here, the second equation must be solved first; it is an implicit equation whose solution yields \( p'(q,\bar{p};\zeta) \). The result is then replaced in (4.11a), that is explicit, and gives \( q'(q,\bar{p};\zeta) \). Finally, we must compose the two transformations replacing the arguments of (4.11) by those of (4.9). All calculations are well served by aberration expansions handled through symbolic computer algorithms [8, 9].

When one or both of the two media \( n \) and \( n' \) is not homogeneous (but still differentiable), then the main problem becomes the analytic or series-expanded computation of the path in those media. This is possible, for example, when the medium is an optical fibre whose refractive index is independent of the \( z \)-coordinate and of the form \( n(q) = \sqrt{\mu - q^2 - q^2} \) [11]. Such a medium yields to easy solution because it is vis-à-vis the mechanical harmonic oscillator.

5. Concluding remarks

Geometric postulates are used often in quantum mechanics with the designation of *symmetries*. The dynamic properties of quantum systems are given by the *potential* in their Schrödinger equation. For some special systems, such as the hydrogen atom or the harmonic oscillator, the dynamics can be incorporated into a statement of *dynamic symmetry*. In other realms, such as elementary particle models, the potentials are not really known or do not exist, so postulating higher symmetries is common currency. Classical optics has lagged behind in this respect, in part because its dynamic law is known precisely— Ibn Sahl's law—and in part because it is perceived as a different discipline, quite unrelated to mechanics. Where cross-breeding has occurred, such as in coherent and squeezed states of light, optical waveguides, paraxial wave optic systems, and the Maxwell fish-eye (the hydrogen atom of optics), the breed has been interesting and useful.

The most distinctive feature of classical optical systems is the presence of sharp refracting interfaces between media; these seldom appear in mechanics. Deriving both *Hamilton* evolution equations and finite refraction conservation laws from common incremental equations brings optics and mechanics conceptually closer. Describing mechanics with the same language as optics is an invitation to further apply the results of one on the other.

References