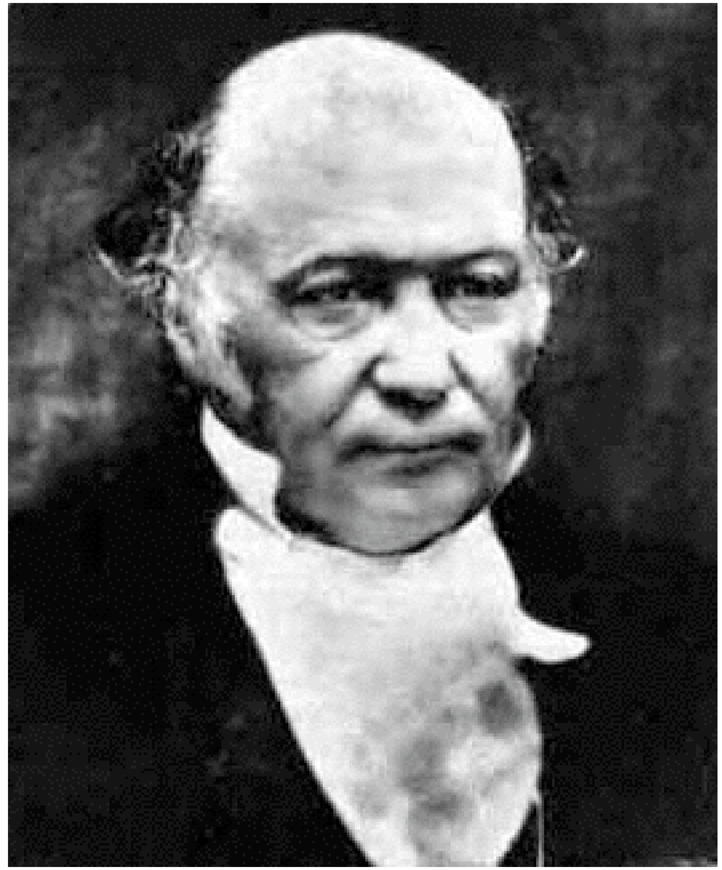
Symplectic Geometry
Lecture 1

Hamilton

Many of the key ideas of symplectic geometry were developed by Hamilton (1805-1865) - first in the context of geometrical optics and then in the context of classical mechanics. His discussion of optics first occurred in in a paper he communicated to Dr. Brinkley in 1823, by whom, under the title "Caustics" it was presented in 1824 to the Royal Irish Academy. It was referred as usual to a "committee" and was finally published in a sequence of papers in the *Transactions of the Royal Irish Academy* starting in 1828 under the title "Essay on the theory of systems of rays, with three supplements": **15** (1828) 69-174, **16** (1830 and 1831) 4-62 and 85-92, and **17** (1837) 1-144. Later he realized that the method he employed for optics could also be used in mechanics. His two great papers this field appeared in the *Philosophical Transactions of the Royal Society of London* under the title "On a general method in dynamics" (1834) 247-308 and (1835) 95-104.

Dr. Brinkley (1763-1835), later the archbishop of Clyne was then the first royal astronomer for Ireland and an accomplished mathematician who perceived the vast talents of young Hamilton. He is said to have remarked in 1823 of this lad of eighteen "This young man, I do not say *will be* but *is*, the first mathematician of the age".

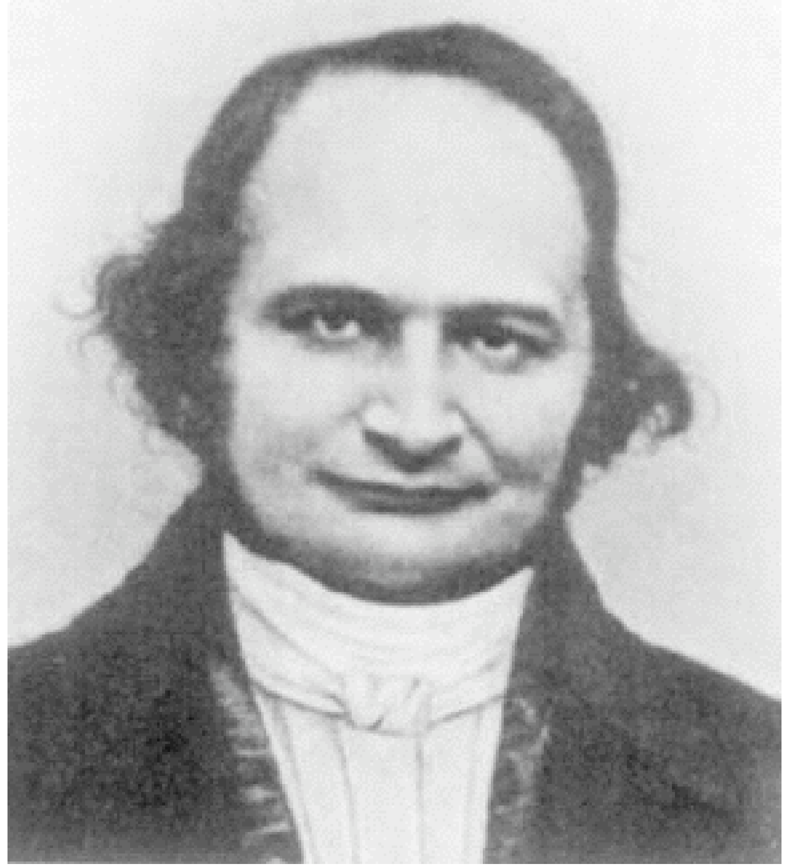
Sir William Rowan Hamilton



Born: 4 Aug 1805 in Dublin, Ireland
Died: 2 Sept 1865 in Dublin, Ireland

In the 1920's the work of Hamilton became one of the inspirations for the development of quantum mechanics. The idea was that just as geometrical optics was a (short wave length) approximation to wave optics (and hence to electrodynamics) classical mechanics as formulated by Hamilton should be an approximation to a "wave mechanics".

Carl Gustav Jacob Jacobi



Born: 10 Dec 1804 in Potsdam, Prussia (now Germany)
Died: 18 Feb 1851 in Berlin, Germany

One of the ingredients in Hamilton's theory was a certain first order non-linear partial differential equation satisfied by the so called "characteristic functions", see later on for a more detailed description of this concept. Hamilton did not ask how to solve the most general non-linear first order partial differential equation of the type he introduced. This was done by Jacobi (1804-1851) who observed that the solution of such an equation could be reduced to a system of certain ordinary differential equations. Since then the theory is known as "Hamilton-Jacobi theory".

Jacobi's results were published in:

"Über die Reduction der Integration der partiellen Differentialgleichungen erster ordnung zwischen irgend einer Zahl Variabeln auf die Integration eines einzigen Systemes gewöhnlicher Differentialgelichungen". *Crelle's Journal für die reine und angewandte Mathematik* **1** (1837) 97-162,

"Nova methodus, equationes differentiales partiales primi orins inter numerum variabilium quemcuque propositas integrandi", *Crelle journal für die reine und angewandte Mathematik* **60** (1862) 1-181, and

Vorlesungen über Dynamik. Gehalten an der Universität Königsberg im Wintersemester 1842-43 und nach einem von C.W. Borchardt ausgearbeiteten Hefte. Verlag G. Reimer, Berlin, 1881.

Marius Sophus Lie

Born: 17 Dec 1842 in Nordfjordeide, Norway
Died: 18 Feb 1899 in Kristiania (now Oslo), Norway

Sophus Lie (1842-1899) developed the idea of viewing the solution of the Hamilton-Jacobi equation as a "Lagrange submanifold" of the "cotangent bundle" in a series of articles in 1872-78, and stressed the point that it is based on the fact that the solutions of the Hamiltonian system of differential equations leaves the "canonical two-form" of the "cotangent bundle" invariant. I will explain the terms in quotation marks later on. For Lie, the use of the group of transformations which leave the Hamiltonian and the two form invariant was analogous to the use of the Galois group in the solution of polynomial equations.

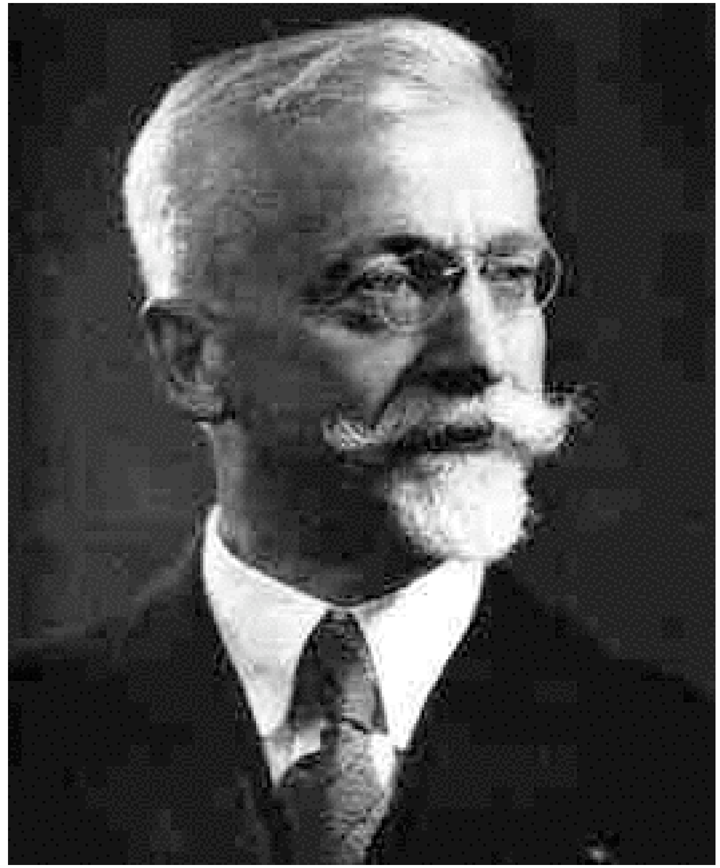
Lie's work is summarized in his book:

Theorie der Transformationsgruppen I-III. Unter Mitwirkung von Prof.dr. F. Engel. Teubner, Leipzig, 1888, 1890, 1893.

An account of his work on first order partial differential equations can be found in:

F. Engel und K. Faber: *Die Liesche Theorie der partiellen Differentialgleichungen Erster Ordnung*. Teubner, Berlin, 1935.

Elie Joseph Cartan



Born: 9 April 1869 in Dolomieu (near Chambéry), Savoie, Rhône-Alpes, France
Died: 6 May 1951 in Paris, France

The subject was completely transformed (as was all of differential geometry) by the work of Élie Cartan (1869-1951) who developed (in 1903) the "exterior differential calculus" from earlier work of Grassmann. Cartan's approach and his calculus were totally innovative. He formulated problems so that they were invariant and did not depend on the particular variables or coordinates. We will briefly review this calculus and use it to give the modern formulation of symplectic geometry.

Symplectic vector spaces.

All of the 19th century work was done before the concept of an abstract vector space became a part of standard mathematics. We will begin with a study of "linear symplectic geometry" which can be regarded as a "linear approximation" to the subject we really want to study.

Let V be a (usually finite dimensional) vector space over the real numbers. A symplectic structure on V consists of an anti-symmetric bilinear form

$$\omega : V \times V \to \mathbf{R}$$

which is non-degenerate.

A vector space equipped with a symplectic structure is called a symplectic vector space.

To say that ω is a bilinear form means that ω assigns a number $\omega(u,v)$ to every pair $u, v \in V$ in such a way that

$$\omega(au_1 + bu_2, v) = a\omega(u_1, v) + b\omega(u_2, v)$$

for all real numbers a and b and $u_1, u_2, v \in V$ and

$$\omega(u, av_1 + bv_2) = a\omega(u, v_1) + b\omega(u, v_2)$$

for all real numbers a and b and $u, v_1, v_2, \in V$.

To say that ω is anti-symmetric means that

$$\omega(v,u) = -\omega(u,v)$$

for all $u, v \in V$.

To say that ω is non-degenerate means that

$$\omega(u,v) = 0 \;\forall v \quad \Rightarrow \quad u = 0.$$

Special kinds of subspaces.

If W is a subspace of symplectic vector space V then $W^\perp$ denotes the symplectic orthocomplement of W:

$$W^\perp := \{v \in V |\ \omega(v, w) = 0,\ \forall w \in W\}.$$

A subspace is called

1. **symplectic** if $W \cap W^\perp = \{0\}$,
2. **isotropic** if $W \subset W^\perp$,
3. **coisotropic** if $W^\perp \subset W$, and
4. **Lagrangian** if $W = W^\perp$.

Symplectic subspaces.

A subspace is called

1. **symplectic** if $W \cap W^{\perp} = \{0\}$,

Since $(W^{\perp})^{\perp} = W$ by the non-degeneracy of ω, it follows that W is symplectic if and only if $W^{\perp}$ is. Also, the restriction of ω to any symplectic subspace W is non-degenerate, making W into a symplectic vector space. Conversely, to say that the restriction of ω to W is non-degenerate means precisely that $W \cap W^{\perp} = \{0\}$.

Isotropic subspaces.

A subspace is called

2. **isotropic** if $W \subset W^{\perp}$

It follows from the anti-symmetry of ω that

$$\omega(u, u) = 0$$

for any $u \in V$. This implies that *any one dimensional subspace of V is isotropic.*

Suppose that $W \subset W^{\perp}$ but $W \neq W^{\perp}$. Let $v \in W^{\perp}$ but $v \notin W$. Consider the subspace $W + \mathbb{R} \cdot v$ spanned by W and v. Then

$$(W + \mathbb{R} \cdot v)^{\perp} = W^{\perp} \cap (\mathbb{R} \cdot v)^{\perp}$$

and we know that $v \in (\mathbb{R} \cdot v)^{\perp}$ and $v \in W^{\perp}$. This shows that

$$W + \mathbb{R} \cdot v \subset (W + \mathbb{R} \cdot v)^{\perp}.$$

Suppose that $W \subset W^{\perp}$ but $W \neq W^{\perp}$. Let $v \in W^{\perp}$ but $v \notin W$. Consider the subspace $W + \mathbb{R} \cdot v$ spanned by W and v. Then

$$(W + \mathbb{R} \cdot v)^{\perp} = W^{\perp} \cap (\mathbb{R} \cdot v)^{\perp}$$

and we know that $v \in (\mathbb{R} \cdot v)^{\perp}$ and $v \in W^{\perp}$. This shows that

$$W + \mathbb{R} \cdot v \subset (W + \mathbb{R} \cdot v)^{\perp}.$$

So $W + \mathbb{R} \cdot v$ is an isotropic subspace of one greater dimension than W. Proceeding inductively, this must come to an end if the dimension of V is finite as we shall now assume. We have proved that every isotropic subspace is contained in a maximal isotropic subspace and that L is a maximal isotropic subspace if and only if $L = L^{\perp}$, which says that L is Lagrangian.

Lagrangian subspaces.

A subspace is called

4. **Lagrangian** if $W = W^{\perp}$.

If A is any subspace of V, the non-degeneracy of ω implies that $\dim A^{\perp} = \dim V - \dim A$. So if $L = L^{\perp}$ then $\dim V = 2\dim L$. We see that every (finite dimensional) symplectic vector space is even dimensional and that an isotropic subspace of V is Lagrangian if and only if its dimension is $\frac{1}{2}\dim V$ if and only if it is maximal isotropic.

Normal form for a symplectic vector space.

For any non-zero $e \in V$ we can find an $f \in V$ such that $\omega(e, f) = 1$ and so the subspace W spanned by e and f is a two dimensional symplectic subspace. Furthermore the map

$$e \mapsto \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad f \mapsto \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

gives a symplectic isomorphism of W with $\mathbf{R}^2$ with its standard symplectic structure. We can apply this same construction to $W^\perp$ if $W^\perp \neq 0$. Hence by induction, we can decompose any symplectic vector space into a direct sum of two dimensional symplectic subspaces:

$$V = W_1 \oplus \cdots W_d$$

$$V = W_1 \oplus \cdots W_d$$

where $\dim V = 2d$ (proving that every symplectic vector space is even dimensional) and where the W_i are pairwise (symplectically) orthogonal and where each W_i is spanned by e_i, f_i with $\omega(e_i, f_i) = 1$. In particular this shows that all $2d$ dimensional symplectic vector spaces are isomorphic, and isomorphic to a direct sum of d copies of $\mathbf{R}^2$ with its standard symplectic structure.

This gives another proof of the fact that every symplectic vector space is even dimensional.

The symplectic group.

A linear transformation $T : V \to V$ of a symplectic vector space is called **symplectic** if

$$\omega(Tu, Tv) = \omega(u, v) \quad \forall\, u, v \in V.$$

Clearly the product of two symplectic linear transformations is symplectic and the identity is a symplectic linear transformation. Also, if T is a symplectic linear transformation and $Tu = 0$ then $\omega(u, v) = 0$ for all $v \in V$ and so $u = 0$. So every symplectic linear trnasormation T is inveritble, and T^{-1} is clearly symplectic. So the set of symplectic linear transformations of V form a group known as the **symplectic group** (of V).

The symplectic group in two dimensions.

For example, if $V = \mathbb{R}^2$ then T is symplectic if and only if it preserves area and orientation which is the same as saying that $\det T = 1$. So the symplectic group of $\mathbb{R}^2$ is the same as $SL(2, \mathbb{R})$, the group of all 2×2 matrices of determinant one. In higher dimensions, the description of the symplectic group is a bit more complicated. We shall discuss this later.

For today's lecture I will need a theorem of Gauss about $SL(2, \mathbb{R})$ which says that every element of $SL(2, \mathbb{R})$ can be written as a product of matrices of the form

$$\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & 0 \\ y & 1 \end{pmatrix}.$$

Proof. First suppose that our matrix $T \in SL(2, \mathbb{R})$ is of the form

$$T = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \text{with} \quad c \neq 0.$$

For real numbers s and t consider the product

$$\begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a + sc & t(a+sc) + b + sd \\ c & ct + d \end{pmatrix}.$$

Since $c \neq 0$ we can choose s so that $a + sc = 1$ and then choose $t = -(b + sd)$. With this choice the matrix on the right has a 1 in the upper left hand corner and a 0 in the upper right hand corner. Since its determinant is 1, the lower right hand corner must also be 1. So for these choices of s and t we have

$$\begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix}$$

or

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & -s \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix} \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}.$$

Now consider a matrix of determinant one of the form

$$\begin{pmatrix} a & b \\ 0 & d \end{pmatrix}.$$

Since the determinant is one, we know that $a \neq 0$. Multiply this matrix by

$$\begin{pmatrix} 1 & 0 \\ -p & 1 \end{pmatrix}, \quad p \neq 0$$

to obtain

$$\begin{pmatrix} 1 & 0 \\ -p & 1 \end{pmatrix} \cdot \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} = \begin{pmatrix} a & b \\ -pa & d \end{pmatrix}$$

or

$$\begin{pmatrix} a & b \\ 0 & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ p & 1 \end{pmatrix} \cdot \begin{pmatrix} a & b \\ -pa & d \end{pmatrix}$$

Completion of the proof of Gauss's theorem.

$$\begin{pmatrix} a & b \\ 0 & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ p & 1 \end{pmatrix} \cdot \begin{pmatrix} a & b \\ -pa & d \end{pmatrix}$$

and we know that the matrix $\begin{pmatrix} a & b \\ -pa & d \end{pmatrix}$ can be written as product of three matrices of our desired type. This proves Gauss' theorem. We shall use this theorem to show that there is a dictionary between "Gaussian optics" (to be described below) and the study of $Sl(2)$.

In the history of physics it is often the case that when an older theory is superseded by a newer one, the older theory retains its validity - either as an approximation to the newer one, an approximation which is valid for an interesting range of circumstances or as a special case of the newer theory.

The currently held theory of light is known as quantum electrodynamics. It describes very successfully and very accurately the interaction of light with charged particle, explaining both the discrete character of light as evinced by the photo-electric effect and the wave-like character of electromagnetic radiation. The triumph of 19th 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 sub-atomic level.

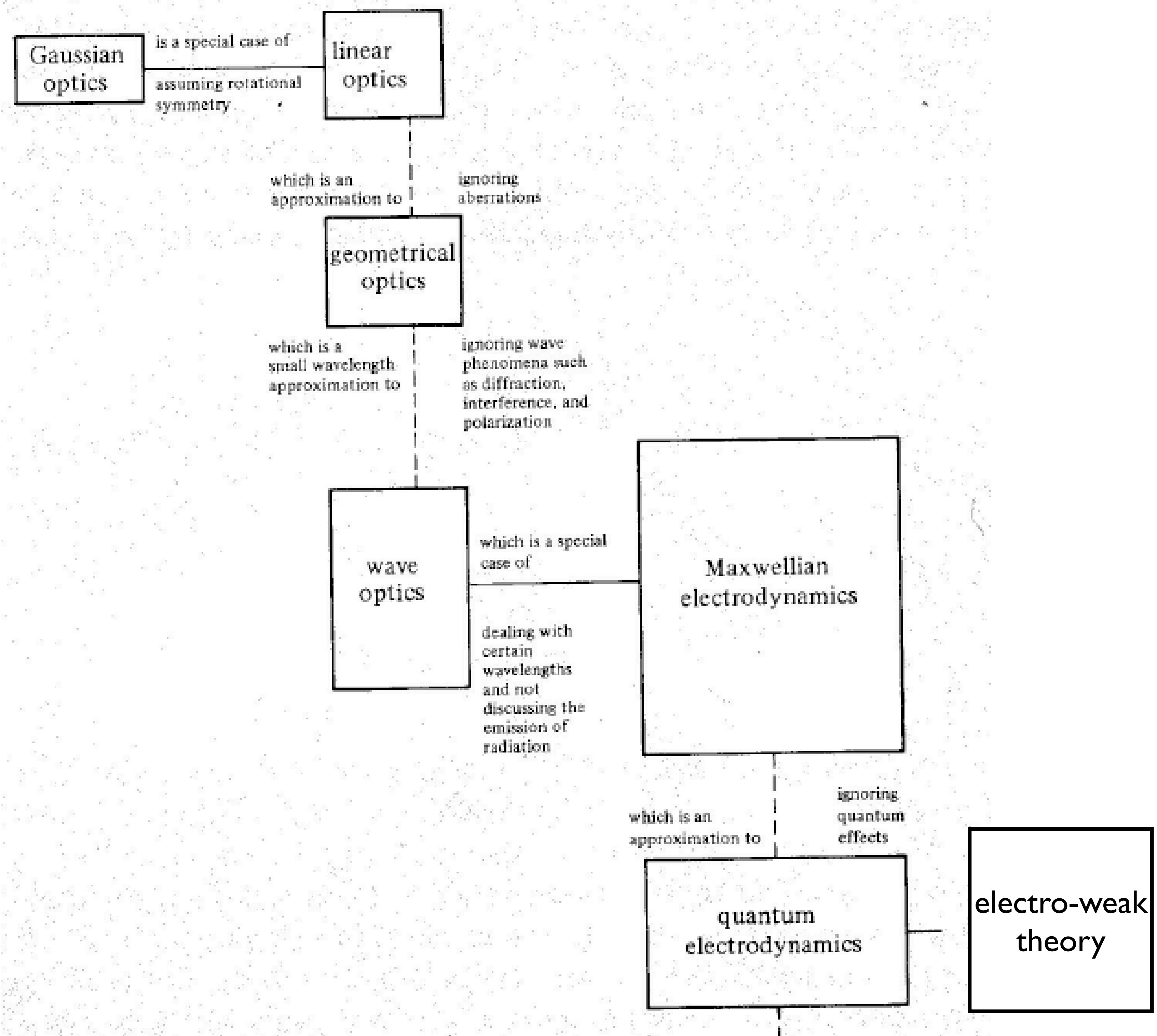
One of Maxwell's remarkable discoveries was that visible light is a form of electromagnetic radiation as is radiant heat. In fact, since Maxwell, optics is a special chapter in the theory of electricity and magnetism. Electromagnetism treats electromagnetic vibrations of all wave lengths. In the flood of invisible electromagnetic radiation that is accessible to the mental eye of the physicist, the physiological eye is almost blind, as the range of wave-lenghts that if converts into sensations is very small.

Maxwell's theory dealt with the source of electromagnetic radiation as well as with its propagation. Before Maxwell, there was a fairly well developed theory of light due mainly to Fresnel, which dealt rather successfully with the propagation of light, but had nothing to say about its production. Fresnel's theory did account for three physical effects which could not be accounted for by earlier theories - diffraction, interference, and polarization.

Geometrical optics is the approximation to wave optics in which the wave character of light is ignored. It is valid when the sizes of the various apertures are large compared to the wave length of the light, and when we do not examine too closely what happens near shadows or foci. It does not account for diffraction, interference, or polarization. We will find that geometrical optics, as formulated by Hamilton is part of symplectic geometry.

Linear optics is an approximation to geometrical optics which is valid when the various angles which 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. So in linear optics all terms which are quadratic or higher in the angles are ignored. For example, in geometrical optics, *Snell's law* says that if light passed from a region where the index of refraction is n to a region where the index of refraction is n', then $n \sin i = n' \sin i'$ where i and i' are the angles that the ray makes with the normal to the surface separating the region. In linear optics we replace Snell's 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 between linear optics and geometrical optics are known as (geometrical) aberrations.

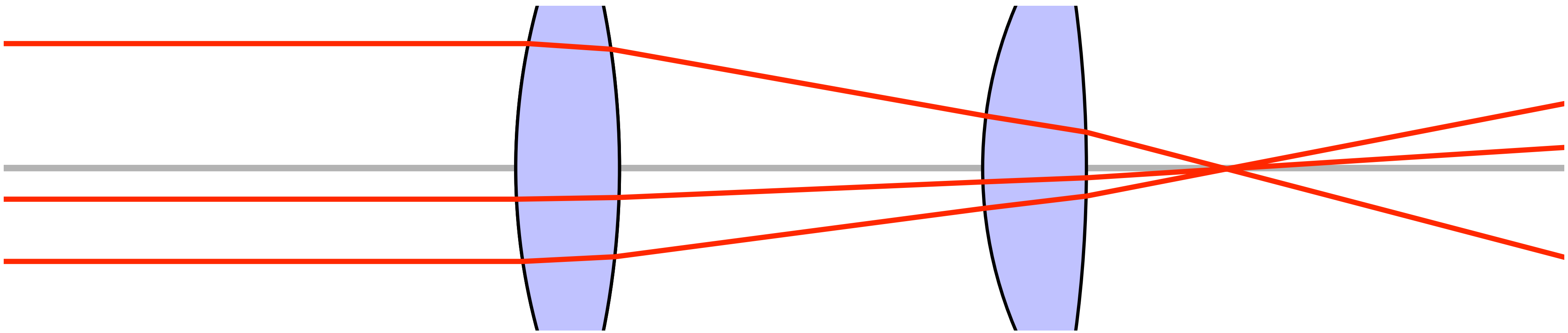
Gaussian optics is a special case of linear optics in which it is assumed that all the surfaces that enter are rotationally symmetric about a fixed axis. This is an important special case since most ground lenses and mirrors have this property.


Gaussian optics
is a special case of
assuming rotational symmetry
linear optics
which is an approximation to
ignoring aberrations
geometrical optics
which is a small wavelength approximation to
ignoring wave phenomena such as diffraction, interference, and polarization
wave optics
which is a special case of
dealing with certain wavelengths and not discussing the emission of radiation
Maxwellian electrodynamics
which is an approximation to
ignoring quantum effects
quantum electrodynamics
electro-weak theory

Gaussian optics

In Gaussian optics we are interested in tracing the trajectory of a light ray as it passes through various refracting surfaces of the optical system or is reflected by reflecting surfaces. We introduce a coordinate system in which the horizontal axis (with coordinate labeled z) coincides with optical axis, that is - with the axis of symmetry of the system. We shall restrict attention to coaxial rays - those which lie in a plane with the optical axis. When we pass from Gaussian optics to linear optics, we will see that the restriction to coaxial rays is harmless, in that the study of the most general ray can be reduced to coaxial rays by orthogonal projection. By rotational symmetry, it is then enough to study what happens for rays lying in one fixed plane containing the optical axis.

“Ray tracing” in Gaussian optics.



Any refracting lens system can be considered as the composition of several systems of two basic types:

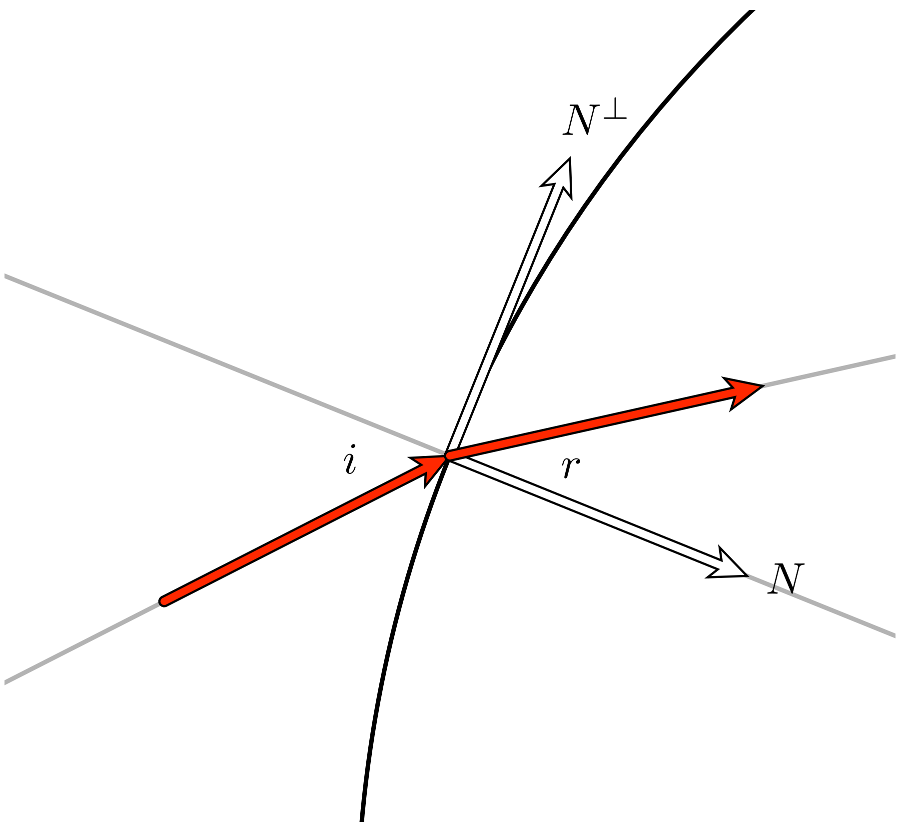
- A translation, in which the ray continues to move (in the same direction) in a straight line in a medium of constant refractive index. Here the angle that the ray makes with the optical axis does not change, but the height of the ray above the axis (denoted by q) does change.

- Refraction at a surface between regions of differing refractive index.

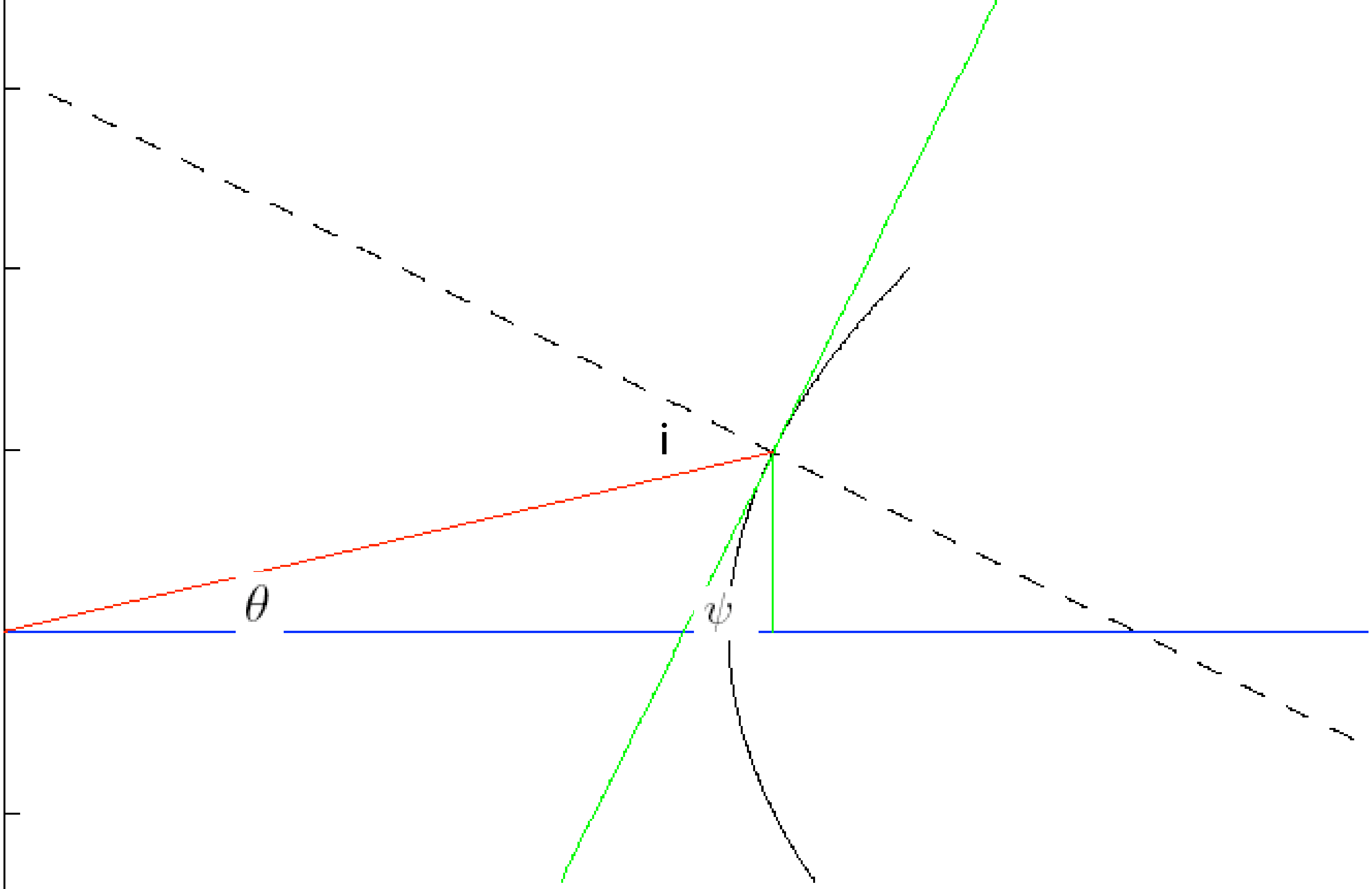
We will see that for the appropriate choice of coordinates, and in the Gaussian approximation, the translations correspond to multiplication by an upper triangular matrix with ones on the diagonal and the refractions correspond to multiplication by a lower triangular matrix with ones on the diagonal. By Gauss's theorem this implies that every optical system can be expressed as a element of $SL(2,\mathbb{R})$ and that every element of $SL(2,\mathbb{R})$ can be realized as an optical system.

Snell's law.

$$n_r \sin r = n_i \sin i$$



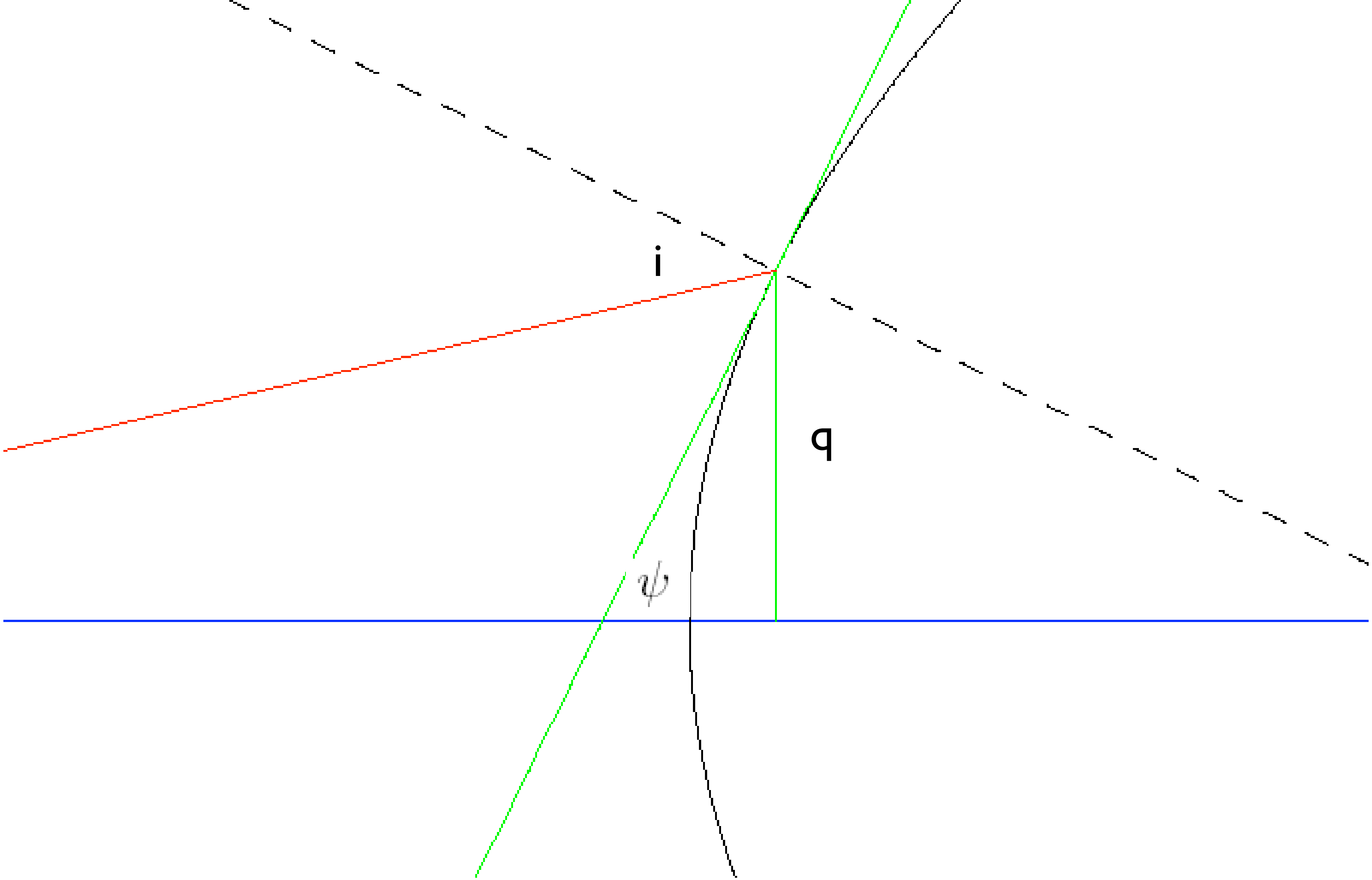
At a surface of refraction the q value does not change.



$$\pi - \psi + (\pi/2 - i) + \theta = \pi$$

or

$$i = \theta + (\pi/2 - \psi).$$



$$i = \theta + (\pi/2 - \psi).$$

We may assume that the curve giving the intersection of our surface of revolution with the plane is a parabola $z = \frac{1}{2}kq^2$ so $z'(q) = kq = \tan(\pi/2 - \psi) \doteq (\pi/2 - \psi)$.

So $i = \theta + kq$.

The matrix of a refraction.

We apply the linearized version of Snell's law which says $n_1 i_1 = n_2 i_2$. This gives

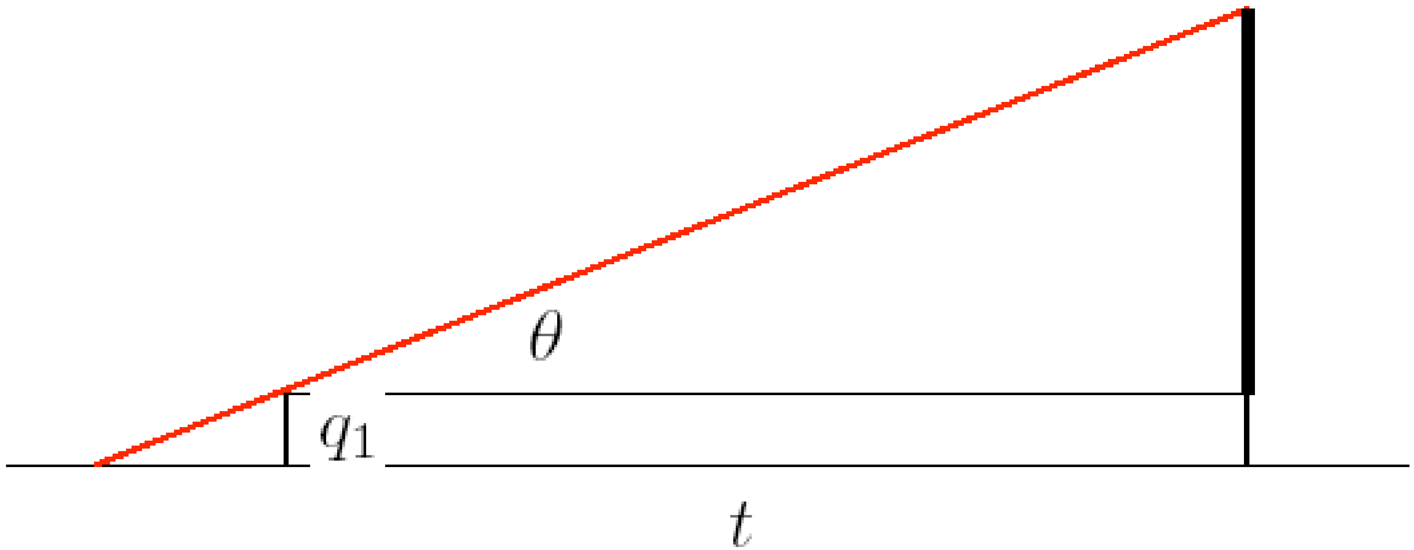
$$n_1\theta_1 = n_2\theta_2 + (n_1 - n_2)kq.$$

Using the variables $p = n\theta$ gives

$$\begin{pmatrix} q_2 \\ p_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -P & 1 \end{pmatrix} \cdot \begin{pmatrix} q_1 \\ p_1 \end{pmatrix} \qquad P = (n_2 - n_1)k.$$

P is called the **power** of the refracting surface.

The matrix of a translation.



In a medium of constant index of refraction the value of $p = n\theta$ does not change. If a distance t along the z-axis has been traversed, then $q_2 = q_1 + t \cdot \tan\theta$. Using the approximation $\tan\theta \doteq \theta$ we obtain

$$\begin{pmatrix} q_2 \\ p_2 \end{pmatrix} = \begin{pmatrix} 1 & T \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} q_1 \\ p_1 \end{pmatrix}, \quad T := t/n$$

The thin lens.

The product of two matrices of the form $\begin{pmatrix} 1 & 0 \\ -P & 1 \end{pmatrix}$ again has this same form. This gives the matrix of a "thin lens" where it is assumed that $z_1 = z_2$ for the two refracting surfaces. If we write $k_1 = \frac{1}{R_1}$ for the first refracting surface and $k_2 = \frac{1}{R_2}$ for the second refracting surface then the matrix of the thin lens is

$$\begin{pmatrix} 1 & 0 \\ -1/f & 1 \end{pmatrix}, \quad 1/f = (n_2 - n_1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right).$$

In case $n_2 - n_1 > 0$ and $R_1 > 0, R_2 < 0$ (a double-convex lens) $f > 0$. Here f is called the **focal length** of the lens.

The focal planes of the thin lens.

Suppose that $n_1 = 1$ and we consider the plane F_1 to the left of the lens at a distance f and the plane F_2 to the right of the lens at a distance f. The matrix between these two planes is

$$\begin{pmatrix} 1 & f \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1/f & 1 \end{pmatrix} \begin{pmatrix} 1 & f \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & f \\ -1/f & 0 \end{pmatrix}.$$

Now

$$\begin{pmatrix} 0 & f \\ -1/f & 0 \end{pmatrix} \begin{pmatrix} q_1 \\ p_1 \end{pmatrix} = \begin{pmatrix} fp_1 \\ -(1/f)q_1 \end{pmatrix}.$$

Here $p_2 = -(1/f)q_1$ does not depend on p_1. All rays passing through F_1 at a given height q_1 emerge as parallel rays at F_2.

Conjugate planes.

Two planes in an optical system are said to be **conjugate** or bd **in focus** with one another if, for all q_1 in first plane all rays emerging from q_1 end up at the same point in the second plane. In other words, q_2 depends only on q_1. In terms of the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ this says that $b = 0$. Suppose we consider the system

consisting of a plane s units to the left of a thin lens (in a region with constant index of refraction $n_1 = 1$) and similarly a plane t units to the right. So the matrix is

$$\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1/f & 1 \end{pmatrix} \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 - t/f & s + t - \frac{st}{f} \\ -1/f & 1 - s/f \end{pmatrix}.$$

The thin lens equation.

$$\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1/f & 1 \end{pmatrix} \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 - t/f & s + t - \frac{st}{f} \\ -1/f & 1 - s/f \end{pmatrix}.$$

So the planes will be in focus if and only if

$$s + t - st/f = 0$$

or

$$\frac{1}{s} + \frac{1}{t} = \frac{1}{f}$$

which is known as the **thin lens equation**. If $s \neq f$ we can solve this equation (uniquely) for t. If $s = f$ then "the conjugate plane is at infinity". If $s \neq f$ then

$$\frac{q_2}{q_1} = 1 - \frac{s}{f} = 1 - t\left(\frac{1}{s} + \frac{1}{t}\right) = -\frac{t}{s}$$

is the "magnification".

The telescope.

Consider a system consisting of two thin lenses of focal lengths f_1 and f_2 separated by a distance ℓ in air (where $n = 1$). The matrix is

$$\begin{pmatrix} 1 & 0 \\ -1/f_2 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & \ell \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ -1/f_1 & 1 \end{pmatrix} = \begin{pmatrix} 1 - \ell/f_1 & \ell \\ \frac{\ell}{f_1 f_2} - \frac{1}{f1} - \frac{1}{f_2} & 1 - \ell/f_2 \end{pmatrix}.$$

Suppose we choose $\ell = f_1 + f_2$. Then the matrix has the form

$$\begin{pmatrix} a & b \\ 0 & d \end{pmatrix}.$$

For such a system $p_2 = dp_1$ so the outgoing directions depend only on the incoming direction. We have

$$d = 1 - \frac{\ell}{f_1} = 1 - \frac{f_1 + f_2}{f_1} = -\frac{f_2}{f_1}.$$

So if we choose f_1 small and f_2 large, the "angular magnification" d will have a large absolute value (and be negative).

The principal planes.

We say that the matrix $\begin{pmatrix} a & b \\ 0 & d \end{pmatrix}$ is telescopic and that a matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with $c \neq 0$ is called **non-telescopic**. For such a matrix we proved that there exist unique s and t such that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ C & 1 \end{pmatrix} \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}.$$

Thus for any non-telescopic system there will be two conjugate planes with magnification one. Gauss called them the **principal planes**.

Hamilton's method in Gaussian optics.

Suppose that z_1 and z_2 are points on the optical axis of a Gaussian system which are *not* conjugate to on another. This means that the upper right hand corner of the optical matrix is not 0. So from the equations

$$\begin{aligned} q_2 &= aq_1 + bp_1 \\ p_2 &= cq_1 + dp_1 \end{aligned}$$

we can solve for p_1 and p_2 in terms of q_1 and q_2.

$$\begin{aligned} q_2 &= aq_1 + bp_1 \\ p_2 &= cq_1 + dp_1 \end{aligned}$$

$$\begin{aligned} p_1 &= \frac{1}{b}(q_2 - aq_1) \\ p_2 &= cq_1 + \frac{d}{b}(q_2 - aq_1) \\ &= \frac{1}{b}(dq_2 - q_1) \end{aligned}$$

where we have used the fact that $ad - bc = 1$. The fact that we can solve for for p_1 and p_2 in terms of q_1 and q_2 means that given a point q_1 on the z_1 "plane" and a point q_2 on the z_2 plane there is a unique light ray which joints them.

Here is still another way of saying the same thing. Consider the four dimensional space

$$\mathbb{R}^2 \oplus \mathbb{R}^2$$

with coordinates

$$\begin{pmatrix} q_1 \\ p_1 \\ q_2 \\ p_2 \end{pmatrix}$$

and with symplectic form

$$\Omega \left(\begin{pmatrix} q_1 \\ p_1 \\ q_2 \\ p_2 \end{pmatrix}, \begin{pmatrix} q_1' \\ p_1' \\ q_2' \\ p_2' \end{pmatrix} \right) := -(q_1 p_1' - p_1 q_1') + (q_2 p_2' - p_2 q_2').$$

$$\Omega\left(\begin{pmatrix} q_1 \\ p_1 \\ q_2 \\ p_2 \end{pmatrix}, \begin{pmatrix} q_1' \\ p_1' \\ q_2' \\ p_2' \end{pmatrix}\right) := -(q_1p_1' - p_1q_1') + (q_2p_2' - p_2q_2').$$

If $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is any matrix, the graph of M consists of all points in $\mathbb{R}^2 \oplus \mathbb{R}^2$ of the form

$$\begin{pmatrix} q_1 \\ p_1 \\ aq_1 + bp_1 \\ cq_1 + dp_1 \end{pmatrix}$$

so the subspace graph(M) is a Lagrangian subspace of $\mathbb{R}^2 \oplus \mathbb{R}^2$ if and only if $M \in SL(2)$.

By the very definition of a graph, we know that that the two dimensional subspace graph(M) projects bijectively onto the subspace given by $q_2 = 0, p_2 = 0$ under the projection $(q_1, p_1, q_2, p_2) \mapsto (q_1, p_1, 0, 0)$. To say that we are able to solve for p_1, p_2 in terms of q_1 and q_2 means that we can regard this two dimensional space as the graph of a map from (q_1, q_2) to (p_1, p_2) i.e. that graph(M) projects bijectively under the projection $(q_1, p_1, q_2, p_2) \mapsto (q_1, 0, q_2, 0)$.

So in terms of graph(M), the condition that we can solve for p_1 and p_2 in terms of q_1 and q_2 is the same as saying that graph(M) has only $\{0\}$ as its intersection with the two dimensional subspace of $\mathbb{R}^2 \oplus \mathbb{R}^2$ given by $q_1 = q_2 = 0$.

Let us return to our condition that we can solve for for p_1 and p_2 in terms of q_1 and q_2 and our equations

$$\begin{aligned} p_1 &= \frac{1}{b}(q_2 - aq_1) \\ p_2 &= \frac{1}{b}(dq_2 - q_1). \end{aligned}$$

Consider the function

$$W(q_1, q_2) = \frac{1}{2b}\left(aq_1^2 + dq_2^2 - 2q_1q_2\right) + K$$

where K is a constant. Then we can write the preceding equations as

$$p_1 = -\frac{\partial W}{\partial q_1}, \quad p_2 = \frac{\partial W}{\partial q_2}.$$

Hamilton called the function W the "point charateristic".

Hamilton's idea.

Here comes Hamilton's idea (in embryonic form): Suppose that we have three points z_1, z_2 and z_3 on the optical axis such that no two are conjugate and such that z_2 does not correspond to a refracting surface. We the have three point characteristics $W_{21} = W_{21}(q_1, q_2)$ for the z_1, z_2 system and similarly W_{32} and W_{31}. Then, up to an additive constant,

$$W_{31}(q_1, q_3) = W_{21}(q_1, q_2) + W_{32}(q_2, q_3)$$

where, in this equation, $q_2 = q_2(q_1, q_3)$ is taken to be the point where the ray joining q_1 and q_3 hits the z_2 "plane".

The characteristic function and the optical length.

Define the **optical length** of a broken line segment as follows: For a line segment of length ℓ contained entirely in a medium of (constant) refractive index ℓ define its optical length to be $L = n\ell$. A general path (in our approximation) is a concatenation of such line segments. Its optical length is the sum of the optical lengths of each segment. So

$$L(\gamma) := \sum n_i \ell_i.$$

A formula for the optical length.

Suppose that γ is a (portion of a) light ray between the z_1 and z_2 planes. I claim that

$$L(\gamma) = L_{\text{axis}} + \frac{1}{2}(p_2 q_2 - p_1 q_1)$$

where $\begin{pmatrix} p_i \\ q_i \end{pmatrix}$, $i = 1, 2$ are the parameters describing the ray at z_1 and z_2. Here L_{axis} denotes the optical length of the portion of the optical axis between z_1 and z_2.

The optical length and the characteristic function.

Before proving this result, let us apply it to the case where z_1 and z_2 are not conjugate, so that we may solve for p_1 and p_2 in terms of q_1 and q_2. Substituting

$$p_1 = \frac{1}{b}(q_2 - aq_1)$$

$$p_2 = \frac{1}{b}(dq_2 - q_1).$$

into $\frac{1}{2}(p_2q_2 - p_1q_1)$ gives

$$L(\gamma) = L_{\text{axis}} + \frac{1}{2b}\left(aq_1^2 + dq_2^2 - 2q_1q_2\right).$$

This is Hamilton's result that the generating function W can be taken to be the optical length of the light ray!

Proof of the formula for the optical length.

To prove that $L(\gamma) = L_{\text{axis}} + \frac{1}{2}(p_2 q_2 - p_1 q_1)$ observe that if the optical ray is the concatenation of two pieces, the expression adds, as it should. So it is enough to prove the result for rays of our two types: 1) a ray lying entirely in a medium of constant refractive index, or 2) a ray where z is at a refracting surface.

Case 1.

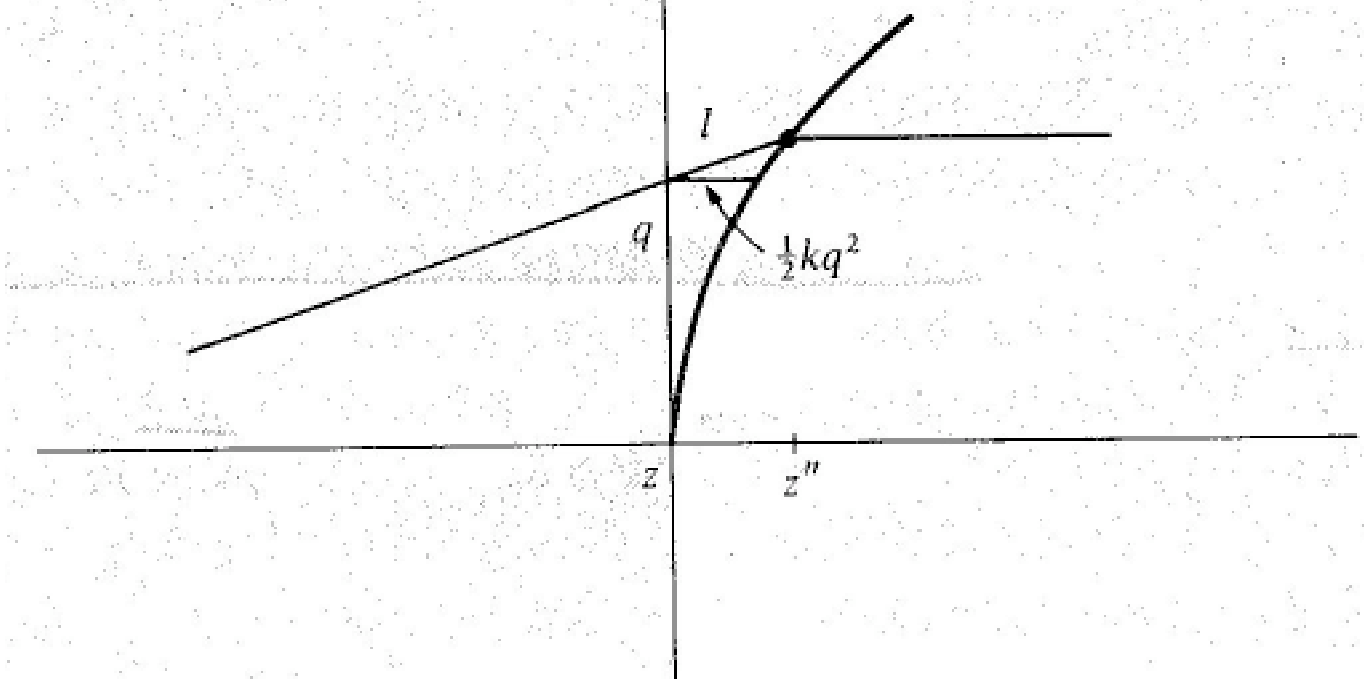
In the first case, if d is the length of the interval between z_1 and z_2 we have

$$\begin{aligned}
L(\gamma) &= n(d^2 + (q_1 - q_2)^2)^{\frac{1}{2}} \\
&\doteq nd + \frac{1}{2}\frac{n}{d}(q_2 - q_1)^2 \\
&= nd + \frac{1}{2}\left[\frac{n}{d}(q_2 - q_1)\right](q_2 - q_1) \\
&\doteq nd + \frac{1}{2}p(q_2 - q_1) \\
&= L_{\text{axis}} + \frac{1}{2}(p_2 q_2 - p_1 q_1)
\end{aligned}$$

since $p_1 = p_2 \doteq p = \frac{n}{d}(q_2 - q_1)$.

Case 2.

Now to the second case - a "surface" of refraction given by $z' - z = \frac{1}{2}kq^2$ with index of refraction n_1 to the left of the "surface" and index of refraction n_2 to the right of the surface. We will consider the case $k \geq 0$. the case $k \leq 0$ is similar.



The computation must be understood in the following sense. Suppose we choose some point $z_3 < z$ and some point $z_4 > z$. If n_1 were equal to n_2, the optical length between z_3 and z_4 would be $n_1\ell_3 + n_2(\ell + \ell_4)$ where ℓ_3 is is the length of the ray from z_3 to z, where $\ell + \ell_4$ is the length of the ray from z to z_4 and where

$$\ell \doteq \frac{1}{2}kq^2$$

is the length of the ray from z to the refracting surface.

The computation must be understood in the following sense. Suppose we choose some point $z_3 < z$ and some point $z_4 > z$. If n_1 were equal to n_2, the optical length between z_3 and z_4 would be $n_1\ell_3 + n_2(\ell + \ell_4)$ where ℓ_3 is is the length of the ray from z_3 to z, where $\ell + \ell_4$ is the length of the ray from z to z_4 and where

$$\ell \doteq \frac{1}{2}kq^2$$

is the length of the ray from z to the refracting surface. If $n_1 \neq n_2$ we must modify this (again up to higher order effects) by $(n_1 - n_2)\ell$. So the contribution of the refracting surface is (up to higher order)

$$(n_1 - n_2) \cdot \frac{1}{2}kq^2 = \frac{1}{2}[(kn_1 - kn_2)q]q = \frac{1}{2}(p_2 - p_1)q$$

since

$$q = q_1 = q_2, \quad \text{and } p_2 = p_1 + (n_1 - n_2)kq$$

at a refracting surface.