CHAPTER I

GENERAL PRINCIPLES

Several approaches are possible starting from either Fermat's principle, or Huyghens principle, or the experimental determination of the laws of refraction and reflection; they all lead to a good agreement with the experimental observations. It thus depends on the problem at hand to choose the most appropriate approach. We will see one example where Fermat's principle is used to study gradient index media (non homogeneous media); another example will be seen in wave optics in studying birefringent (anisotropic) media where Huyghens construction is much used.

It may be useful to recall the notions of homogeneous and isotropic media, which will often be used in these lectures.

**A medium is <u>homogeneous</u> for a light wave if its index does not depend on the location inside the medium at the scale of the wavelength.*

**A medium is <u>isotropic</u> for a light wave if its index does not depend on the direction of polarization of the light.*

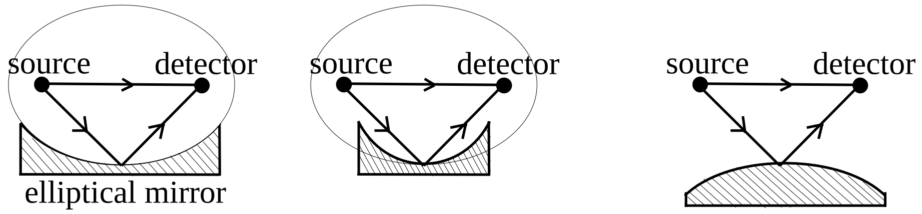
Examples: An optical fiber with an index of refraction decreasing with the distance to the optical axis (gradient-index fiber) is an example of inhomogeneous but isotropic medium. A perfect crystal with a lattice that is not symmetric at the atomic level (birefringent medium such as quartz) is an anisotropic but homogeneous medium.

Here Fermat's principle has been chosen as a starting point, and we will see how we can deduce from it the fundamental laws of propagation of light.

I. Fermat's principle

When you are searching for a general principle describing propagation of light, it is natural at first to assume that light takes the path corresponding to the shortest time. Indeed, such a principle can describe well-known properties, such as linear propagation of light in a homogeneous and isotropic medium, or the properties of light to retrace its path in the reverse direction. It also leads to the laws of reflection and refraction at a plane interface.

However, it is not able to describe the situation where a source and a detector are placed next to each other close to a curved mirror. First, there are two possible paths for the light: the absolute shortest is the straight line connecting the source to the detector, but another one, longer, is possible including a reflection on the mirror. In addition, depending on the curvature of the mirror, this second path can be a local minimum, a local maximum or a constant. The constant corresponds to the case when the mirror is an ellipse with the source and the detector as its foci, the local maximum to a mirror with a curvature larger than that of the ellipse, and the local minimum to a smaller curvature (such as a plane or convex mirror).



elliptical mirror constant time except for direct path

local maximum time for reflected path

local minimum for reflected path

The physicist Pierre de Fermat expressed a more complete principle for light propagation, including in particular the above situation:

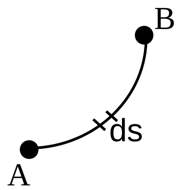
Fermat's principle:

Light takes the path corresponding to a stationary optical path length.

Let us first recall the definition of the terms used:

1) definition of the optical path length

We consider a random trajectory between two points A and B (i.e. not necessarily a trajectory followed by light), in a medium characterized by its index n, which depends on the position in the medium.



If light were to follow this path, it would take a time T to travel from A to B that can be written:

$$T=\int_A^B \frac{ds}{v}$$

where $v=c/n$ is the light wave propagation velocity in a medium with index n.

The optical path length L corresponding to this trajectory is the distance that light would travel during time T if it were propagating in vacuum. L thus is given by:

$$L=cT=\int_A^B n\,ds$$

2) stationary optical path length

Stationary means that the optical path length varies slowly when we deform the trajectory, which in mathematical terms is written as:

$$\partial L=0 \text{ in the first order in } \partial M$$

where ∂M is an elementary displacement of a point M chosen randomly along the AB trajectory, apart from the extremities A and B.

3) discussion

This expression of Fermat's principle now allows us to describe the reflection on the curved mirror: a minimum (absolute or local) of travel time, or a maximum, or a constant are all included in "stationary". We will also see in chapter III that Fermat's principle is well suited to determine the trajectories followed by light in inhomogeneous media (gradient-index fibers for example).

We will often abusively talk about optical path between points A and B without specifying for which trajectory, implying the path effectively followed by light (thus the one that is stationary). In the particular case of a homogeneous and isotropic medium with a uniform index n, this optical path will simply be $L = n\overline{\mathrm{AB}}$. In the case of an anisotropic medium, rays and wave vectors are no longer collinear and the optical path must be calculated along the wave vectors taking into account the index corresponding to the phase velocity of the light wave.

The physical picture that we get from Fermat's principle goes further than the framework of ray optics. Indeed, it suggests that light « tries » different paths to find out which is the shortest. This happens over a length scale of the order of the wavelength, and it allows us to interpret in these terms the experiment of diffraction through a circular aperture (spreading of light under the influence of diffraction is larger if the size of the aperture is closer to the wavelength of light). This type of variational approach in terms of possible paths is found also in other fields of physics: principle of « least action » in classical mechanics, Feynman path integrals in quantum mechanics (I recommend reading the chapters relative to optics in the Feynman lectures in physics).

II. Consequences of Fermat's principle - Snell-Descartes laws

We have already seen two immediate consequences of Fermat's principle:

- linear propagation of light **in the case of a homogeneous medium**

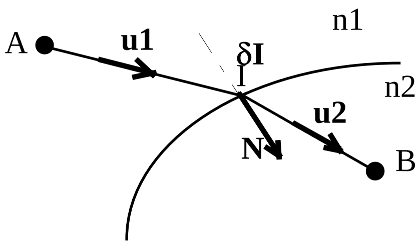
- the path followed by light does not depend on the direction of propagation, for any propagation medium (note that this does not specify what happens in terms of energy in the beams going one way or the other, in particular when you take into account the polarization of light).

We will now see how Fermat's principle allows us to find the laws of refraction and reflection through a surface between two homogeneous media.

1) Vectorial and angular expression for Snell-Descartes laws

Depending on the country, these laws can be called Descartes laws ('lois de Descartes' in France for example) or Snell's laws (in English speaking countries, e.g.). Indeed there were established independently by the Dutch Willebrord Snell in 1621 and by the French René Descartes in 1637. A few centuries earlier, the Greeks had already established empirically the law for reflection and done precise measurements on refraction (without finding the actual law). We will here invert the chronology since we will demonstrate these laws starting from the principle stated by the French Fermat in 1657.

We will first consider the case of refraction that we will then generalize to reflection. Let us consider two points A and B located on each side of a refractive surface S, surface separating two homogeneous media with indices n_1 and n_2. We know that the paths in homogeneous media are straight lines, we are then left with the task of determining the point of impact I on the refractive surface corresponding to an optical path from A to B that is stationary.



The optical path AIB is given by:

$$L = n_1\overline{AI} + n_2\overline{IB} = n_1\mathbf{AI}.\mathbf{u_1} + n_2\mathbf{IB}.\mathbf{u}_2$$

Here as in the rest of this text, vectors will be represented in bold font (not with arrows above).

The optical path actually followed by light must be stationary with respect to a displacement of I of a small amount ∂I along the refractive surface, which is given by:

$$\Delta L = n_1\partial\mathbf{I}.\mathbf{u_1} - n_2\partial\mathbf{I}.\mathbf{u_2} = 0$$

(the terms of the type $\mathbf{AI}.\partial\mathbf{u_1}$ are equal to zero since the modulus of $\mathbf{u_1}$ is a constant).
The vector $n_1\mathbf{u_1} - n_2\mathbf{u_2}$ is thus orthogonal to the vector $\partial\mathbf{I}$, that is to say it is collinear to the normal to the surface **N**, that can be written as (where a is a real number):

$$n_1\mathbf{u_1} - n_2\mathbf{u_2} = a\mathbf{N}$$

We can write the expression of the coefficient a as a function of the angles i_1 and i_2 (defined between -90° and +90°, and counted algebraically from the normal towards the ray) by taking the scalar product of the above relation with the vector **N**, and we get:

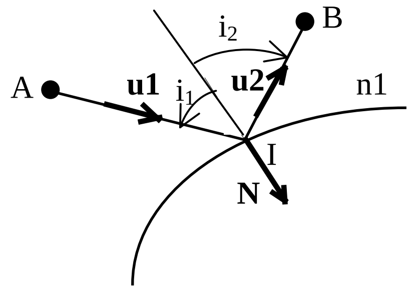
$$n_1\mathbf{u_1} - n_2\mathbf{u_2} = (n_1\cos i_1 - n_2\cos i_2)\mathbf{N}$$

This is Snell-Descartes' law in its vectorial form, relative to refraction. We can write it in an equivalent way using the following two properties:

- the refracted ray is included in the incidence plane, plane defined by the incident ray and the normal to the refractive surface;
- when you are set in the plane of incidence and if you take the vector product of the previous law with **N**, you get the law for refraction:

$$n_1\sin i_1 = n_2\sin i_2$$

The <u>case for reflection</u> can be treated in an analogous way as refraction defining the vectors $\mathbf{u_1}$ and $\mathbf{u_2}$ in the following way:



We show that: $\mathbf{u_1}-\mathbf{u_2}=2\cos i_1\mathbf{N}$

Or in other words:

- the reflected ray is included in the plane of incidence
- $i_1=-i_2$

We can note that the laws of reflection are included in the laws of refraction if we write $n_1=n_2$ and the rule specified above about unitary vectors.

2) A few practical consequences

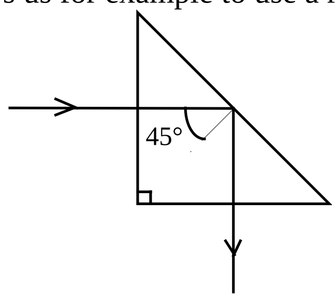
a) the ray gets closer to the normal when it refracts inside a medium more refringent (i.e. with a higher index of refraction) and it moves away from the normal when it enters a less refringent medium.

b) the fact that light takes the same path when it travels in the opposite direction shows that after refraction through a plane parallel plate (or several such plates stacked), a ray always comes out parallel to itself.

c) when entering through a less refringent medium (for example from water with index 1.3 into air with index 1), there is a limit angle above which there will not be any more refraction. This is the phenomenon called total internal reflection which happens for a limit incidence angle i_1 that is given by:

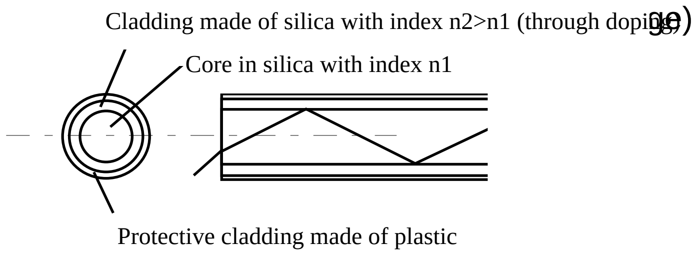
$$\sin i_1=\frac{n_2}{n_1}$$

This angle is about 42° for a glass (n=1.5)/air interface and it is 49° for a water/air interface. This phenomenon allows us for example to use a right-angle prism as a mirror:



This type of mirror is interesting because it is perfectly reflecting (as long as the surface is clean) for any wavelength, and it can handle high power light since there is no absorption (except for that of glass, which can be very small if you use fused silica). It is thus used in the ultraviolet part of the spectrum where good reflecting coatings are difficult to make and also for pulse lasers when instantaneous power is very high.

Another important application of this phenomenon of total internal reflection is the propagation of light inside optical fibers (or inside waveguides in general). The following schematic shows an example of such propagation for a step-index fiber (core with a slightly higher index that the cladding):



d) from an energetic point of view, the only information that geometrical optics gives us is the existence of this limit angle above which the reflection coefficient is equal to 100%. The electromagnetic theory is necessary to determine the energy distribution between the reflected and the refracted waves, which depends on the angle of incidence and of the polarization state of the incident light. Without entering into the detailed expressions for reflection and transmission coefficients (they can be found in the electromagnetic course or in reference books), it is useful to know the following results:

- close to normal incidence, the reflection coefficient in intensity between two media with indices n_1 and n_2 is:

$$R=\left(\frac{n_1-n_2}{n_1+n_2}\right)^2$$

In particular on a glass surface with index 1.5, 4% of the incident intensity is reflected when on the edges of a laser diode made of AsGa with index 3.5 that same reflection coefficient is equal to 30% (for visible wavelengths).

- there is an angle of incidence for which the reflection coefficient is equal to zero for one specific polarization state (the polarization, i.e. direction of the electric field, parallel to the plane of incidence): it is the Brewster's angle given by $\tan i_B = n_2 / n_1$. This angle is equal to approximately 56° for an air/glass interface.

III. Relationship between rays and wave surfaces: Malus-Dupin's theorem.

As we mentioned in the previous chapter, geometrical optics is an approximation that allows us to characterize in a simpler way the propagation of light in terms of light rays. We can for example, thanks to the Snell-Descartes' laws, calculate the trajectory of any ray through an optical system, and do that for a large number of rays (basis for computer-based optical design). It is then important to connect these rays to the wave nature of light, in order to be able to include the interference and diffraction effects.

We will show below how rays can be related to the wave surfaces **in the case of isotropic medium.**

1) Reminder about wave surfaces

A wave surface is an equal phase surface for the electromagnetic field. It is constructed starting from a point source reporting the same optical path along all possible trajectories. For a point source S at a finite distance in a homogeneous and isotropic medium, the wave surfaces are thus spheres centered at S. If the point source is located at infinity, the wave surfaces become wave planes.

2) Malus-Dupin's theorem

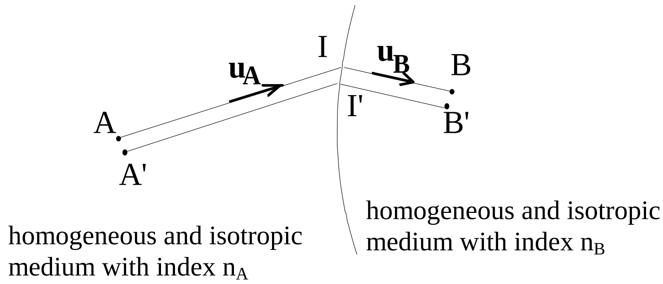
In an isotropic medium, the wave surfaces are orthogonal to the light rays.

This property can be demonstrated starting from Maxwell's equations: we then show that the equal phase surfaces are orthogonal to the wave vector, parallel to the Poynting vector in the case of an isotropic medium. Here we are going to prove it using Fermat's principle.

a) differential of the optical path along a light ray

We will consider here only the case of a succession of homogeneous and isotropic media separated by refractive surfaces. We will see in chapter III that this expression also extends to the case of inhomogeneous media.

Let us consider a light ray connecting point A to point B, passing through a refractive surface:



The optical path AB is equal to:

$$L=(AB)=n_A\mathbf{u_A}.\mathbf{AI}+n_B\mathbf{u_B}.\mathbf{IB}$$

The differential of this optical path, i.e. the difference to a nearby trajectory A'B', is given by:

$$\partial L=n_A\mathbf{u_A}.(\partial\mathbf{I}-\partial\mathbf{A})+n_B\mathbf{u_B}.(\partial\mathbf{B}-\partial\mathbf{I})$$

$$\partial L=(n_A\mathbf{u_A}-n_B\mathbf{u_B}).\partial\mathbf{I}+n_B\mathbf{u_B}.\partial\mathbf{B}-n_A\mathbf{u_A}.\partial\mathbf{A}$$

Since AIB is a path followed by light, the Snell-Descartes' law is verified at the interface: the first term is ∂L is thus equal to zero. We are left with the following expression for the differential of the optical path:

$$\partial L=n_B\mathbf{u_B}.\partial\mathbf{B}-n_A\mathbf{u_A}.\partial\mathbf{A}$$

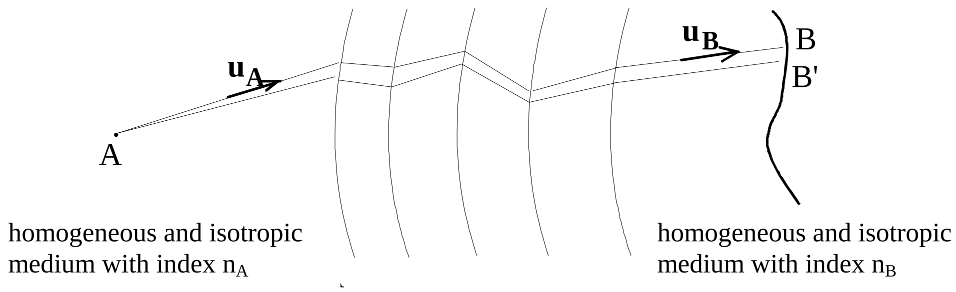
Along the light ray, the differential of the optical path depends only on the displacements of its extremities. This expression will be used several times in the following paragraphs.

b) proof of Malus-Dupin's theorem

The wave surfaces are constructed starting from a point source A reporting along the light rays (trajectories that can be followed by light) equal optical paths. We will thus prove that, from one homogeneous medium to the next, the wave surfaces are necessarily orthogonal to the rays.

Let us consider a point source A in a first homogeneous and isotropic medium. The light rays are straight lines, and the wave surfaces originating from A are spheres: there are indeed orthogonal to the light rays. The property is true in that first medium.

Let us consider now two points B and B', close to each other, located on the same wave surface in an other medium:



The light rays connecting A to B and A' to B' correspond to the same optical path by definition of the wave surface. On the other hand we can write the difference between these two optical paths using the above relation, using the fact that the initial point A did not change:

$$\partial L = n_B \mathbf{u_B} . \mathbf{BB'}$$

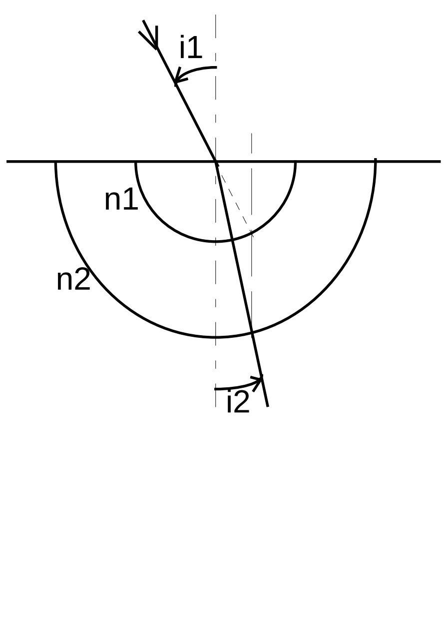
∂L is equal to zero, thus the ray is orthogonal to the wave surface.

IV Huyghens' construction and construction using the index surfaces

We will see two graphical methods to trace rays refracted by a plane refractive surface: the first one used Snell-Descartes' laws; the second one is based on Huyghens' principle, that we will remind you here.

1) Construction using the index surfaces

Snell-Descartes' laws in terms of angles allow us to construct the refracted and reflected rays geometrically starting from the incident ray as shown on the following figure:

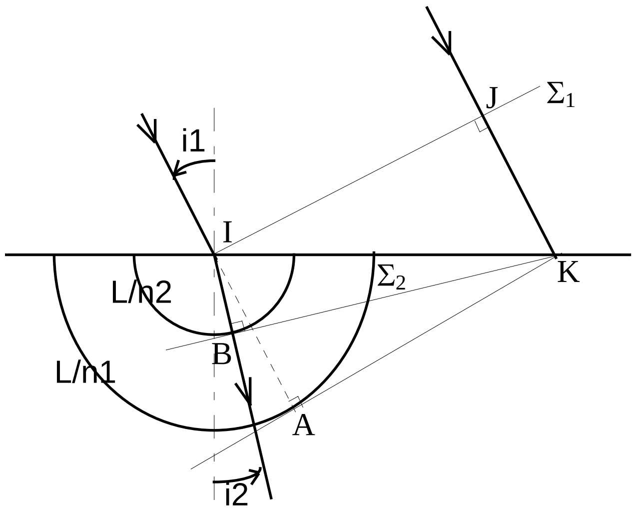


This type of construction will not be often used in the next chapters of this course. However it will be found again in the wave optics course when propagation through anisotropic media will be studied.

2) Huyghens' construction

Huyghens' principle can be expressed in the following form: light propagates through successive steps. Each point of a wave surface acts as a secondary source that emits small spherical wavelets (in an isotropic medium). The envelope of the wavelets corresponding to the same optical path makes a new wave surface.

Let us consider a plane wave incident on a plane refractive surface with an angle of incidence i_1 with respect to the normal.



Let us consider one ray corresponding to this wave and trace from the impact point I on the surface a circle with radius L/n_1 (L is an arbitrary length). In the absence of the refractive surface, light would have reached point A with a delay L/c with respect to point I.

We then trace a second incident ray, parallel to the first one, such that it reaches the surface at a point K at the same time when we reach point A. We construct this ray by tracing the tangent to the circle L/n_1 at A: K is at the intersection of the tangent and the refractive surface.

We apply Huyghens' principle starting from the incident wave plane Σ_1 passing through I (J on the second ray). We want to construct the wave surface Σ_2 reached after a time L/c. The distance JK is exactly equal to L/n_1, so K belongs to the new wave surface Σ_2. From point I, light propagates through the medium of index n_2 thus the distance covered equals L/n_2. The wavelet originating from I corresponds to a sphere with center I and radius L/n_2. The new wave surface is the envelope of the wavelets: it is thus tangent to the sphere of radius L/n_2. At last, we obtain Σ_2 by tracing the tangent to the circle with radius L/n_2 passing through K.

The refracted ray is normal to the new wave surface: it thus passes through point B, where Σ_2 tangents the L/n_2 sphere.

Note: we can also justify this construction using Snell-Descartes' laws noting that:

$$\frac{IA}{IK}=\sin i_1 \quad \text{and} \quad \frac{IB}{IK}=\sin i_2$$