Foundations of Geometrical Optics

Section 1

Introduction

Optics

Optics is the field of science and engineering encompassing the physical phenomena associated with the generation, transmission, manipulation, detection and utilization of light.

From National Research Council Report: “Harnessing Light”

OPTI-502 Optical Design and Instrumentation I
© Copyright 2019 John E. Greivenkamp
College of Optical Sciences
The UNIVERSITY OF ARIZONA

History of Optics

In the beginning… "Let there be light"

~300 BC	Euclid (Alexandria) in his *Optica* noted that light travels in straight lines and described the law of refraction. He believed that vision involved rays going from the eye to the object.
100-150 BC	Hero (Alexandria) studied reflection for a plane mirror and showed that the actual ray path was the shortest of all possible paths.
140 AD	Ptolemy (Alexandria) studied refraction and suggested that the angle of refraction was proportional to the angle of incidence.
1000	Alhazen (Bazra) investigated magnification produced by lenses, atmospheric refraction, and spherical and parabolic mirrors. He was aware of spherical aberration. His work was translated into Latin and available to later European scholars.
1267	Roger Bacon (England) considered the speed of light to be finite. He studied the magnification of small objects using convex lenses, and he attributed rainbows to the reflection of sunlight from individual raindrops.
1268-1289	The invention of eyeglasses in Italy. The lenses were made of natural crystal.
1590	Zacharius Jensen (Netherlands) constructed a compound microscope.
1608	Hans Lipperhay (Netherlands) constructs the first telescope using a positive objective and a negative eye lens.

History of Optics

1609	Galileo Galilei (Italy) constructs a version of Liperhay's telescope and begins astronomical investigations, reporting the moons of Jupiter in 1610.
1611	Johannes Kepler (Germany) in his book, *Dioptrics*, he presents the use of positive and negative lenses in telescopes and microscopes. He proposes the design of a telescope using two positive lenses.
1621	Willebrord Snell (Netherlands) discovers the relationship between the angle of incidence and the angle of refraction.
1657	Pierre de Fermat (France) states the principle of least time for determining ray paths.
1666	Issac Newton (England) proves that white light is composed of individual colors by using prisms.
1668	Issac Newton (England) constructs the first reflecting telescope to solve the problem of chromatic aberration.
1678	Christian Huygens (Netherlands) publishes his wave theory light based upon wavelets propagating in an all-pervading ether.
1704	Issac Newton (England) publishes his book, *Opticks*.
1733	Chester More Hall (England) invents the achromatic doublet to correct chromatic aberration in lenses.
1758	John and Peter Dollond (England) patent and commercialize that achromatic doublet.

History of Optics

1801	Thomas Young (England) proves the wave nature of light by demonstrating interference.
1821	Augustin Jean Fresnel (France) presents the relationships describing the intensity and polarization of reflected and refracted light.
1823	Joseph Fraunhofer (Germany) publishes his theory of diffraction.
1835	George Airy (England) calculates the diffraction pattern produced by a circular aperture.
1873	James Clerk Maxwell (England) publishes *A Treatise on Electricity and Magnetism* which contains his four equations relating electric and magnetic fields.
1879	Thomas Alva Edison (US) develops the electric light.
1887	Michelson and Morely (US) are unsuccessful in detecting the motion of the earth through the stationary luminiferous ether.
1896	Wilhelm Wein (Germany) described the spectrum of a thermal or black body source.
1899	Lord Rayleigh (England) develops scattering theory to describe the blue color of the sky.
1900	Max Planck (Germany) explains the spectrum of a black body source by requiring a quantum of action (Planck's Constant).
1905	Albert Einstein (Germany) explain the photoelectric effect on the basis that light is quantized into what are now called photons.

History of Optics

1916	Albert Einstein (Germany) proposes the process of stimulated emission.
1932	Edwin Land (US) invents polarizing film (Polaroid).
1948	Dennis Gabor (Hungary) invents holography.
1958	Arthur Schawlow and Charles Townes (US) propose the principles that will give rise to the visible laser.
1960	Theodore Maiman (US) constructs the first laser using a rod of synthetic ruby.
1961	Ali Javan, W. Bennett and Donald Herriott (US) demonstrate the first gas laser using helium and neon.
1990	The Hubble Space Telescope is launched.

Portions of this history are taken from R. Victor Jones:
http://people.seas.harvard.edu/~jones/ap216/local_copies/history_of_optics/hist.htm

OPTI-502 Optical Design and Instrumentation I
© Copyright 2019 John E. Greivenkamp
College of Optical Sciences
The University of Arizona

Theories of Optics

Quantum Theory:
- Energy levels
- Probability densities
- Photons

Wave Theory:
- E & M
- Interference
- Diffraction
- Polarization

Geometrical Theory:
- Reflection
- Refraction
- Optical design
- Aberrations
- Radiometry

Geometrical optics is the study of light in the limit of short wavelengths.

Light propagates as rays.

Geometrical optics usually ignores interference, diffraction, polarization and quantum effects.

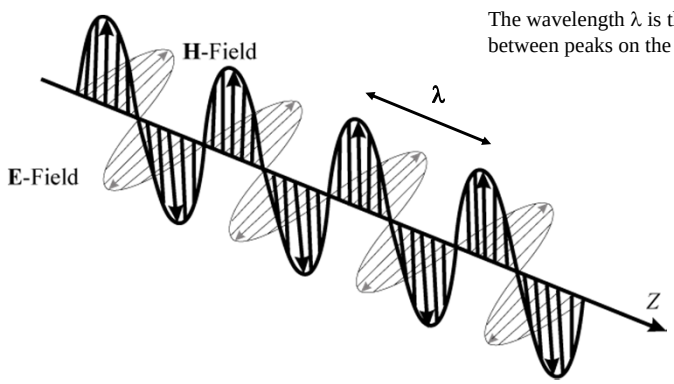
Geometrical optics includes:

- imaging properties (location, magnification)
- aberrations (image quality)
- radiometry (how much light)

What is Light??

Light is a self-propagating electro-magnetic wave. The electric and magnetic fields are perpendicular or transverse to the direction of propagation.

The wavelength λ is the distance between peaks on the wave.



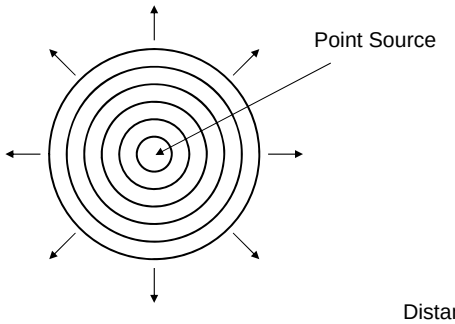
In vacuum, the EM wave propagates a the speed of light:

$$c = 2.99792458 \times 10^8 \text{ m/s}$$

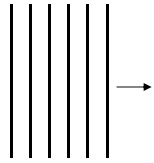
OPTI-502 Optical Design and Instrumentation I
© Copyright 2019 John E. Greivenkamp
College of Optical Sciences
The University of Arizona

Propagation and Wavefronts

A point source in a homogeneous medium will produce an expanding spherical wave. A wavefront is a surface of constant propagation time from the source.



At a great distance from the source, a portion of the spherical wave will appear to be a plane wave.



There is a direct analogy to water waves, which are also transverse waves. The water moves up and down – the water does not propagate, but the wave does.

Wavelength and Frequency

If the wave is propagating with a velocity or speed V, the frequency ν of the wave is the number of cycles or wavelengths that pass a point in one second.

The time for one wavelength to pass is $T = \frac{\lambda}{V}$

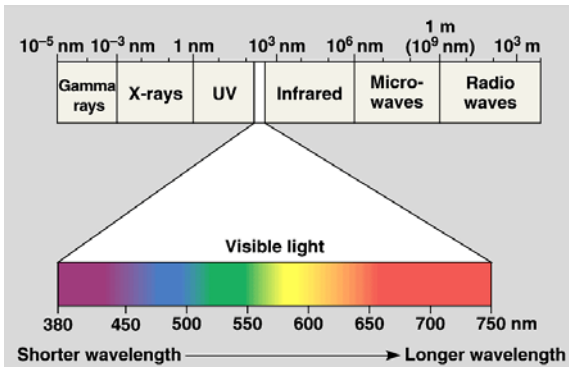
The frequency is then $\frac{1}{T} = \nu = \frac{V}{\lambda}$

$$\lambda = \frac{V}{\nu} \quad \text{in vacuum: } \lambda = \frac{c}{\nu}$$

Units:		
	λ	m
	V	m/sec
	ν	1/sec or Hz

OPTI-502 Optical Design and Instrumentation I
© Copyright 2019 John E. Greivenkamp
College of Optical Sciences
The University of Arizona

Electro-Magnetic Spectrum



~ 350 nm < λ < ~ 750 nm for Visible Light

~ 0.35 µm < λ < ~ 0.75 µm for Visible Light

The field of optics often implies only the visible portion of the spectrum, but more generally optics includes both the ultraviolet and infrared portions of the spectrum.

Example:

$$\lambda = 550\ nm = 0.55\ \mu m \quad \text{(Green)}$$

$$c = 2.99792458 \times 10^{8}\ \text{m/s}$$

$$\nu = 5.45 \times 10^{14}\ Hz$$

Index of Refraction

Speed of light in vacuum

$$c = 2.99792458 \times 10^{8}\ \text{m/s}$$

Index of Refraction n

$$n \equiv \frac{\text{Speed of Light in Vacuum}}{\text{Speed of Light in Medium}} = \frac{c}{V} \qquad V = \frac{c}{n}$$

The index of refraction tells how much light slows down in a medium compared to its speed in a vacuum.

Note that in a medium, the frequency does not change as this is determined by the source oscillation. The wavelength changes with the index.

$$\lambda_M = \frac{\lambda_0}{n}$$

Example:

$$\lambda_0 = 550\ nm = 0.55\ \mu m \quad \text{(Green)}$$

$$\nu = 5.45 \times 10^{14}\ Hz$$

$$n = 1.5$$

$$V = 1.993 \times 10^{8}\ \text{m/s}$$

$$\lambda_M = 367\ nm = 0.367\ \mu m$$

Typical Indices of Refraction

vacuum		1.0
helium		1.000036
hydrogen		1.000132
air		1.000293
water		1.33
fused silica		1.46
plastics		1.48-1.6
borosilicate crown glass		1.51
crown glass		1.52
light flint glass		1.57
dense barium crown glass		1.62
dense flint glass		1.72
diamond		2.4
Silicon	@ 10 µm	3.4
Germanium	@ 13 µm	4.0

Geometrical Optics

Any object is comprised of a collection of independently radiating point sources. Each source is infinitesimally small; there is no interference. Each point source is independently imaged through the system, and the image is the superposition of intensity patterns from all of the point images.

First-order optics is the study of perfect optical systems, or optical systems without aberrations. Analysis methods include Gaussian optics and paraxial optics. Small angle approximations are used.

Imaging properties (image location and magnification)
Radiometric properties
Preliminary system design and layout

Aberrations are the deviations from perfection of the optical system. These aberrations are inherent to the design of the optical system, even when perfectly manufactured. Additional aberrations can result from manufacturing errors.

Third-order optics (and higher-order optics) includes the effects of aberrations on the system performance. The effects of diffraction are sometimes included in the analysis.

Image quality
Optical design

Geometrical optics principles can be applied to radiation of any wavelength.

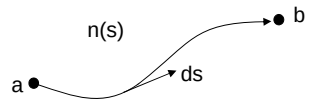
Optical Path Length

The Optical Path Length OPL is proportional to the time it takes light to propagate from point a to point b along a ray.

$$\Delta T = \int_a^b \frac{ds}{V(s)}$$

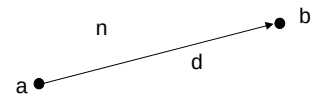
$$\Delta T = \frac{1}{c}\int_a^b n(s)ds$$

$$OPL = \int_a^b n(s)ds$$



In a homogeneous medium:

$$OPL = nd$$



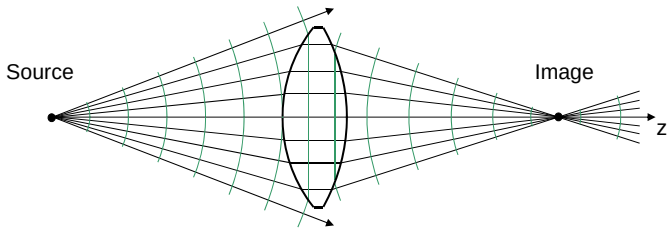
The OPL is the distance that light travels in a vacuum in the same time the light travels a distance d in the medium.

OPTI-502 Optical Design and Instrumentation I
© Copyright 2019 John E. Greivenkamp
College of Optical Sciences
The UNIVERSITY of ARIZONA

Wavefronts and Rays

Wavefronts are surfaces of constant OPL from a source point.

Rays indicate the direction of energy propagation and are normal to the wavefront surfaces. If n is constant, the rays are straight lines.



In a perfect optical system or a first-order optical system:

- All wavefronts are spherical or planar.
- The OPL along each ray from the object point to the image point is constant.

If n is not constant, the rays may be bent. An example is GRIN or Gradient Index materials.

Fermat's Principle

A method for determining valid ray paths.

Popular form:

The path taken by a light ray in going from point a to point b through any set of media is the one that takes the least time.

Correct Form:

The path taken by a light ray in going from point a to point b through any set of media is the one that renders its OPL equal, in the first approximation, to other paths closely adjacent to the actual path.

The OPL of the actual ray is either an extremum (a minimum or a maximum) with respect to the OPL of adjacent paths or equal to the OPL of adjacent paths. The actual ray path is stationary with respect to closely adjacent paths

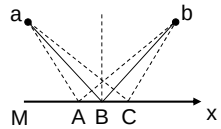
In a first-order or paraxial imaging system, all of the light rays connecting a source point to its image have equal OPLs.

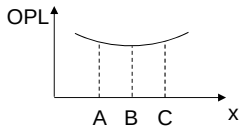
Fermat's principle implies that rays are straight lines in a material of constant index since a straight line connecting two points is shorter than a curved line connecting the points.

OPTI-502 Optical Design and Instrumentation I
© Copyright 2019 John E. Greivenkamp
College of Optical Sciences

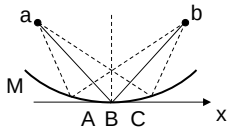
Fermat Examples

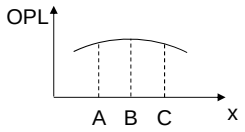
Minimum OPL – a plane mirror



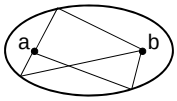


Maximum OPL – a concave mirror





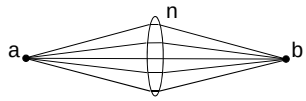
Equal OPL – an elliptical reflector



(a and b are the foci of the ellipse)

An ellipse can be drawn with a piece of string of fixed length with its ends fixed at the foci.

Equal OPL – an imaging lens



(a and b are object and image)

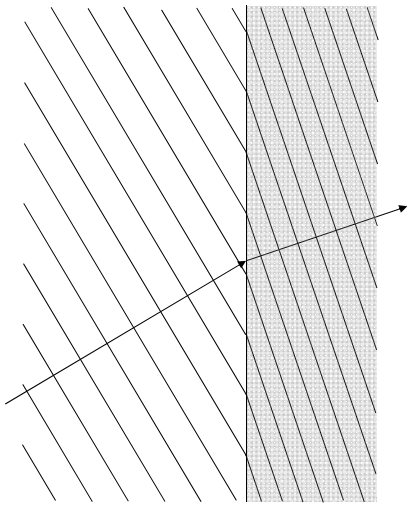
The extra OPL through the center of the lens (nt) compensates for the extra distance to the edge of the lens.

Refraction and Wavefronts

The wavefront spacing is different across the boundary, and the wavefronts must meet at the interface.

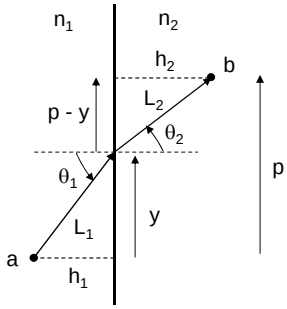
The Law of Refraction or Snell's Law can be easily derived from this condition.

The same arguments and analysis lead to the Law of Reflection.



Refraction at a Surface

Use Fermat's Principle to determine the valid ray path across a refractive boundary. The ray goes from point a to point b. The variable y defines the ray intersection at the interface:



$$OPL = n_1 L_1 + n_2 L_2$$

$$L_1 = \sqrt{h_1^2 + y^2}$$

$$L_2 = \sqrt{h_2^2 + (p-y)^2}$$

$$\frac{dOPL}{dy} = n_1 \frac{dL_1}{dy} + n_2 \frac{dL_2}{dy} = 0 \quad \text{for a valid ray path}$$

$$\frac{dL_1}{dy} = \frac{y}{\sqrt{h_1^2 + y^2}} = \frac{y}{L_1} = \sin\theta_1$$

$$\frac{dL_2}{dy} = \frac{-(p-y)}{\sqrt{h_2^2 + (p-y)^2}} = -\frac{(p-y)}{L_2} = -\sin\theta_2$$

Then

$$n_1 \sin\theta_1 - n_2 \sin\theta_2 = 0$$

$$n_1 \sin\theta_1 = n_2 \sin\theta_2$$

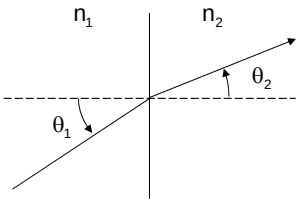
This is, of course, Snell's Law

Snell's Law of Refraction

For refraction at a surface,

$$n_2 \sin\theta_2 = n_1 \sin\theta_1$$

where n_1 and n_2 are the indices of refraction in the two media and θ_1 and θ_2 are the angles between the rays and the surface normal.



An important second part of this law is that the incident ray, the refracted ray and the surface normal must be coplanar.

For refraction at a curved surface, the surface slope and angles at the ray intercept are used.

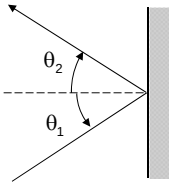
When propagating through a series of parallel interfaces, the quantity n sinθ is conserved.

Law of Reflection

For reflection at a surface, the angle of incidence equals the angle of reflection:

$$\theta_2 = -\theta_1$$

Both angles are measured relative to the surface normal, and the minus sign is due to sign conventions.



Once again, the two rays and the surface normal must be coplanar.

The law of reflection is often considered a special case of Snell's law where $n_2 = - n_1$. The sign of the index of refraction is now a function of propagation direction.

Following a reflection, light propagates from right to left, and its velocity can be considered to be negative. Using velocity instead of speed in the definition of *n*, the index of refraction is now also negative.

$$n \equiv \frac{\text{Speed of Light in Vacuum}}{\text{Speed of Light in Medium}} = \frac{c}{V}$$

OPTI-502 Optical Design and Instrumentation I
© Copyright 2019 John E. Greivenkamp
College of Optical Sciences
The UNIVERSITY of ARIZONA

Critical Angle and Total Internal Reflection TIR

When a ray propagates from a high index medium to a low index medium, the angle of refraction is larger than the angle of incidence. For a sufficiently large incident angle, there is no solution to Snell's law:

$$n_1 \sin\theta_1 = n_2 \sin\theta_2 \quad \text{and} \quad n_1 > n_2$$

There is no solution when

$$\sin\theta_2 = \frac{n_1}{n_2}\sin\theta_1 > 1$$

Under this condition, the light ray is reflected at the surface by *total internal reflection* (TIR). There is no refracted beam and the beam is completely reflected. This process achieves 100% reflection, and is often used in prisms to achieve high reflectivity.

The incident angle that just satisfies this inequality is the *critical angle*:

$$\sin\theta_1 = \sin\theta_C = \frac{n_2}{n_1}$$

Incident rays above the critical angle undergo TIR; rays at angles less than the critical angle are refracted.

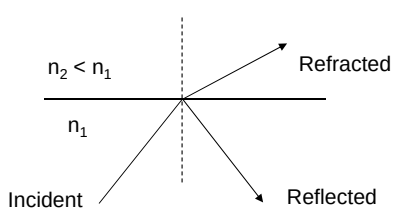
Critical angles for $n_2 = 1.0$

1.3	50.3°
1.4	45.6°
1.5	41.8°
1.6	38.7°
1.7	36.0°
1.8	33.7°
1.9	31.8°
2.0	30.0°

Partial Reflection

For common glasses (n = 1.5), the critical angle is about 42°. Even at angles below the critical angle there is a significant portion of the ray that is reflected. The ratio of reflection to transmission is governed by the Fresnel reflection coefficients, and the reflectance increases rapidly as the critical angle is approached.

Care must be taken in the design of optical systems with steep surface slopes; TIR instead of refraction may occur at glass/air interfaces.



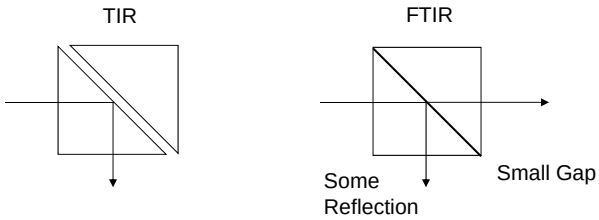
At normal incidence, with no absorption, the reflectance of an interface is

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

OPTI-502 Optical Design and Instrumentation I
© Copyright 2019 John E. Greivenkamp
College of Optical Sciences

Frustrated TIR (FTIR)

When a ray undergoes TIR, the electric field extends a short distance into the lower index medium (evanescent wave). If another high index surface is brought in close proximity to the TIR interface, a transmitted ray is produced. The transmitted intensity is a strong function of the gap size. The gap must be a fraction of a wavelength.



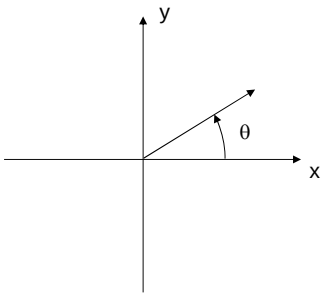
This phenomenon has practical importance in providing a switching mechanism from high to low transmission by pushing two prisms together.

OPTI-502 Optical Design and Instrumentation I
© Copyright 2019 John E. Greivenkamp
College of Optical Sciences

Cartesian Coordinates

Distances:

UP: Positive
DOWN: Negative
RIGHT: Positive
LEFT: Negative



Angles (measured from the +x axis):

COUNTER-CLOCKWISE: Positive
CLOCKWISE: Negative

The origin is the reference location for distances.

The x-axis is the reference direction for angles.

In a real-world measurement, the origin can be placed anywhere.

The reference direction for angles can be any direction.

Right-Hand Rule

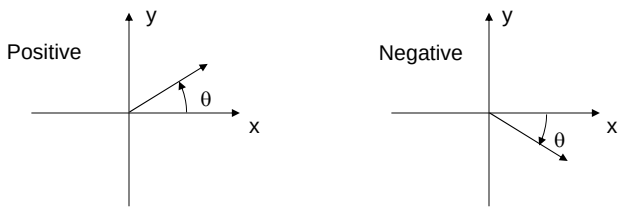
Rather than using clockwise and counterclockwise, it is common to use what is known as the "right-hand rule" for defining the signs of angles.

Place the fingers of your right hand along the reference direction.

Curl your fingers towards the desired direction.

If your thumb points up (or out of the plane of the paper) it is a positive angle.

If your thumb points down (or into the plane of the paper), it is a negative angle.



The right-hand rule is completely consistent with Cartesian measurements and the clockwise and counter-clockwise conventions.

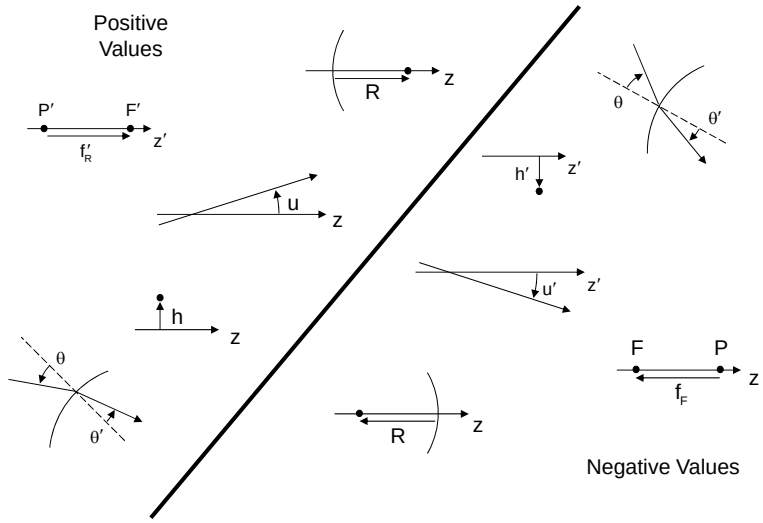
Reference Definitions and Sign Conventions

Throughout this course, a set of fully-consistent reference definitions and sign conventions is utilized. This allows the signs of results and variables to be easily related to the diagram or to the physical system.

- The axis of symmetry of a rotationally symmetric optical system is the optical axis and is the z-axis.
- All distances are measured relative to a reference point, line, or plane in a Cartesian sense: directed distances above or to the right are positive; below or to the left are negative.
- All angles are measured relative to a reference line or plane in a Cartesian sense (using the right-hand rule): counter-clockwise angles are positive; clockwise angles are negative.
- The radius of curvature of a surface is defined to be the directed distance from its vertex to its center of curvature.
- Light travels from left to right (from -z to +z) in a medium with a positive index of refraction.
- The signs of all indices of refraction following a reflection are reversed.

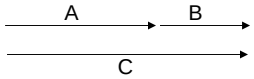
To aid in the use of these conventions, all directed distances and angles are identified by arrows with the tail of the arrow at the reference point, line, or plane.

Sign Conventions



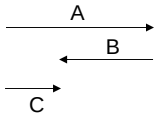
The value and sign for an angle or position depends on the reference. For example, a ray angle can be measured relative to either the optical axis or a surface normal.

Use of Sign Conventions



$$C = A + B$$

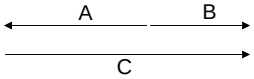
C is defined from the tail of A to the head of B



If the direction of B changes to negative, the same equation holds:

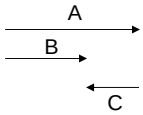
$$C = A + B \qquad B < 0$$

Redefine the reference for A: C is now defined from the head of A to the head of B



$$C = -A + B \qquad A < 0$$

If A changes direction to negative, the same equation holds:

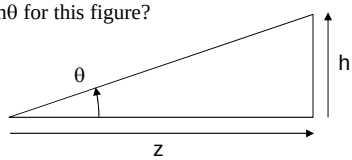


$$C = -A + B \qquad A > 0$$
$$C < 0$$

Set the equation up for the figure as drawn. When the signs of quantities change, the equations are still valid and the correct answer is obtained.

Use of Sign Conventions for Angles

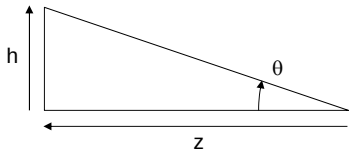
What is tanθ for this figure?



$\theta > 0 \qquad h > 0 \qquad z > 0$

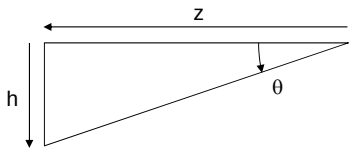
$$\tan\theta = \frac{h}{z}$$

What about these variations?



$\theta < 0 \qquad h > 0 \qquad z < 0$

$$\tan\theta = \frac{h}{z} < 0$$

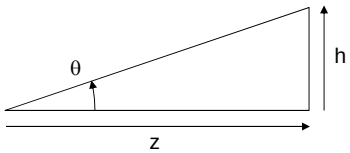


$\theta > 0 \qquad h < 0 \qquad z < 0$

$$\tan\theta = \frac{h}{z} > 0$$

In all of these Figures, z, h and θ are defined from the same reference locations (apex of the angle; base of the triangle), and the same equation holds independent of the signs of the quantities.

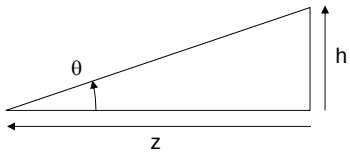
Choice of Reference



$\theta > 0 \qquad h > 0 \qquad z > 0$

$$\tan\theta = \frac{h}{z}$$

If the quantity z is redefined to have its reference changed to the opposite side of the triangle:



$\theta > 0 \qquad h > 0 \qquad z < 0$

$$\tan\theta = \frac{h}{-z} = -\frac{h}{z}$$

The choice of the reference location changes the required equation. It is essential to know the reference location for all quantities.

OPTI-502 Optical Design and Instrumentation I
© Copyright 2019 John E. Greivenkamp
College of Optical Sciences