The Abdus Salam International Centre for Theoretical Physics

United Nations Educational, Scientific and Cultural Organization

IAEA International Atomic Energy Agency

2130-3

Preparatory School to the Winter College on Optics and Energy

1 - 5 February 2010

Radiometry and photometry

M. Alonso

University of Rochester
U.S.A.

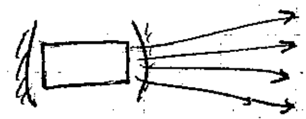
Radiometry (& Photometry)

1.) Ray Picture of coherence.

• Spatially Coherent sources:

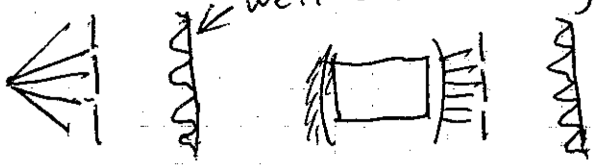
pt. source

Laser



To test coherence, we use Young's 2 pinhole experiment.

well-defined fringes

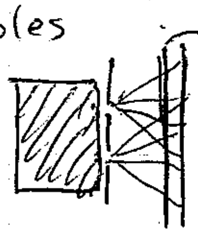


• Spatially incoherent sources

eg. thermal sources

2 pinholes

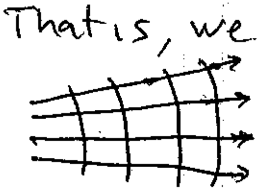
no fringes



This is because each point is a source that is statistically uncorrelated to all others.
Then, at each instant of time there might be fringes, but their position fluctuates so fast that they average out.

Rays and phase fronts

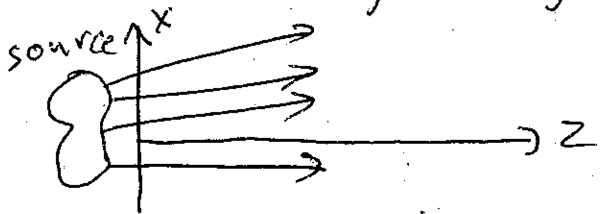
For a <u>coherent</u> source, we only use a 2-parameter (1 parameter in 2D) family of rays. That is, we can draw phase fronts normal to them.



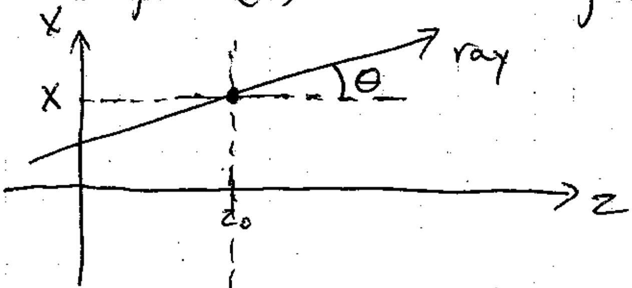
For an <u>incoherent</u> source, we use all 4 possible parameters (2 parameters in 2D), so phase fronts are not defined

Phase space

Let us first work in 2D. Let z coincide with the optical axis, and x to the transverse axis. Assume all light is going towards larger z:



At a given z, each ray can be identified by its height (x) and angle (θ):



Let us define p (sometimes called the optical momentum) as

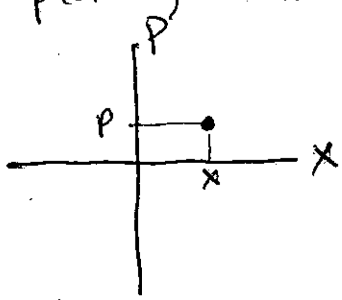
$$p = n \sin\theta$$

refractive index.

Then, the ray at z is fully characterized by x & p.

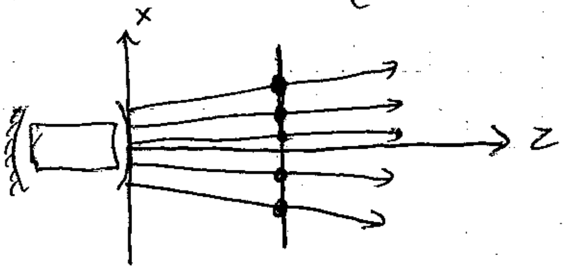
Let us define the vector $\underline{V} = \begin{pmatrix} x \\ p \end{pmatrix}$

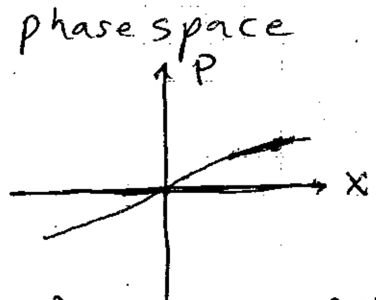
The ray can then be represented by a point in the x vs p plane, called "phase space".



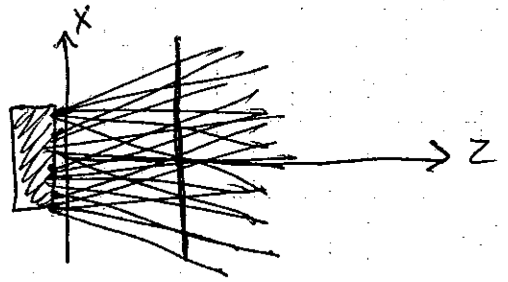
Of course, as the ray propagates in z, this point generally moves.

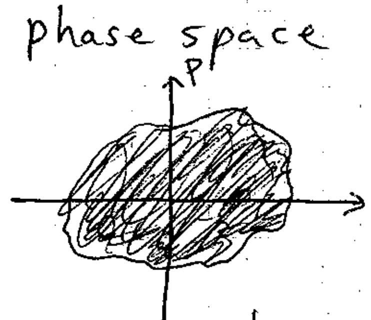
For a coherent source (in 2D), all rays form a 1-parameter family, so their points in phase space form a curve (can be straight!):





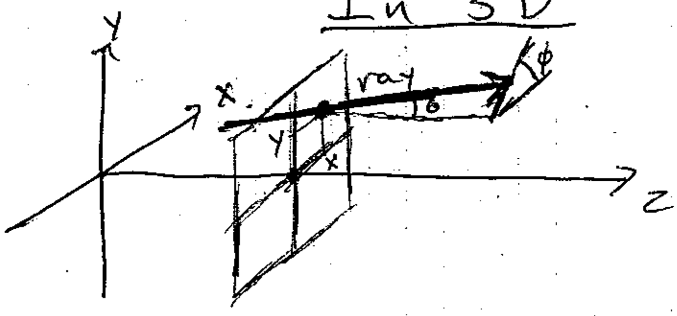
For an incoherent source (in 2D), the rays fill an area in phase space:





Propagation in free space or an optical system causes these curves and areas to change form.

In 3D



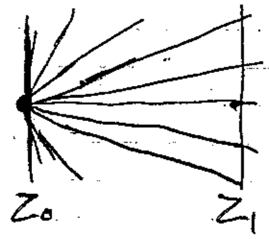
the transverse position is $\underline{x} = (x, y)$ instead of just x. The optical momentum is $\underline{p} = (p_x, p_y)$, where $p_x = n \sin\theta \cos\phi$, $p_y = n \sin\theta \sin\phi$.

Therefore $\underline{V} = \begin{pmatrix} x \\ y \\ p_x \\ p_y \end{pmatrix} = \begin{pmatrix} \underline{x} \\ \underline{p} \end{pmatrix}$, and phase space is four-dimensional! The rays of a coherent source form a surface (called the Lagrange manifold), while the rays of an incoherent source fill hypervolumes.

Exercise:

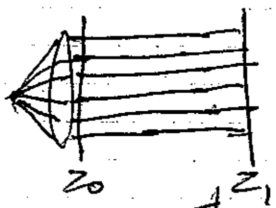
Represent in phasespace (2D) the rays generated by the following sources:

1) pointsource

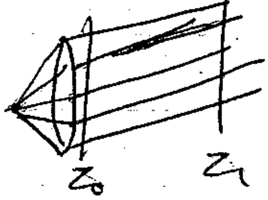


Draw it both for $z = z_0$, $z = z_1$

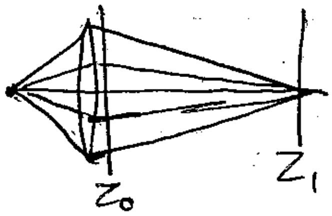
2) collimated beam



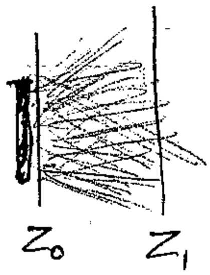
3) tilted collimated beam



4) focused beam



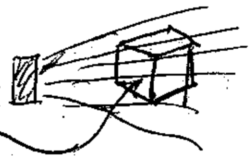
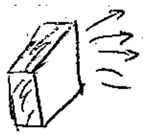
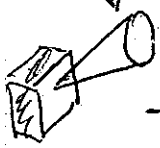
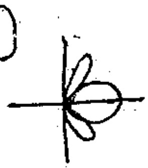
5) Extended incoherent source



Radiometry

- Phenomenological theory that describes the energy content in optical radiation fields, and how it flows through optical systems.
- Assumes spatially incoherent light.

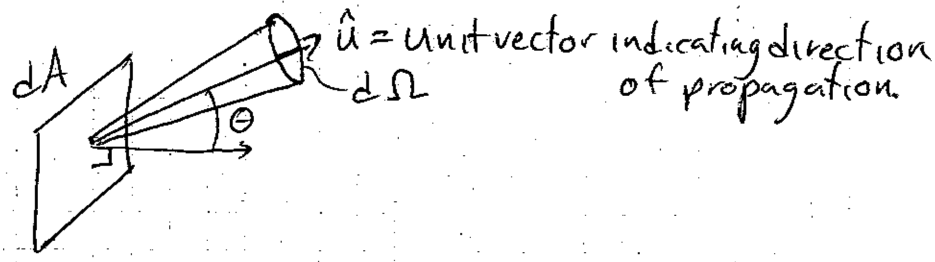
Quantities

Type		Name	Symbol	units
Total energy inside a volume		Radiant energy	Q	Joules
Energy per volume		Radiant energy density	$u = \frac{dQ}{dV}$	J/m^3
Energy per time		Radiant flux or Power	$\Phi = \frac{dQ}{dt}$	$\frac{J}{s} = W$
Power per area at source:		Radiant Exitance	$M = \frac{d\Phi}{dA}$	$\frac{W}{m^2}$
Power per area at detector/screen		Irradiance	$E = \frac{d\Phi}{dA}$	$\frac{W}{m^2}$
Power per solid angle	 	Radiant Intensity	$I = \frac{d\Phi}{d\Omega}$	$\frac{W}{sr}$
Power per solid angle per $\perp$ area		Radiance (brightness, specific intensity)	$L = \frac{d\Phi}{dA_\perp d\Omega}$	$\frac{W}{m^2 sr}$

The radiance

Although perhaps the less intuctive and difficult to measure, the radiance is in a way the most fundamental of all these quantities.

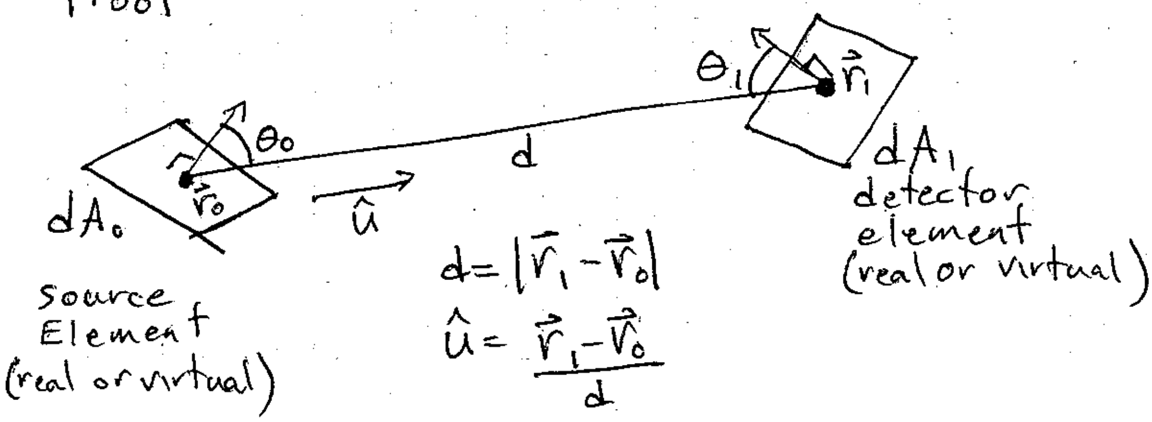
$$L(\vec{r},\hat{u}) = \frac{d\Phi}{dA_\perp d\Omega} = \frac{d\Phi}{d\vec{A}\cdot\hat{u}\, d\Omega} = \frac{d\Phi}{dA\cos\theta\, d\Omega}$$



$L(\vec{r},\hat{u})$ = amount of flux traversing an area dA and traveling in directions within a cone of solid angle $d\Omega$ centered at the direction of $\hat{u}$, per transverse area per unit angle.

- In free space / transparent homogeneous media, the radiance L is conserved along all points $\vec{r}$ in a line of direction $\hat{u}$. In other words, it is constant along <u>rays</u>.

Proof



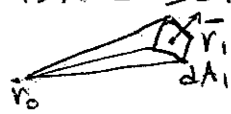
$d = |\vec{r}_1 - \vec{r}_0|$

$\hat{u} = \frac{\vec{r}_1 - \vec{r}_0}{d}$

At the source, the total power leaving dA_0 traveling towards dA_1 can be written as

$$d\Phi_0 = L(\vec{r}_0, \hat{u})\, dA_0 \cos\theta_0 \, d\Omega_1$$

where $d\Omega_1$ is the solid angle subtended by dA_1 from $\vec{r}_0$

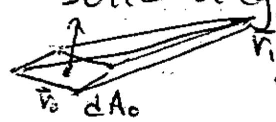


$d\Omega_1 = \dfrac{dA_1 \cos\theta_1}{d}$

At the detector, the total power arriving from dA_0 can be written as

$$d\Phi_1 = L(\vec{r}_1, \hat{u})\, dA_1 \cos\theta_1 \, d\Omega_0$$

where $d\Omega_0$ is the solid angle subtended by dA_0 from $\vec{r}_1$



$d\Omega_0 = \dfrac{dA_0 \cos\theta_0}{d}$

Since both powers must be the same, we get

$$d\Phi_0 = d\Phi_1$$

$$L(\vec{r}_0, \hat{u})\, dA_0 \cos\theta_0 \, d\Omega_1 = L(\vec{r}_1, \hat{u})\, dA_1 \cos\theta_1 \, d\Omega_0$$

using $d\Omega_0$ & $d\Omega_1$ from above:

$$L(\vec{r}_0, \hat{u})\, dA_0 \cos\theta_0 \frac{dA_1 \cos\theta_1}{d} = L(\vec{r}_1, \hat{u})\, dA_1 \cos\theta_1 \frac{dA_0 \cos\theta_0}{d}$$

$$\underline{L(\vec{r}_0, \hat{u}) = L(\vec{r}_1, \hat{u})} \quad \text{for } \vec{r}_1 - \vec{r}_0 \parallel \hat{u}.$$

Can also write as

$$\boxed{L(\vec{r}, \hat{u}) = L(\vec{r} + \delta\hat{u}, \hat{u}) \text{ for any } \delta} \quad \text{or}$$

$$\boxed{\hat{u} \cdot \nabla L(\vec{r}, \hat{u}) = 0}$$

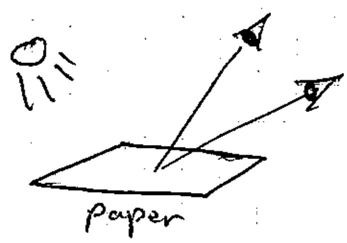
Lambertian sources/reflectors

$L(\vec{r}, \hat{u})$ independent of $\hat{u}$

For small (or very distant) sources:
radiant intensity $I(\hat{u}) = L\, dA \cos\theta$
Lambert's law.

Examples: Lambertian source → thermal sources

Lambertian reflector → matte paper



Looks equally bright in all directions.

- Radiant exitance for Lambertian sources:

$$M = \int L \cos\theta \, d\Omega = \int_0^{2\pi}\int_0^{\pi/2} L \cos\theta \sin\theta \, d\theta \, d\phi$$

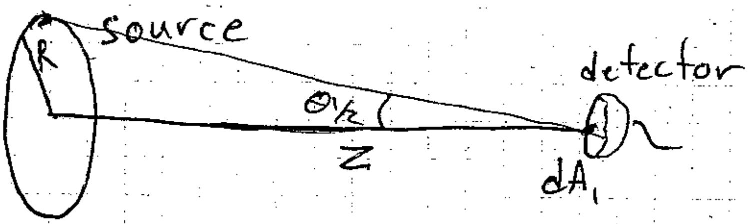
$$= 2\pi L \int_0^{\pi/2} \underbrace{\sin\theta}_{p} \underbrace{\cos\theta \, d\theta}_{dp} = 2\pi L \int_0^1 p\, dp = \underline{\pi L}$$

- Exercise: can you think of non-Lambertian:

sources	reflectors ?
•	•
•	•
•	•

Exercises: Extended sources.

1. consider a Lambertian uniform source of radiance L_o Whose shape is a disk of radius R, and a detector of area dA_1 at a distance z from the source and coaxial to it. Their surfaces are parallel. Find $d\Phi$ and E at the detector.



Power: $d\Phi = \iint L \cos\theta \, dA_1 \, d\Omega \approx dA_1 L_o \int_{\Omega_{source}} \cos\theta \, d\Omega$

where Ω_{source} is the solid angle subtended by the source at the detector.

$$d\Phi = dA_1 L_o \int_0^{2\pi} \int_0^{\theta_{1/2}} \cos\theta \sin\theta \, d\theta \, d\phi = dA_1 L_o 2\pi \int_0^{\sin\theta_{1/2}} \rho \, d\rho$$

use $\rho = \sin\theta$
$d\rho = \cos\theta \, d\theta$

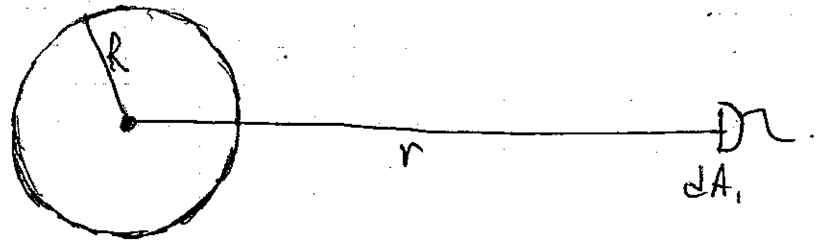
$$= dA_1 L_o 2\pi \left.\frac{\rho^2}{2}\right|_0^{\sin\theta_{1/2}} = \pi L_o dA_1 \sin^2\theta_{1/2}$$

$$= \pi L_o dA_1 \left(\frac{R^2}{R^2+Z^2}\right)$$

similarly

$$E = \frac{d\Phi}{dA_1} = \pi L_o \sin^2\theta_{1/2} = \pi L_o \frac{R^2}{R^2+Z^2}$$

2. Now consider a uniform Lambertian spherical source with radiance L_0, whose radius is R. The center of this source is at a distance r from an infinitesimal detector of area dA_1.



To solve this problem we use a trick, based on the spherical symmetry of the source. Because the source is Lambertian, the radiant exitance is

$$M = \pi L_0$$

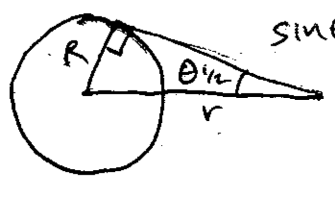
The irradiance E at any point must only depend on the distance r and not on direction. Further, since the power gets spread over larger areas as r increases, E must vary as r^{-2}: $E \propto \frac{1}{r^2}$.

But at the source's surface $E = M = \pi L_0$, so

$$\underline{E = \pi L_0 \frac{R^2}{r^2}}$$

Therefore, the power reaching the detector is $\underline{d\Phi = E\,dA_1 = \pi dA_1 L_0 \frac{R^2}{r^2}}$

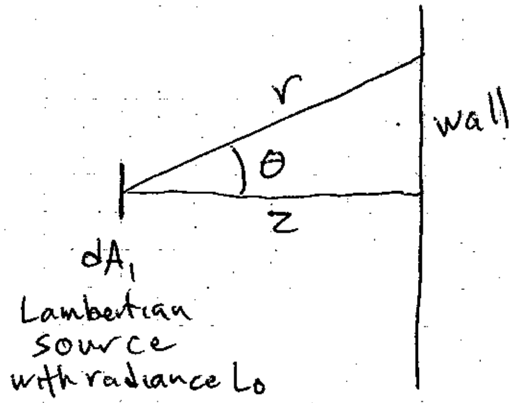
Notice



$\sin\theta_{1/2} = \frac{R}{r}$, so $\underline{E = \pi L_0 \sin^2\theta_{1/2},\ d\Phi = \pi L_0 dA_1 \sin^2\theta_{1/2}}$

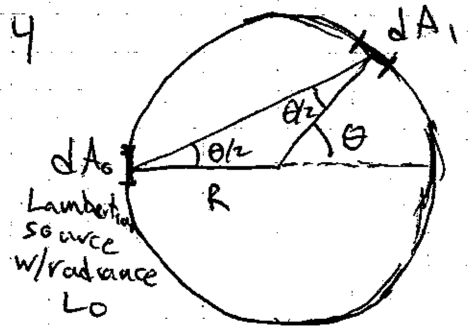
which coincides with previous example.
Why?

3.



Find E as a function of θ and z.

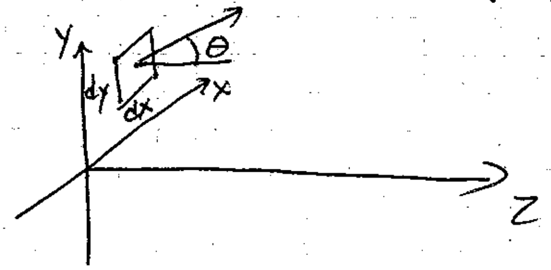
4



Find the flux going from dA_0 to dA_1 as a function of R and θ.

Connection to phase space

- Assume dA_0 is in x-y plane

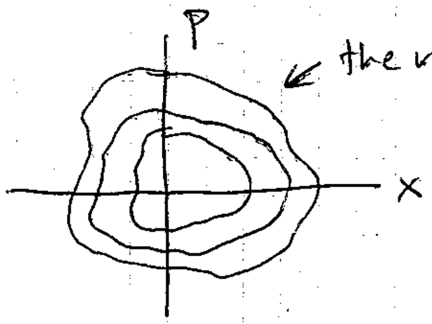


$$L = \frac{d\Phi}{dA\cos\theta\, d\Omega}, \quad \text{but } dA = dx\,dy,\ d\Omega = \sin\theta\, d\theta\, d\phi$$

so
$$\begin{aligned} dA\cos\theta\, d\Omega &= dx\,dy \sin\theta\cos\theta\, d\theta\, d\phi \\ &= dx\,dy \sin\theta\, d(\sin\theta)\, d\phi \\ &= \frac{dx\,dy}{n^2} \underbrace{|p|\,d|p|\,d\phi}_{\text{polars}} \\ &= \frac{dx\,dy}{n^2} dp_x\, dp_y \end{aligned}$$

Therefore, if $n=1$ (free space):

$$L = \frac{d\Phi}{dx\,dy\,dp_x\,dp_y} = \text{power per unit phase space volume}$$



the radiance is the "weight" of a ray.

If $n \neq 1$

$$\frac{L}{n^2} = \frac{d\Phi}{dx\,dy\,dp_x\,dp_y}$$

Ètendue

$$d\mathcal{E} = dA_\perp d\Omega = dx\,dy\,dp_x\,dp_y$$

$$\mathcal{E} = \iint dA_\perp d\Omega = \iint dx\,dy\,dp_x\,dp_y$$

phasespace volume.

For uniform Lambertian sources (freespace)

$$\Phi = L\mathcal{E}.$$

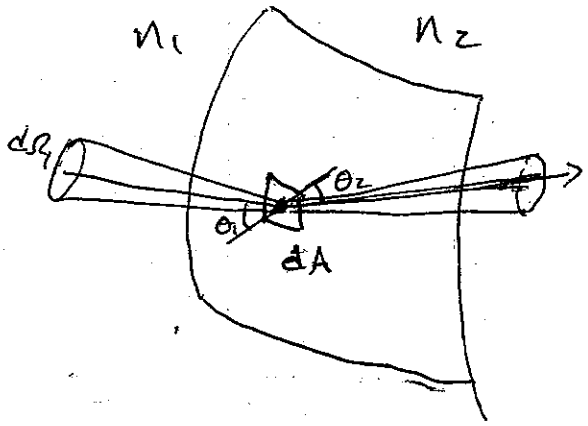
In a medium

$$\Phi = \frac{L}{n^2}\mathcal{E}, \text{ where } \mathcal{E} = n^2 \iint dA_\perp d\Omega$$

= phasespace volume

Radiance theorem at interfaces

Let us now see what happens to the Étendue and the radiance at an interface between two media with refractive indices n_1, n_2:



$$d\Omega_{\substack{1\\2}} = \sin\theta_{\substack{1\\2}}\, d\theta_{\substack{1\\2}}\, d\phi_{\substack{1\\2}}$$

But because the incident, transmitted rays and the surface normal are coplanar:

$$\phi_1 = \phi_2$$

Also from Snell's law: $n_1 \sin\theta_1 = n_2 \sin\theta_2$.

Differenciating this $n_1 \cos\theta_1 d\theta_1 = n_2 \cos\theta_2 d\theta_2$

$$d\mathcal{E}_1 = n_1^2 dA \cos\theta_1\, d\Omega_1 = dA\, \underbrace{n_1 \sin\theta_1}_{= n_2 \sin\theta_2}\, \underbrace{n_1 \cos\theta_1 d\theta_1}_{= n_2 \cos\theta_2 d\theta_2}\, \underbrace{d\phi_1}_{= d\phi_2}$$

$$= n_2^2 dA \cos\theta_2\, d\Omega_2 = \underline{d\mathcal{E}_2}$$

This means that the <u>étendue is conserved across interfaces</u>. Since it is also conserved on propagation through transparent media, then the <u>etendue is conserved along optical systems</u> (An exception is turbid media)

What about the radiance?

Let us neglect first reflections at the interface. then the power exiting the interface must equal that incident on it, that is

$$d\Phi_1 = d\Phi_2,$$

but $d\Phi = \frac{L\,d\mathcal{E}}{n^2}$, so

$$\frac{L_1 \cancel{d\mathcal{E}_1}}{n_1^2} = \frac{L_2 \cancel{d\mathcal{E}_2}}{n_2^2}$$

$$\boxed{\frac{L_1}{n_1^2} = \frac{L_2}{n_2^2}} \quad \text{or} \quad \frac{L_1(\vec{r}, \hat{u}_1)}{n_1^2} = \frac{L(\vec{r}, \hat{u}_2)}{n_2^2}$$

point at interface

related by Snell's law.

$\frac{L}{n^2}$ is called the "basic radiance".

If there are reflections

$$\frac{L_2(\vec{r}, \hat{u}_2)}{n_2^2} = \frac{L_1(\vec{r}, \hat{u}_1)}{n_1^2} \underbrace{T(\hat{u}_1)}_{\text{Fresnel transmission coefficient.}}$$

Spectral quantities

$$\Phi_\nu = \frac{d\Phi}{d\nu} \qquad \nu = \text{frequency}$$

$$\lambda = \frac{c}{\nu}, \quad \nu = \frac{c}{\lambda}$$

$$\Phi_\lambda = \frac{d\Phi}{d\lambda} \qquad \lambda = \text{wavelength}$$

$$\Phi = \int \Phi_\nu \, d\nu = \int \Phi_\lambda \, d\lambda$$

$$\Phi_\nu |d\nu| = \Phi_\lambda |d\lambda| \rightarrow \Phi_\nu = \Phi_\lambda \left|\frac{d\lambda}{d\nu}\right| = \Phi_\lambda \frac{c}{\nu^2}$$

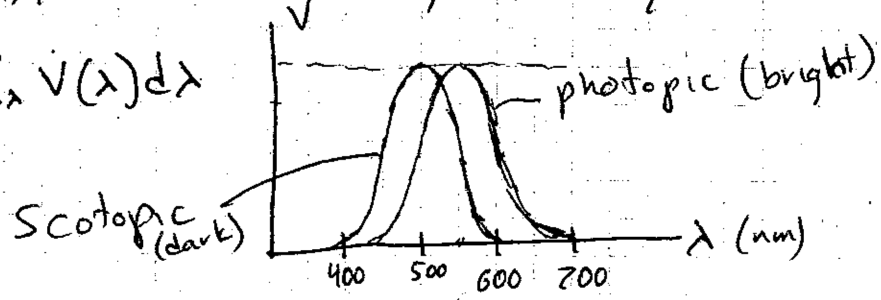
$$\therefore \underline{\nu \Phi_\nu = \lambda \Phi_\lambda}$$

Same for Q, u, E, M, I, L.

Photometry

Like radiometry, but as seen by the eye.

$$\underset{\substack{\uparrow \\ \text{visual}}}{\Phi_v} = \int \Phi_\lambda \, V(\lambda) \, d\lambda$$



Quantities

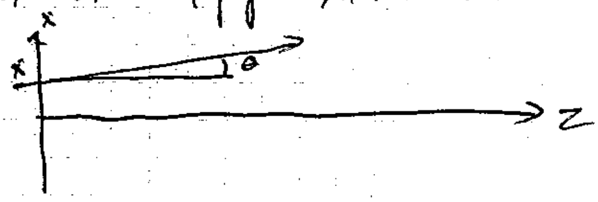
name	symbol	units
Luminous Energy	Q_v	talbot
Luminous density	u_v	talbot/m^3
Luminous flux	Φ_v	lumen (lm = talbot/s)
Illuminance	E_v	lux ($lx = lm/m^2$); foot-candle ($fc = lm/ft^2$); phot ($ph = lm/cm^2$)
Luminous exitance	M_v	(same as E_v)
Luminous intensity	I_v	candela ($cd = lm/rs$)
Luminance	L_v	($nt = lm/m^2 sr$)

Some luminance levels (in cd/cm^2)

sun (zenith)	$\sim 2 \times 10^5$
sun (horizon)	$\sim 6 \times 10^2$
Blue sky	$\sim .8$
Overcast sky	$\sim .2$
night sky	$\sim 5 \times 10^{-9}$
moon	$\sim .25$
least perceptible	$\sim 5 \times 10^{-11}$

Geometrical optics

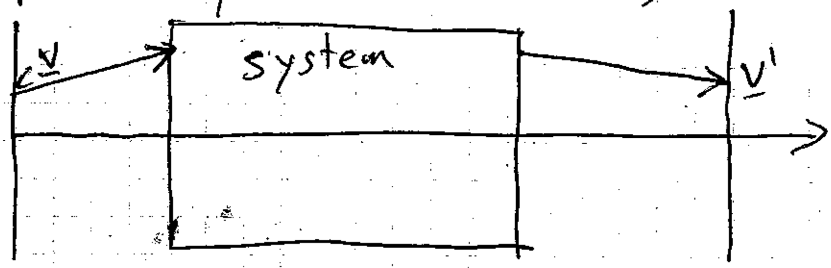
- Paraxial approximation



$\theta \ll \pi/2$, $\sin\theta \approx \theta \approx \tan\theta$

$\underline{p} \approx (\theta\cos\phi, \theta\sin\phi)$

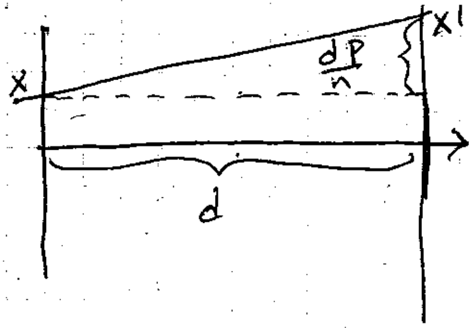
- Remember the ray vector $\underline{V} = \begin{pmatrix} \underline{x} \\ \underline{p} \end{pmatrix}$
- Optical systems (1^{st} order):



$\underline{V}' = \mathbb{M}\,\underline{V}$

matrix describing the system

Examples: free propagation in a homogeneous medium of index n:



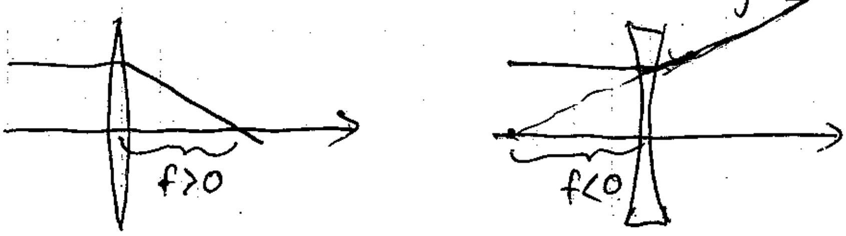
$\underline{x}' = \underline{x} + \frac{d}{n}\underline{p}$

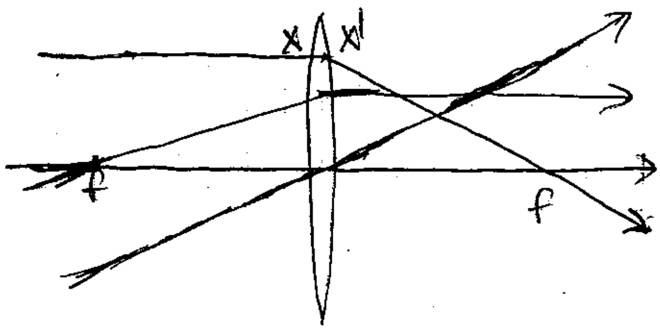
$\underline{p}' = \underline{p}$

so

$$\mathbb{M} = \mathbb{T}(d/n) = \begin{pmatrix} \mathbb{I} & d/n\,\mathbb{I} \\ \mathbb{O} & \mathbb{I} \end{pmatrix}$$

Thin lenses of focal length f



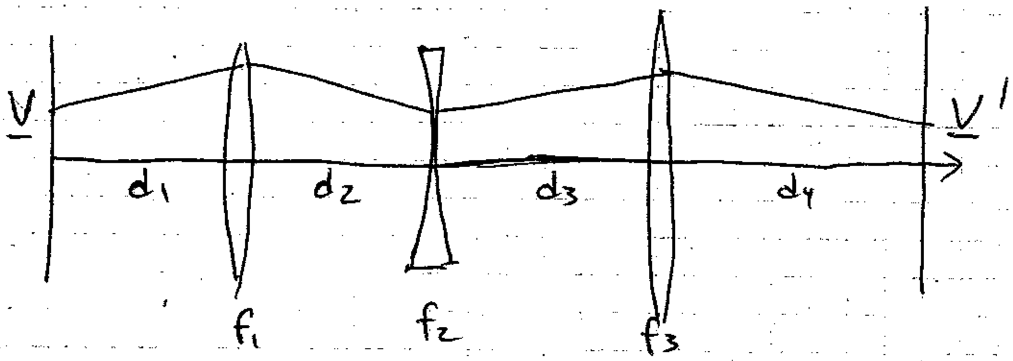


$\underline{x}' = \underline{x}$

$\underline{p}' = \underline{p} - \underline{x}/f$

$$\mathbb{M} = \mathbb{L}(f) = \begin{pmatrix} \mathbb{I} & \mathbb{O} \\ -\frac{\mathbb{I}}{f} & \mathbb{I} \end{pmatrix}$$

For a full system



$$\underline{V}' = \underbrace{\mathbb{T}(d_{N+1})\mathbb{L}(f_N)\ldots\mathbb{L}(f_2)\mathbb{T}(d_2)\mathbb{L}(f_1)\mathbb{T}(d_1)}_{\mathbb{M}}\underline{V}$$

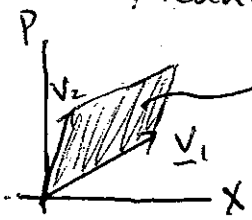
(ABCD systems because $\mathbb{M} = \begin{pmatrix} \mathbb{A} & \mathbb{B} \\ \mathbb{C} & \mathbb{D} \end{pmatrix}$)

Note: $\text{Det}\{\mathbb{T}(d)\} = \text{Det}\{\mathbb{L}(f)\} = 1$

therefore

$$\text{Det}\{\mathbb{M}\} = 1$$

Meaning? Think of 2D:



Think of the phase space area in a rhomboid defined by the rays $\underline{V}_1, \underline{V}_2$.

This area, which is an étendue in 2D, is given by

$$\mathcal{E} = |x_1 p_2 - x_2 p_1| = |\underline{V}_1^{+} \sigma \underline{V}_2|, \text{ where}$$

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

After propagation through the system,

$$\underline{V}_{\substack{1\\2}} \rightarrow \underline{V}'_{\substack{1\\2}} = \mathbb{M}\,\underline{V}_{\substack{1\\2}}, \text{ so}$$

$$\mathcal{E}' = |\underline{V}_1'^{+} \sigma \underline{V}_2| = |\underline{V}_1^{+} \mathbb{M}^{+} \sigma \mathbb{M} \underline{V}_2|$$

but $\mathbb{M}^{+}\sigma\mathbb{M} = \begin{pmatrix} m_{11} & m_{21} \\ m_{12} & m_{22} \end{pmatrix}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} = \begin{pmatrix} m_{11} & m_{21} \\ m_{12} & m_{22} \end{pmatrix}\begin{pmatrix} m_{21} & m_{22} \\ -m_{11} & -m_{12} \end{pmatrix}$

$$= \begin{pmatrix} m_{11}m_{21} - m_{21}m_{11}, & m_{11}m_{22} - m_{12}m_{21} \\ m_{12}m_{21} - m_{11}m_{22}, & m_{12}m_{22} - m_{22}m_{12} \end{pmatrix} = \text{Det}\{\mathbb{M}\}\,\sigma$$

therefore:

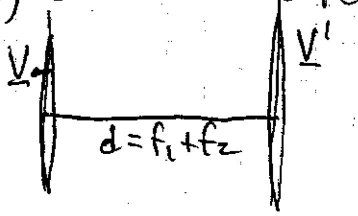
$$\mathcal{E}' = |\mathrm{Det}\{\mathbb{M}\}|\, |\underline{V}_1^{+} \sigma \underline{V}_2| = |\mathrm{Det}\{\mathbb{M}\}|\, \mathcal{E}$$

because $\mathrm{Det}\{\mathbb{M}\} = 1$, $\underline{\mathcal{E}' = \mathcal{E}}$

therefore $\mathrm{Det}\{\mathbb{M}\} = 1$ is analogous to Liouville's theorem.

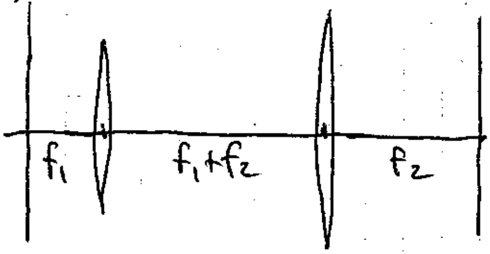
In 3D, if the system has rotational symmetry around the z axis, there are other invariants besides the étendue. More on this later...

Exercise: a) consider a telescope setup



find $\mathbb{M}$. What does the system do to both x & p?

b) Now surround this system with some free space as

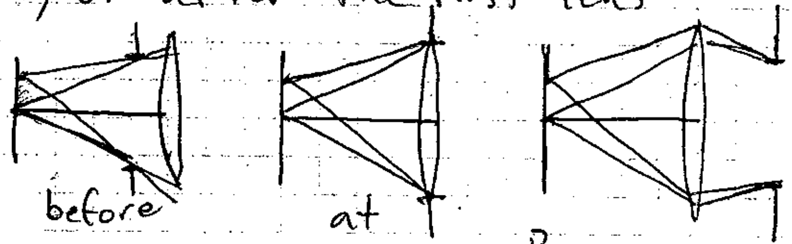


What is the new matrix? what does it do to the étendue occupied by a bundle of rays in phase space?

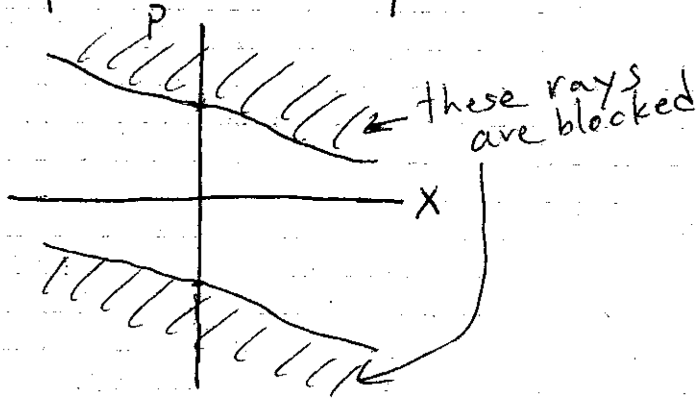
Stops and pupils

They limit the total étendue (or throughput) of the system.

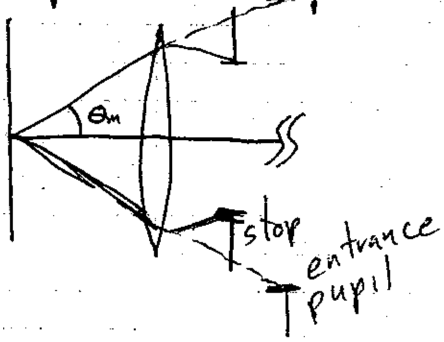
- Aperture stop: puts an angular limit to the rays entering the system. It can be before, at, or after the first lens



In phase space:
(at the plane of the object)



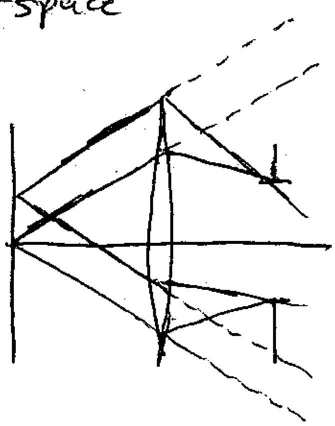
- Entrance pupil: image, viewed from the object, of the aperture stop



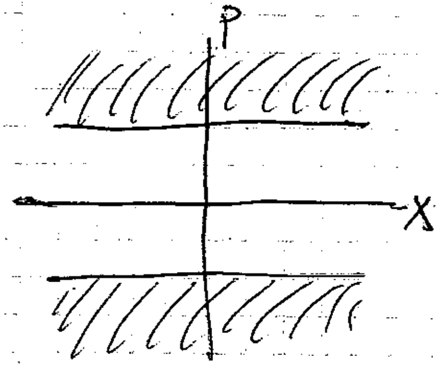
Note: if the stop is before or at the first lens, the stop itself is the entrance pupil.

$n \sin\theta_m = NA$
numerical aperture in object space

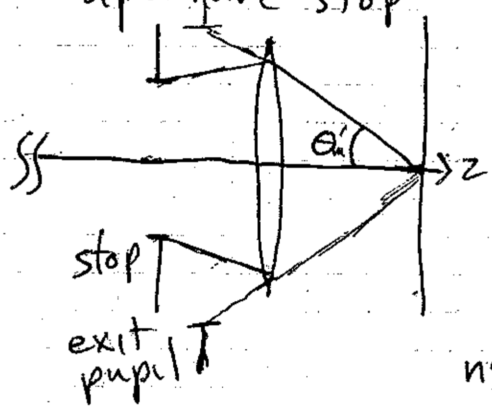
If the exit pupil is at infinity, e.g. if the stop is at the back focal plane of the 1st lens, the system is called "entrance telecentric".



For entrance telecentric systems, the angular limit is independent of object position



• <u>Exit pupil</u>: image, viewed from the image space, of the aperture stop

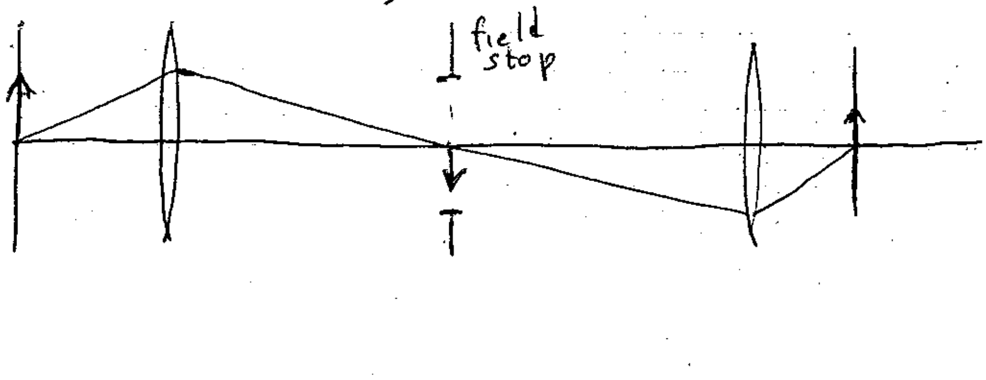


If the stop is at or after the last lens, the stop itself is the exit pupil

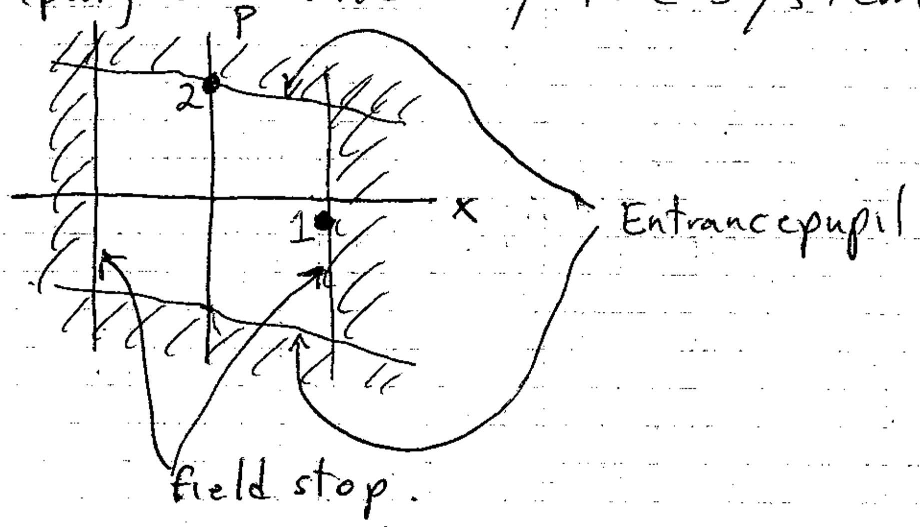
$n' \sin\theta_m' = NA'$, <u>numerical aperture</u> in image space

If the exit pupil is at infinity, then the system is called "exit telecentric". The image size is then insensitive to defocus.

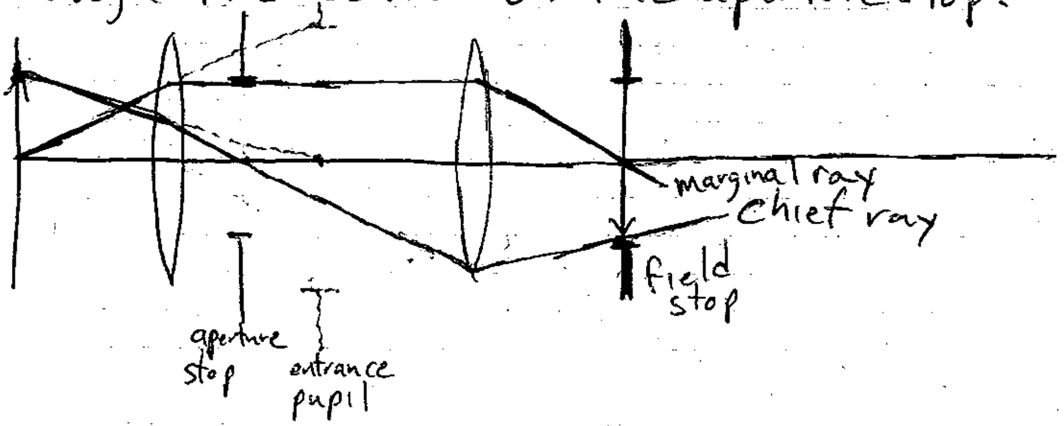
• <u>Field stop</u>: determines the size of the object that can be imaged. It can be at the object, the final image (e.g. at the film or ccd) or at an intermediate image:



The field and aperture stops (entrance pupil) determine the total étendue (also called throughput) admitted by the system.



1. <u>Chief ray</u>: A ray coming from the edge of the object (field stop) headed to the center of the entrance pupil, which then passes through the center of the aperture stop.



2. <u>Marginal ray</u>: A ray coming from the center of the object, headed for the edge of the exit pupil, and touching the edge of the aperture stop.

These are "extreme" rays in the system.

Lagrange invariant

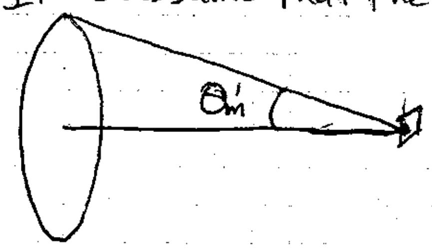
$$H = X_m P_c - X_c P_m = \underline{V}_c^{+} \underline{\sigma}\, \underline{V}_m$$

constant along the system.
In 2D this is the total $\mathcal{E}$.
In 3D, the total $\mathcal{E}$ is given by

$$\mathcal{E} = \pi^2 H^2$$

Irradiance of an image

Suppose we are imaging a Lambertian object of radiance L_o.
If we assume that the transmissivity of the medium is T, then the radiance at the exit pupil is TL_o.



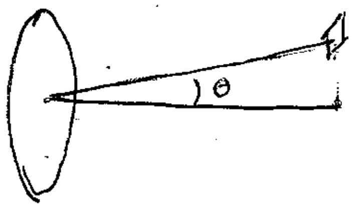
The irradiance at the image plane, on axis, is given by (see Problem 1)

$$E_{image} = \pi L_o T \sin^2\theta_m' \quad \text{(assuming uniform } T\text{)}.$$

Since $NA' = n \sin\theta_m'$,

$$E_{image} = \pi T \frac{L_o}{n'^2} NA'^2$$

Off axis, we get a factor of $\cos^4\theta$.

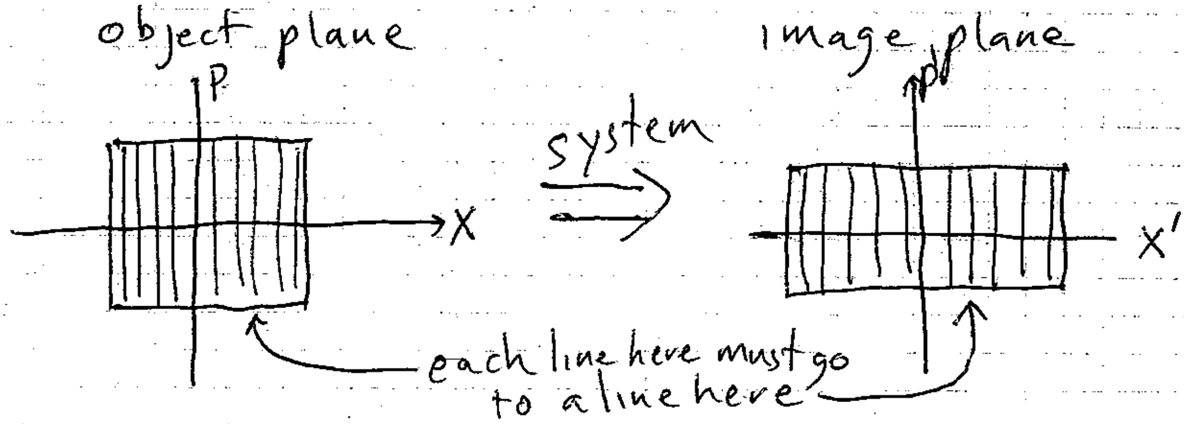


(see prob. 3)

Imaging vs. nonimaging (illumination) optics

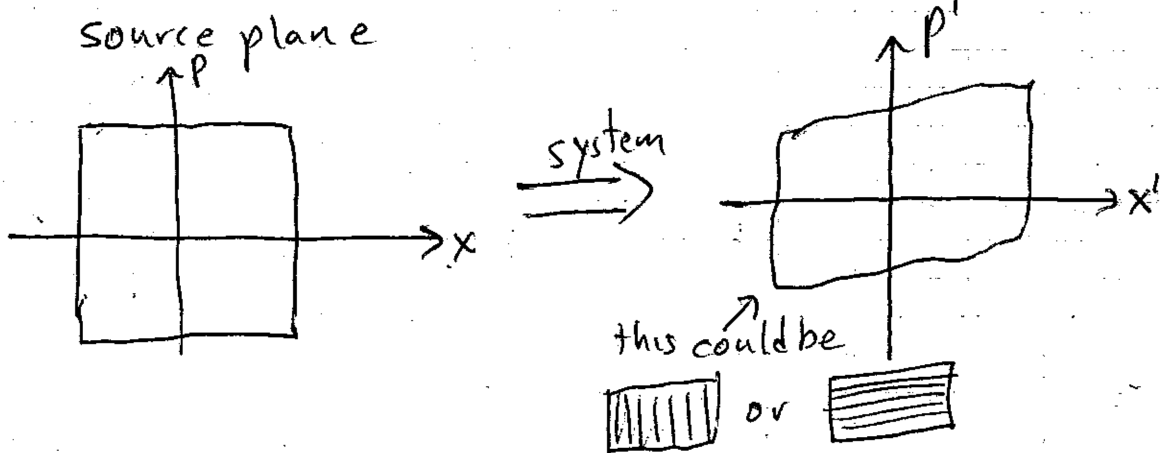
Imaging optics

The goal is to form good images of an object, i.e. to make all rays from the object go (at least approximately) to a corresponding point at the image plane. In phase space this means



Nonimaging optics

The goal is to take the light from a source and use it to illuminate an object in a given way. In phase space, this means



It is not important what ray goes to what point. The "imaging" solution is not forbidden, though.

Edge ray principle: it is enough to map the rays that define the edge of the region to solve the nonimaging problem.

3rd order theory (aberrations)

Let us assume that the system has axial symmetry. It is convenient to use, instead of $\underline{x}, \underline{f}$, the following normalized coordinates

$$\underline{h} = \frac{\underline{x}}{X_{max}} \quad \text{at the object plane}$$

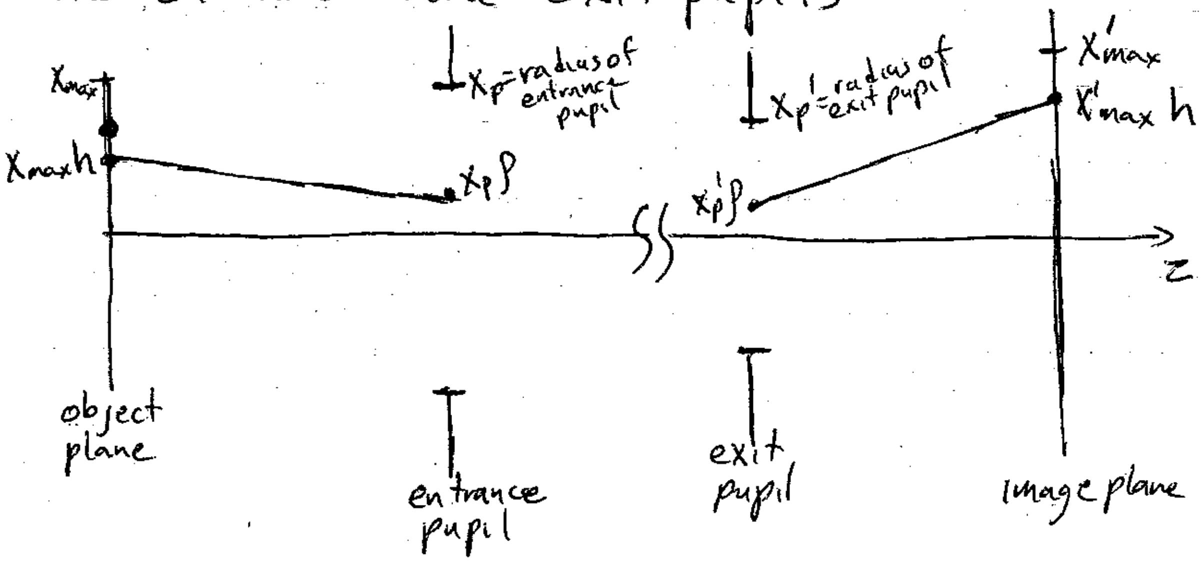
$$\underline{\rho} = \frac{\underline{x}}{X_{max}} \quad \text{at the aperture stop.}$$

Therefore, due to axial (rotational) symmetry, the three meaningful parameters are

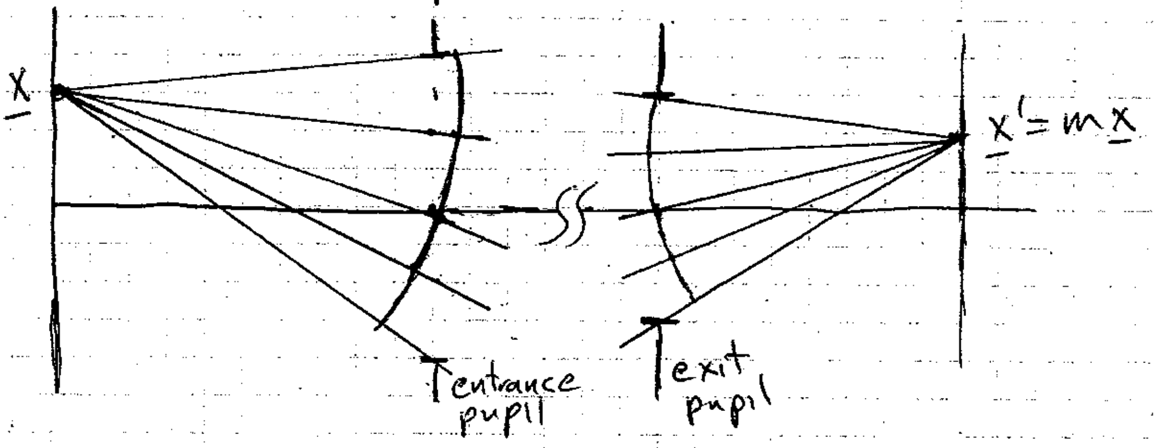
$$h = |\underline{h}|, \quad \rho = |\underline{\rho}|, \quad \phi = \arccos\left(\frac{\underline{h}\cdot\underline{\rho}}{h\rho}\right).$$

(Note that, for telecentric systems, $\underline{\rho} \propto \underline{f}$.)

We can think of an ideal system in terms of the entrance and exit pupils:



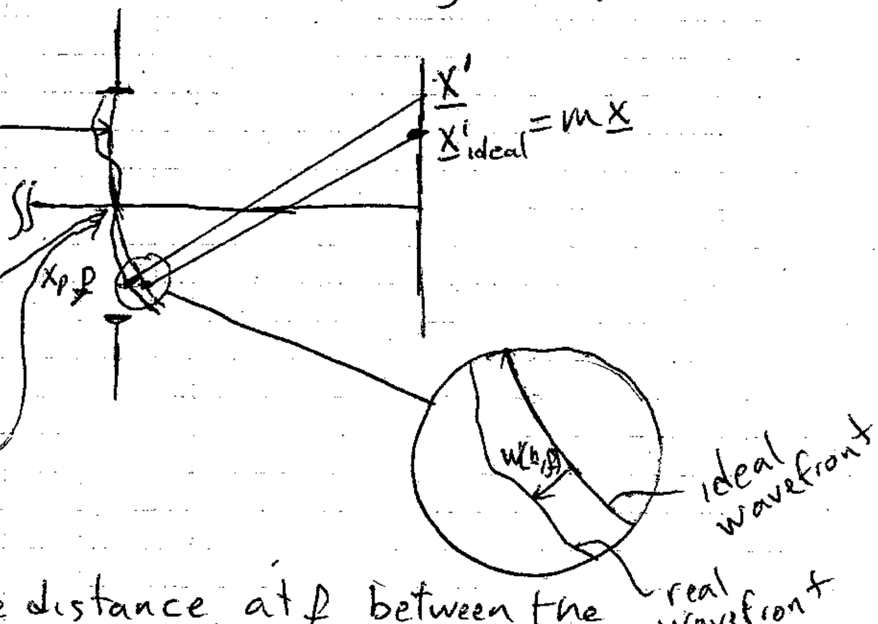
Ideally, all rays from an object point $\underline{x}$ go to the corresponding image point $\underline{x}' = m\underline{x}$ (where m is the magnification), so the phase fronts in both spaces are spherical:



However, in practice the phase front exiting the system is imperfect

We choose as the reference ideal wavefront the one that crosses the axis at the pupil plane.

We also choose the real one to do this



Let $W(\underline{h},\underline{\rho})$ be the distance at $\underline{\rho}$ between the real and the ideal wavefront (with $W>0$ if the real wavefront is to the right of the ideal one).

W is called the "wave aberration function".

The transverse error can be found:

$$\underline{\varepsilon}(\underline{h},\underline{\rho}) = \underline{x}' - \underline{x}'_{ideal} = -\frac{R}{n' x'_{max}} \frac{\partial W}{\partial \underline{\rho}}$$

refractive index

To study the system, we expand $W(\underline{h}, \underline{\rho})$ in a Taylor series in $\underline{h}$ and $\underline{\rho}$ around $\underline{0}$:

$$
\begin{aligned}
W(\underline{h},\underline{\rho}) = \; & W_{000} && \text{constant term} \\
& + W_{020}\,\rho^2 + W_{111}\,\underline{\rho}\cdot\underline{h} + W_{200}\,h^2 && \text{quadratic terms} \\
& \left.\begin{aligned} & + W_{040}\,\rho^4 + W_{131}\,\rho^2\,\underline{\rho}\cdot\underline{h} + W_{220}\,\rho^2 h^2 \\ & + W_{222}\,(\underline{\rho}\cdot\underline{h})^2 + W_{311}\,\underline{\rho}\cdot\underline{h}\,h^2 + W_{400}\,h^4 \end{aligned}\right\} && 4^{rth}\text{ order terms.} \; * \\
& + \ldots
\end{aligned}
$$

Notice that, when calculating $\underline{\varepsilon}$, because of the derivative, the order is reduced by one.
For this reason the quadratic terms describe "linear" behavior, and the 4^{rth} order ones (also called aberrations) lead to the "3^{rth} order theory".

Recall that we chose the reference and real wavefronts to cross at the axis ($\underline{\rho}=\underline{0}$), therefore $W(\underline{h},\underline{0})=0$, which implies $W_{000}=0$, $W_{200}=0$, $W_{400}=0$. Even if we did not set these constants to zero, they do not affect the image because these terms are independent of $\underline{\rho}$.

* Note that the notation for the coefficients W_{ijk} is such that:

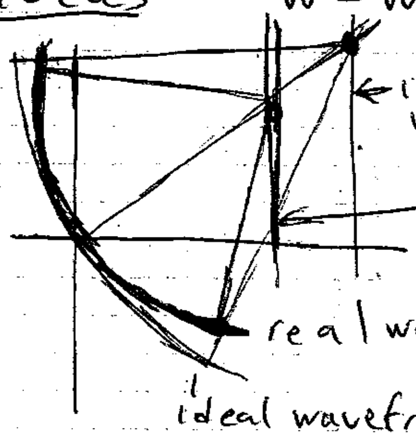
i = power of $|\underline{h}|$

j = power of $|\underline{\rho}|$

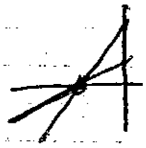
k = power of $\cos(\phi_h - \phi_\rho) = \dfrac{\underline{h}\cdot\underline{\rho}}{|\underline{h}||\underline{\rho}|}$.

The meaning of the aberrations is clarified if we first understand the two linear terms.

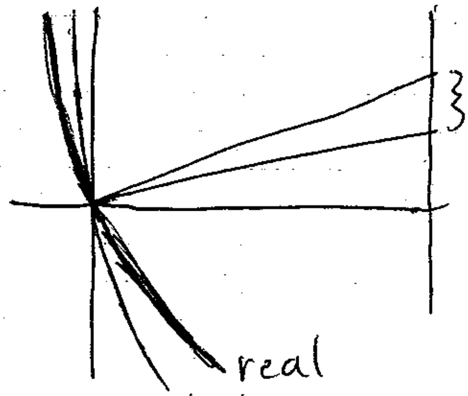
- <u>Defocus</u> $W = W_{020}\,\rho^2$ ← this changes the curvature so the image plane is closer or further away.



$$\underline{\varepsilon} = \frac{-2R\,W_{020}\,\underline{\rho}}{n' X'_{max}}$$



- <u>Magnification error</u> $W = W_{111}\,\underline{\rho}\cdot\underline{h}$ ← this is linear, so it tilts the wavefront



$$\underline{\varepsilon} = -\frac{R}{n' X'_{max}}\, W_{111}\, \underline{h}$$

$$\underline{X}' = m\,\underline{X} + \underline{\varepsilon} = \underbrace{\left(m - \frac{R W_{111}}{n' X'_{max}}\right)}_{\text{correction in magnification}} \underline{X}$$

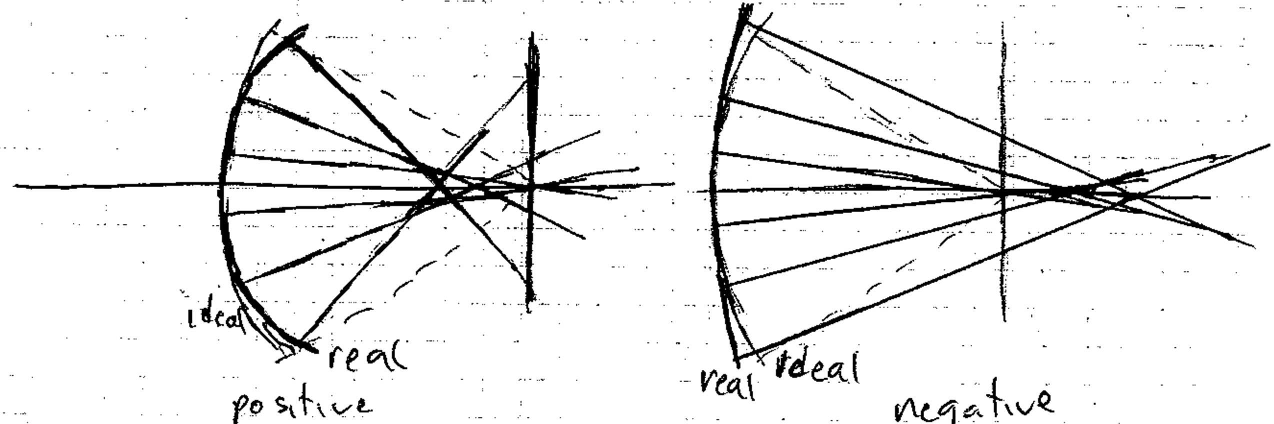
Aberrations

- <u>Spherical</u> $W = W_{040}\rho^4$

Only one that is independent of $\underline{h}$, therefore only one that degrades the image on-axis; can think of it as ρ-dependent defocus:

$$W = \underbrace{(W_{040}\rho^2)}_{\text{like } W_{020}}\rho^2$$

so rays at the edge of the pupil are more defocused than those at the center:



- Coma $\quad W = W_{131}\, \underline{h}\cdot\underline{\rho}\, \rho^2$

This is the only aberration linear in $\underline{h}$. If a system has no spherical aberration ($W_{040}=0$) or coma ($W_{131}=0$), then the image quality is good at and near the axis, since all other aberrations go as h^2 or h^3. Such a system is called "aplanatic".

Notice that coma looks both like a defocus and a magnification error:

$$W = \underbrace{\left(W_{131}(\underline{h}\cdot\underline{\rho})\right)}_{\text{like } W_{020}} \rho^2 = \underbrace{\left(W_{131}\,\rho^2\right)}_{\text{like } W_{111}} \underline{h}\cdot\underline{\rho}$$

Transverse error:

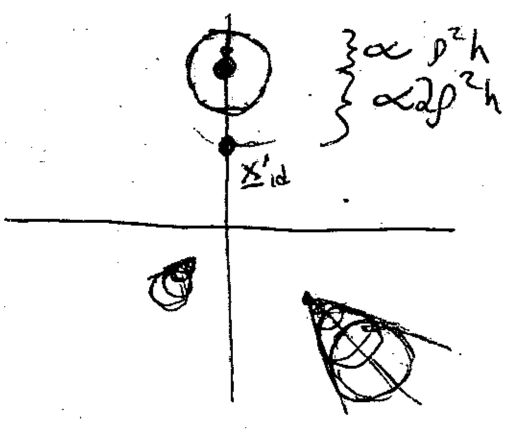
$$\underline{\varepsilon} = \frac{-R}{n' X'_{max}} \frac{\partial W}{\partial \underline{\rho}} = \frac{-2R W_{131}(\underline{h}\cdot\underline{\rho})}{n' X'_{max}}\, \underline{\rho} - \frac{R}{n' X'_{max}} W_{131}\, \rho^2\, \underline{h}$$

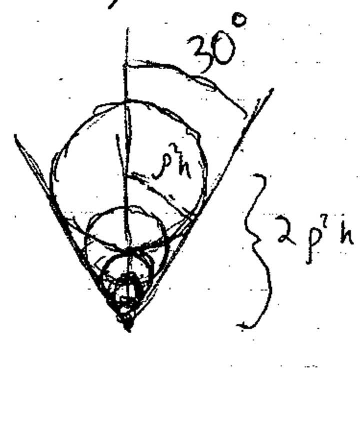
Consider an object point at the y axis: $\underline{h} = h(0,1)$

Let $\underline{\rho} = \rho(\cos\phi, \sin\phi)$. Then $\underline{h}\cdot\underline{\rho} = h\rho\sin\phi$ and

$$\underline{\varepsilon} = \frac{-R}{n' X'_{max}} W_{131}\, \rho^2 h \left[(2\sin\phi\cos\phi, 2\sin^2\phi) + (0,1)\right]$$

$$= \frac{-R}{n' X'_{max}} W_{131}\, \rho^2 h\, (\sin 2\phi,\ 2 - \cos 2\phi)$$

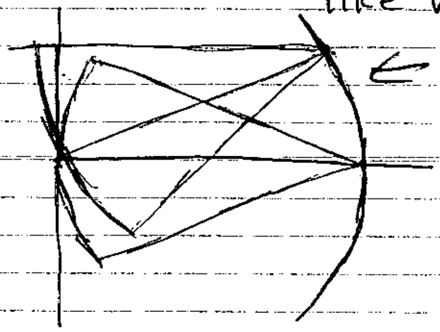




• Field curvature $W = W_{220}\, h^2 \rho^2$

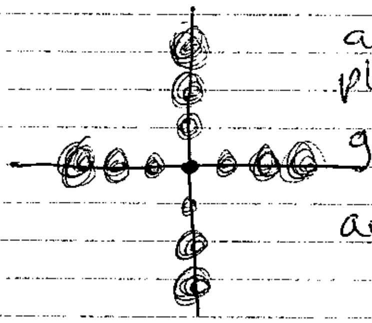
Like a defocus proportional to h^2:

$$W = \underbrace{(W_{220}\, h^2)}_{\text{like } W_{020}} \rho^2$$



Makes the ideal image plane be not a "plane" but a curved surface. Can be >0 or <0.

$$\underline{\varepsilon} = -\frac{2R}{n' x'_{max}} W_{220}\, h^2 \underline{\rho}$$



at ideal image plane, images get out of focus away from axis.

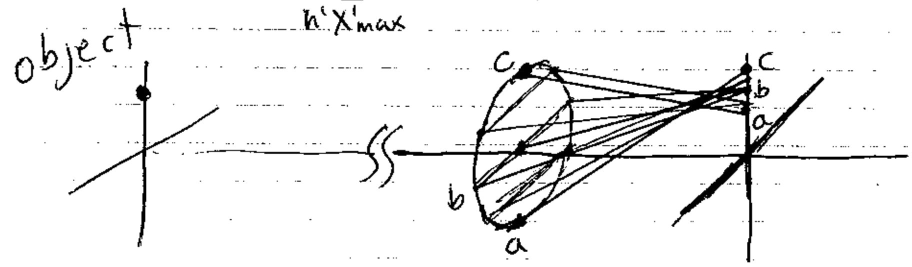
• Astigmatism $W = W_{222} (\underline{\rho} \cdot \underline{h})^2$

Like a magnification error proportional to $\underline{h} \cdot \underline{\rho}$:

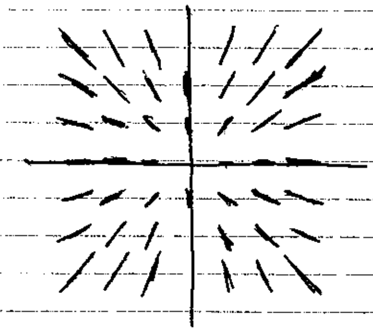
$$W = \underbrace{(W_{222}\, \underline{h} \cdot \underline{\rho})}_{\text{like } W_{111}}\, \underline{h} \cdot \underline{\rho}$$

Since this "change in magnification" is $\propto \underline{h} \cdot \underline{\rho}$, at a given point $\underline{h}$, the intersection of the rays moves from the ideal image an amount proportional to the component of ρ in the direction of $\underline{h}$

$$\underline{\varepsilon} = -\frac{2R}{n' x'_{max}} W_{222}\, \underline{h} \cdot \underline{\rho}\, \underline{h}$$

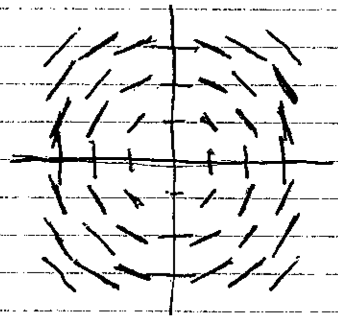


the image at the ideal plane then looks like

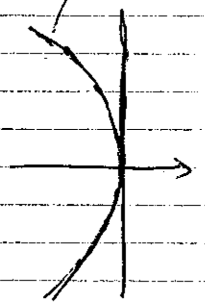


this plane is called the "sagittal image surface".

at a curved surface, the intersections are perpendicular

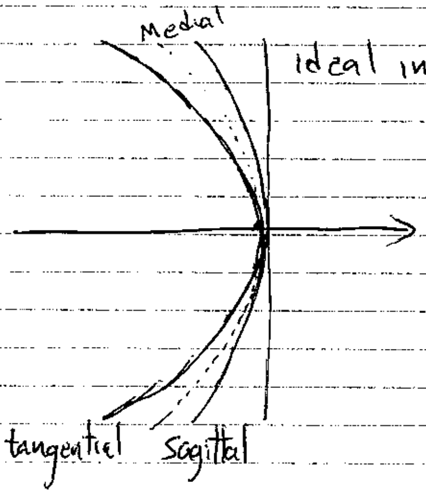


shape of this plane



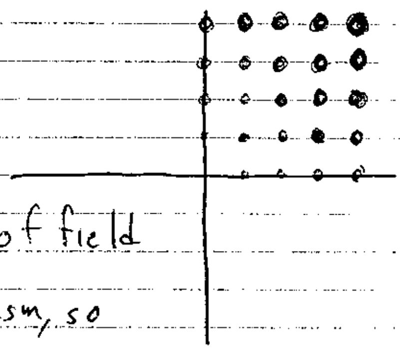
this is called the "tangential image surface".

If we combine field curvature and astigmatism the sagittal image surface gets curved, so we have:



ideal image plane

Right between the sagittal and tangential surfaces, at the so-called "Medial image surface", the spread of each point is smaller and round:



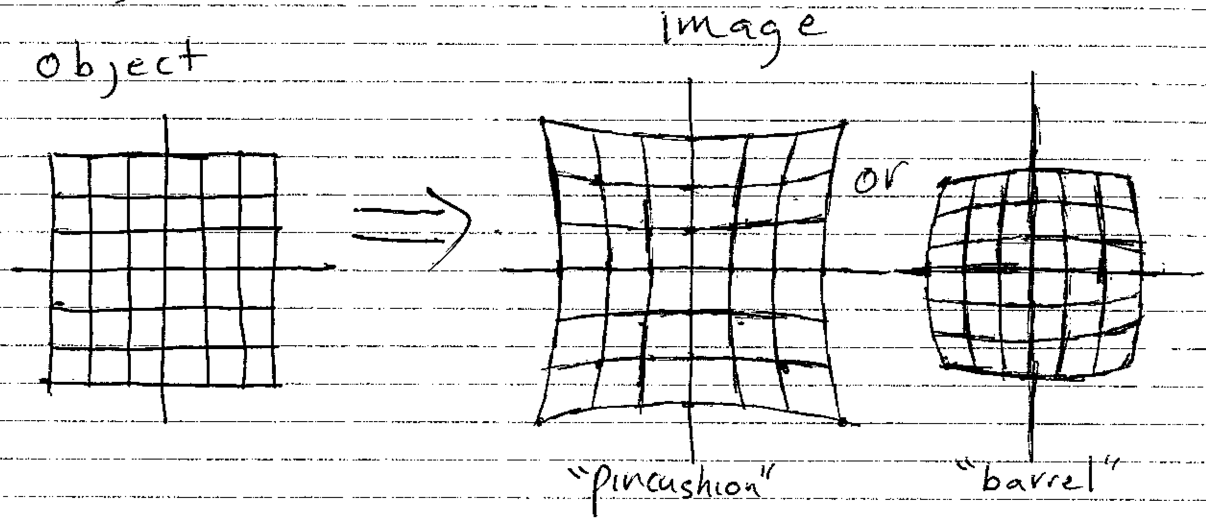
In some references, an amount of field curvature is grouped with astigmatism, so that, on its own, the medial image surface coincides with the flat image plane. The remainder of the field curvature is called "Petzval curvature".

• **Distorsion** $W = W_{311} h^2 \underline{h} \cdot \underline{\rho}$

This looks like a magnification error proportional to h^2:

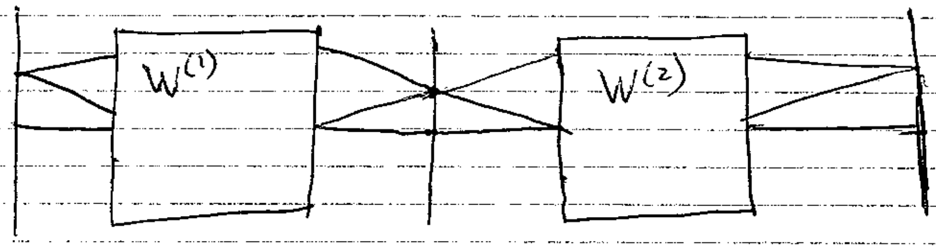
$$W = \underbrace{(W_{311} h^2)}_{\text{like } W_{111}} \underline{h} \cdot \underline{\rho}$$

That is, the magnification is different for object points away from the axis than for those near the axis:



Distorsion is the only aberration that does not affect the quality of the image at the image plane, but only its shape. Nowadays it can be corrected easily on the computer.

Concatenation of systems:



$W = W^{(1)} + W^{(2)}$, so each aberration is the sum of the aberrations.

Stop shifts

Shifting the stop changes some of the aberrations. Let us restrict ourselves to the case when the stop is also rescaled so that the output numerical aperture is preserved, i.e. so that the old and new exit pupil look like:

Notice that, then, the new pupil coordinate ρ' does not in general coincide with the old one, ρ, unless $\underline{h} = \underline{0}$. more generally: $\rho = \rho' + \varepsilon \underline{h}$ (see figure above). Find the new aberrations W^*_{ijk} in terms of the old ones W_{ijk} by substituting $\rho = \rho' + \varepsilon \underline{h}$ in $W = W_{040}\rho^4 + W_{131}\rho^2 \rho\cdot\underline{h} + W_{220}\rho^2 h^2$

$$+ W_{222}(\rho\cdot\underline{h})^2 + W_{311}\,\rho\cdot\underline{h}\,h^2.$$