6 Astronomical optics

Because astronomical sources are faint, we need to collect light. We use telescopes/cameras to make images of astronomical sources. Example: a 20th magnitude star gives $\sim 0.01\text{photons/s/cm}^2$ at 5000 A through a 1000 A filter! However, using a 4m telescope gives 1200 photons/s.

Telescopes/optics are the bread and butter tool of the observational astronomer, so it is worthwhile to be familiar with how they work.

6.1 Single surface optics and definitions

We will define an optical system as a system which collects light; usually, the system will also make images. This requires the bending of light rays, which is accomplished using lenses (refraction) and/or mirrors (reflection), using curved surfaces.

The operation of refractive optical systems is given by Snell’s law of refraction:

\[ n \sin i = n' \sin i' \]

where $n$ are the indices of refraction, $i$ are the angles of incidence, relative to the normal to the surface. For reflection:

\[ i' = -i \]

An optical element takes a source at $s$ and makes an image at $s'$. The source can be real or virtual. A real image exists at some point in space; a virtual image is formed where light rays apparently emanate from or converge to, but at a location where no light actually appears. For example, in a Cassegrain telescope, the image formed by the primary is virtual, because the secondary intercepts the light and redirects it before light gets to the focus of the primary.

Considering an azimuthally symmetric optic, we can define the optical axis to go through the center of the optic. The image made by the optic will not necessarily be a perfect image: rays at different height at the surface, $y$, might not cross at the same point. This is the subject of aberrations, which we will get into in a while. For a “smooth” surface, the amount of aberration will depend on how much the different rays differ in $y$, which depends on the shape of the surface. We define paraxial and marginal rays, as rays near the center of the aperture and those on the edge of the aperture. We define the chief ray as the ray that passes through the center of the aperture. To define nominal (unaberrated) quantities, we consider the paraxial regime, i.e. a small region near the optical axis, surrounding the chief ray. In this regime, all angles are small, aberrations vanish, and a surface can be wholly specified by its radius of curvature $R$.

The field angle gives the angle formed between the chief ray from an object and the z-axis. Note that paraxial does not necessarily mean a field angle of zero; one can have an object at a field angle and still consider the paraxial approximation.
Note also that for the time being, we are ignoring diffraction. But we’ll get back to that too. We are considering geometric optics, which is what you get from diffraction as wavelength tends to 0. For nonzero wavelength, geometric optics applies at scales \( x >> \lambda \).

We can derive the basic relation between object and image location as a function of a surface where the index of refraction changes (Schroeder, chapter 2).

\[
\frac{n'}{s'} - \frac{n}{s} = \frac{(n' - n)}{R}
\]

The points at \( s \) and \( s' \) are called conjugate; the behavior is independent of which direction the light is going. If either \( s \) or \( s' \) is at infinity (true for astronomical sources for \( s \)), the other distance is defined as the focal length, \( f \), of the optical element. For \( s = \infty \), \( f = s' \).

We can define the quantity on the right side of the equation, which depends only the the surface parameters (not the image or object locations), as the power, \( P \), of the surface:

\[
P \equiv \frac{(n' - n)}{R} = \frac{n'}{f'} = \frac{n}{f}
\]

We can make a similar derivation for the case of reflection:

\[
\frac{1}{s'} + \frac{1}{s} = \frac{2}{R}
\]

This shows that the focal length for a mirror is given by \( R/2 \).

Note that one can treat reflection by considering refraction with \( n' = -n \), and get the same result:

\[
\frac{n'}{s'} + \frac{n'}{s} = \frac{(n' + n')}{R}
\]

Given the focal length, we define the focal ratio to be the focal length divided by the aperture diameter. The focal ratio is also called the \( F \)-number and is denoted by the abbreviation \( f/ \). Note \( f/10 \) means a focal ratio of ten; \( f \) is not a variable in this! The focal ratio gives the beam “width”; systems with a small focal ratio have a short focal length compared with the diameter and hence the incoming beam to the image is wide. Systems with small focal ratios are called “fast” systems; systems with large focal ratios are called “slow” systems.

The magnification of a system gives the ratio of the image height to the object height:

\[
\frac{h'}{h} = \frac{(s' - R)}{(s - R)} = \frac{ns'}{n's}
\]
The magnification is negative for this case, because object is flipped. The magnification also negative for reflection: \( n' = -n \). Magnification is an important quantity for multi-element systems.

We define the scale as the motion of image for given incident angle of parallel beam from infinity. From a consideration of the chief rays for objects on-axis and at field angle \( \alpha \), we get:

\[
\tan \alpha \approx \alpha = \frac{x}{f}
\]

or

\[
\text{scale} \equiv \frac{\alpha}{x} = \frac{1}{f}
\]

In other words, the scale, in units of angular motion per physical motion in the focal plane, is given by \( 1/f \). For a fixed aperture diameter, systems with a small focal ratio (smaller focal length) have a larger scale, i.e. more light in a patch of fixed physical size: hence, these are "faster" systems.

Exercise: the APO 3.5m telescope is a f/10 system. A typical CCD might have 15 micron pixels. What angle in the sky would one pixel subtend? Once you get this, comment on whether you think this is a good pixel scale and why or why not?

Know the terminology: real/virtual images, paraxial/marginal/chief rays, field angle, focal ratio, magnification. Understand what is meant by the paraxial approximation. Know the basic lens/mirror equation. Know how to calculate the scale of an optical system.

6.2 Multi-surface systems

To combine surfaces, one just takes the image from the first surface as the source for the second surface, etc., for each surface. We can generally describe the basic parameters of multi-surface systems by equivalent single-surface parameters, e.g. you can define an effective focal length of a multi-surface system as the focal length of some equivalent single-surface system. The effective focal length is the focal length of the first element multiplied by the magnification of each subsequent element. The two systems (single and multi) are equivalent in the paraxial approximation ONLY.

6.2.1 a lens (has two surfaces)

Consider a lens in air \((n \sim 1)\). The first surface give

\[
\frac{n}{s'_1} - 1 = \frac{(n - 1)}{R_1} = P_1
\]
The second surface gives:
\[ \frac{1}{s'_2} - \frac{n}{s_2} = \frac{1 - n}{R_2} = P_2 \]

but we have \( s_2 = s'_1 - d \) (remember we have to use the plane of the second surface to measure distances for the second surface).

After some algebra, we find the effective focal length (from center of lens):
\[
P = \frac{1}{f'} = P_1 + P_2 - \frac{d}{n} P_1 P_2
\]

\[
P = \frac{(n - 1)}{R_1} + \frac{(1 - n)}{R_2} - \frac{d}{n} \frac{(n - 1)(1 - n)}{R_1 R_2}
\]

From this, we derive the thin lens formula:
\[
\frac{1}{f'} = \frac{1}{f_1} + \frac{1}{f_2} - \frac{(n - 1)(1 - n)}{R_1 R_2}
\]

6.2.2 plane-parallel plate

Zero power, but moves image laterally: \( \Delta = d[1 - (1/n)] \). Application to filters: variation of focus.

6.2.3 Two-mirror telescopes:

In astronomy, most telescopes are two-mirror telescopes of Newtonian, Cassegrain, or Gregorian design. All 3 types have a concave primary. The Newtonian has a flat secondary, the Cassegrain a convex secondary, and the Gregorian a concave secondary. The Cassegrain is the most common for research astronomy; it is more compact than a Gregorian and allows for magnification by the secondary. Basic parameters are outlined in Figure 12.

Each of these telescope types defines a family of telescopes with different first-order performances. From the usage/instrumentation point of view, important quantities are:

- the diameter of the primary, which defines the light collecting power
- the scale of the telescope, which is related to the focal length of the primary and the magnification of the secondary:
  \[ f_{eff} = f_1 m \]
  (alternatively, the focal ratio of the telescope, which gives the effective focal length with the diameter)
the back focal distance, which is the distance of the focal plane behind the telescope

From the design point of view, we need to specify:

- the radii of curvature of the mirrors
- the separation between the mirrors

The relation between the usage and design parameters can be derived from simple geometry. First, accept some basic definitions:

- ratio of focal lengths, $\rho$:
  \[ \rho = \frac{R_2}{R_1} = \frac{f_2}{f_1} \]

- magnification of the secondary, $m$ (beware that $s_2'$ is negative for a Cassegrain!):
  \[ m = -\frac{s_2'}{s_2} \]

- back focal distance, the distance from the primary vertex to the focal plane (often expressed in units of the primary focal length, or primary diameter):
  \[ f_1\beta = D\eta \]
• primary focal ratio, \( F_1 \):
   \[ F_1 = f_1 / D \]

• ratio of marginal ray heights, \( k \) (directly related to separation of mirrors):
   \[ k = y_2 / y_1 \]

Using some geometry, we can derive some basic relations between these quantities, in particular:

\[ \rho = \frac{mk}{m - 1} \]

and

\[ (1 + \beta) = k(m + 1) \]

Usually, \( f_1 \) is limited by technology/cost. Then choose \( m \) to match desired scale. \( k \) is related to separation of mirrors, and is a compromise between making telescope shorter and blocking out more light vs. longer and blocking less light; in either case, have to keep focal plane behind primary!

One final thing to note is how we focus a Cassegrain telescope. Most instruments are placed at a fixed location behind the primary. Ideally, this will be at the back focal distance, and everything should be set as designed. However, sometimes the instrument may not be exactly at the correct back focal distance, or it might move slightly because of thermal expansion/contraction. In this case, focusing is usually then done by moving the secondary mirror.

The amount of image motion for a given secondary motion is given by:

\[ \frac{d\beta}{dk} = \frac{d}{dk} k(m + 1) - 1 \]

Working through the relations above, this gives:

\[ \frac{d\beta}{dk} = m^2 + 1 \]

so the amount of focal plane motion \((f_1 d\beta)\) for a given amount of secondary motion \((f_1 dk)\) depends on the magnification of the system.

If you move the secondary you change \( k \). Since \( \rho \) is fixed by the mirror shapes, it’s also clear that you change the magnification as you move the secondary; this is expected since you are changing the system focal length, \( f = mf_1 \). So it’s possible that a given instrument could have a slightly varying scale if its position is not perfectly fixed relative
to the primary. Alternatively, if you need to independently focus and set the scale (e.g., SDSS!), then you need to be able to move two things!

Note that even if the instrument is at exactly the back focal distance, movement of the secondary is required to account for mechanical changing of spacing between the primary and secondary as a result of thermal expansion/contraction.

### 6.2.4 Definitions for multi-surface system: stops and pupils

- **aperture stop**: determines the amount of light reaching an image (usually the primary mirror)
- **field stop**: determines the angular size of the field. This is usually the detector, but for a large enough detector, it could be the secondary.
- **pupil**: location where rays from all field angles fill the same aperture.
- **entrance pupil**: image of aperture stop as seen from source object (usually the primary).
- **exit pupil**: image of aperture stop formed by all subsequent optical elements.

In a two-mirror telescope, the location of the exit pupil is where the image of the primary is formed by the secondary. This can be calculated using $s = d$ as the object distance (where $d$ is the separation of the mirrors), then with the reflection equation, we can solve for $s'$ which gives the location of the exit pupil relative to the secondary mirror. If one defines the quantity $\delta$, such that $f_1 \delta$ is the distance between the exit pupil and the focal plane, then (algebra not shown):

$$\delta = \frac{m^2k}{m + k - 1} = \frac{m^2(1 + \beta)}{m^2 + \beta}$$

This pupil is generally not accessible, so if one needs access to a pupil, additional optics are used.

The exit pupil is an important concept. When we discuss aberrations, it is the total wavefront error at the exit pupil which gives the system aberration. Pupils are important for aberration compensation. They can also be used to put light at a location that is independent of pointing errors.

*Understand in principle how you can calculate multi-surface systems as a sequence of individual surface. Know the different types of two-mirror telescopes (Cassegrain, Gregorian, Newtonian). Know the terminology: aperture stop, field stop, pupil.*
6.3 Aberrations

6.3.1 Surface requirements for unaberrated images

Next we consider non-paraxial rays. We first consider what surface is required to make an unaberrated image.

We can derive the surface using Fermat’s principle. Fermat’s principle states that light travels in the path such that infinitessimally small variations in the path doesn’t change the travel time to first order: \(\frac{d(\text{time})}{d(\text{length})}\) is a minimum. For a single surface, this reduces to the statement that light travels the path which takes the least time. An alternate way of stating Fermat’s principle is that the optical path length is unchanged to first order for a small change in path. The OPL is given by:

\[
\text{OPL} = \int c\,dt = \int \frac{c}{v}\,dt = \int nds
\]

Fermat’s principle has a physical interpretation when one considers the wave nature of light. It is clear that around a stationary point of the optical path light, the maximum amount of light can be accumulated over different paths with a minimum of destructive interference. By the wave theory, light travels over all possible paths, but the light coming over the “wrong” paths destructively interferes, and only the light coming over the “right” path constructively interferes.

Fermat’s principle can be used to derive the basic laws of reflection and refraction (Snell’s law).

Now consider a perfect imaging system that takes all rays from an object and makes them all converge to an object. Since Fermat’s principle says the only paths taken will be those for which the OPL is minimally changed for small changes in path, the only way a perfect image will be formed is when all optical path lengths along a surface between an image and object point are the same - otherwise the light doesn’t get to this point!

Instead of using Fermat’s principle, we could solve for the parameters of a perfect surface using analytic geometry, but this would require an inspired guess for the correct functional form of the surface.

We find that the perfect surface depends on the situation: whether the light comes from a source at finite or infinite distance, and whether the mirror is concave or convex. We consider the various cases now, quoting the results without actually doing the geometry. In all cases, consider the z-axis to be the optical axis, with the y-axis running perpendicular. We want to know the shape of the surface, \(y(z)\), that gives a perfect image.

**Concave mirror with one conjugate at infinity**

Sample application: primary mirror of telescope looking at stars.

Fermat’s principle gives:

\[
y^2 = 2Rz
\]
where \( R = 2f \), the radius of curvature at the mirror vertex. This equation is that of a parabola. Note, however, that a parabola makes a perfect image only for on axis images (field angle=0).

**Concave mirror with both conjugates at finite distance**

Sample application: Gregorian secondary looking at image formed by primary.

For a concave mirror with both conjugates finite, we get an ellipse. Again, this is perfect only for field angle = 0.

\[
\frac{(z-a)^2}{a^2} + \frac{y^2}{b^2} = 1 \\
y^2 - 2zb^2/a + z^2b^2/a^2 = 0
\]

where

\[
a = (s + s')/2. \\
b = \sqrt{ss'} \\
R = ss'/(s + s') = 2b^2/a
\]

**Convex mirror with both conjugates at finite distance**

Sample application: Cassegrain secondary looking at image formed by primary.

For a convex mirror with both conjugates finite, we get a hyperbola:

\[
\frac{(z-a)^2}{a^2} - \frac{y^2}{b^2} = 1 \\
y^2 + 2zb^2/a - z^2b^2/a^2 = 0
\]

where

\[
a = (s + s')/2 \\
b^2 = -ss' \\
(s \text{ is negative}) \\
R = -2b^2/a
\]

**Convex mirror with one conjugate at infinity**

For a convex mirror with one conjugate at infinity, we get a parabola.

**2D to 3D**

Note that in all cases we’ve considered a one-dimension surface. We can generalize to 2D surfaces by rotating around the z-axis; for the equations, simply replace \( y^2 \) with \( (x^2 + y^2) \).

**Conic sections**

As you may recall from analytic geometry, all of these figures are conic sections, and it is possible to describe all of these figures with a single equation:
\[ \rho^2 - 2Rz + (1 + K)z^2 = 0 \]

where \[ \rho^2 = x^2 + y^2 \]

and \( R \) is the radius of curvature at the mirror vertex, \( K \) is called the conic constant \((K = -e^2, \text{ where } e \text{ is the eccentricity for an ellipse, } e(b, a)).\)

- \( K > 0 \) gives a prolate ellipsoid
- \( K = 0 \) gives a sphere
- \(-1 < K < 0 \) gives an oblate ellipsoid
- \( K = -1 \) gives a paraboloid
- \( K < -1 \) gives a hyperboloid

Know what optical shapes produce perfect images for different situations. Know the terminology: conic constant.

6.3.2 Aberrations: general description and low-order aberrations

Now consider what happens for surfaces that are not perfect, e.g. for the cases considered above for field angle \( \neq 0 \) (since only a sphere is symmetric for all field angles), or for field angle 0 for a conic surface which doesn’t give a perfect image?

You get aberrations; the light from all locations in aperture does not land at any common point.

One can consider aberrations in either of two ways:

1. aberrations arise from all rays not landing at a common point,
2. aberrations arise because wavefront deviates from a spherical wavefront.

These two descriptions are equivalent. For the former, one can talk about the transverse aberrations, which give the distance by which the rays miss the paraxial focus, or the angular aberration, which is the angle by which the rays deviate from the perfect ray which will hit paraxial focus. For the latter, one discusses the wavefront error, i.e., the deviation of the wavefront from a spherical wavefront as a function of location in the exit pupil.

In general, the angular and transverse aberrations can be determined from the optical path difference between a given ray and that of a spherical wavefront. The relations are given by:

\[ \text{angular aberration} = \frac{d(2\Delta z)}{d\rho} \]
transverse aberration = \frac{s'd(2\Delta z)}{d\rho}

If the aberrations are not symmetric in the pupil, then we could define angular and transverse \( x \) and \( y \) aberrations separately by taking derivatives with respect to \( x \) or \( y \) instead of \( \rho \).

**Spherical aberration**

First, consider the axisymmetric case of looking at an object on axis (field angle equal zero) with an optical element that is a conic section. We can consider where rays land as \( f(\rho) \), and derive the effective focal length, \( f_e(\rho) \), for an arbitrary conic section: in Figure 13

\[
\begin{align*}
  z_0 &= \rho/\tan(2\phi) = \rho(1 - (\tan \phi)^2)/(2 \tan \phi) \\
  \tan \phi &= dz/d\rho \\
  \text{from conic equation:} & \\
  \rho^2 - 2Rz + (1 + K)z^2 &= 0 \\
  z &= \frac{R}{(1 + K)} \left[ 1 - \left( 1 - \frac{\rho^2}{R^2} (1 + K) \right)^{1/2} \right] \\
  z &\approx \frac{\rho^2}{2R} + (1 + K)\frac{\rho^4}{8R^3} + (1 + K)^2\frac{\rho^6}{16R^5} + \ldots \\
  \frac{dz}{d\rho} &= \rho/(R - (1 - K)z) \\
  z_0 &= \frac{\rho}{2} \left[ \frac{R - (1 + K)z}{\rho} - \frac{\rho}{R - (1 + K)z} \right]
\end{align*}
\]
Astr 535 Class Notes – Fall 2019

Figure 14: Spherical aberration

\[ f_e = z + z_0 \]
\[ = \frac{R}{2} + \frac{(1 - K)z}{2} - \frac{\rho^2}{2(R - (1 + K)z)} \]
\[ = \frac{R}{2} - (1 + K) \frac{\rho^2}{4R} - (1 + K)(3 + K) \frac{\rho^4}{16R^3} - \ldots \]

\[ \Delta f = f_e - \frac{R}{2} \]

Note that \( f_e \) is independent of \( z \) only for \( K = -1 \), a parabola. Also note that \( \Delta f \) is symmetric with respect to \( \rho \).

We define spherical aberration as the aberration resulting from \( K \neq -1 \). Rays from different radial positions in the entrance aperture focus at different locations. It is an aberration which is present on axis as seen in Figure 14.

Spherical aberration is symmetric in the pupil. There is no location in space where all rays focus at a point. Note that the behavior (image size) as a function of focal position
is not symmetric. One can define several criteria for where the “best focus” might be, leading to the terminology paraxial focus, marginal focus, diffraction focus, and the circle of least confusion.

The asymmetric nature of spherical aberration as a function of focal position distinguishes it from other aberrations and is a useful diagnostic for whether a system has this aberration. This is shown in Figure 15.

We define transverse spherical aberration (TSA) as the image size at paraxial focus. This is not the location of the minimum image size.

\[
\frac{TSA}{\Delta f} = \frac{\rho}{(f - z(\rho))}
\]

\[
TSA = -(1 + K) \frac{\rho^3}{2R^2} - 3(1 + K)(3 + K) \frac{\rho^5}{8R^4} + \ldots
\]

The difference in angle between the “perfect” ray from the parabola and the actual ray is called the angular aberration, in this case angular spherical aberration, or ASA.

\[
ASA = 2(\phi_p - \phi) \approx \frac{d}{d\rho} (2\Delta z) \approx -(1 + K) \frac{\rho^3}{R^3}
\]

where \(2\Delta z\) gives the optical path difference between the two rays.

This is simply related to the transverse aberration:

\[
TSA = \frac{R}{2} ASA
\]

We can also consider aberration as the difference between our wavefront and a spherical wavefront, which in this case is the wavefront given by a parabolic surface.

\[
\Delta z = z_{\text{parabola}} - z(K) = -\frac{\rho^4}{8R^3}(1 + K) + \ldots
\]

This result can be generalized to any sort of aberration: the angular and transverse aberrations can be determined from the optical path difference between a given ray and that of a spherical wavefront. The relations are given by:

\[
\text{angular aberration} = \frac{d(2\Delta z)}{d\rho}
\]

\[
\text{transverse aberration} = s' \frac{d(2\Delta z)}{d\rho}
\]
Figure 15: Spherical aberration image sequence as function of focal position
Figure 16: Spherical aberration diagram

Figure 17: Spherical aberration diagram

Figure 18: Spherical aberration diagram
If the aberrations are not symmetric in the pupil, then we could define angular and transverse $x$ and $y$ aberrations separation by taking derivatives with respect to $x$ or $y$ instead of $\rho$.

**General aberration description**

We can describe deviations from a spherical wavefront generally. Since all we care about are optical path differences, we write an expression for the optical path difference between an arbitrary ray and the chief ray, and in doing this, we can also include the possibility of an off-axis image, and get

$$OPD = OPL - OPL(\text{chief ray})$$

where we’ve kept terms only to fourth order and chosen our coordinate system such that the object lies in the $y$-$z$ plane. The coefficients, $A_i$, depend on lots of things, such as $(\theta, K, n, R, s, s')$.

Note that rays along the $y$-axis are called *tangential* rays, while rays along the $x$-axis are called *sagittal* rays.

Analytically, people generally restrict themselves to talking about *third-order* aberrations, which are fourth-order (in powers of $x, y, \rho, \theta$) in the optical path difference, because of the derivative we take to get transverse or angular aberrations. In the third-order limit, one finds that $A_2 = A_2'$, and $A_1 = -A_1'$. Working out the geometry, we find for a mirror that:

$$A_0 = 0$$
$$A_1 = \frac{n\theta^2}{R}$$
$$A_2 = -\frac{n\theta}{R^2} \left( \frac{m+1}{m-1} \right)$$
$$A_3 = \frac{n}{4R^3} \left[ K + \left( \frac{m+1}{m-1} \right)^2 \right]$$

From the general expression, we can derive the angular or the transverse aberrations in either the $x$ or $y$ direction. Considering the aberrations in the two separate directions, we find:

$$AA_y = 2A_1y + A_2(x^2 + 3y^2) + 4A_3y\rho^2$$
$$AA_x = 2A_1'x + 2A_2xy + 4A_3x\rho^2$$

The first term is proportional to $\theta^2 y$ and is called *astigmatism*. The second term is proportional to $\theta(x^2 + 3y^2)$ and is called *coma*. The final term, proportional to $y\rho^2$ is...
Figure 19: Rays in the presence of astigmatism.

spherical aberration, which we’ve already discussed (note for spherical, \( AA_x = AA_y \) and in fact the AA in any direction is equal, hence the aberration is circularly symmetric).

**Astigmatism**

For astigmatism, rays from opposite sides of the pupil focus in different locations relative to the paraxial rays. At the paraxial focus, we end up with a circular image. As you move away from this image location, you move towards the tangential focus in one direction and the sagittal focus in the other direction. At either of these locations, the astigmatic image looks like an elongated ellipse. Astigmatism goes as \( \theta^2 \), and consequently looks the same for opposite field angles. Astigmatism is characterized in the image plane by the transverse or angular astigmatism (TAS or AAS), which refer to the height of the marginal rays at the paraxial focus. Astigmatism is symmetric around zero field angle.

Figure 19 shows the nature of astigmatism.

Figure 20 shows the behavior of astigmatism as one passes through paraxial focus.

**Coma**

For coma, rays from opposite sides of the pupil focus at the same focal distance. However, the tangential rays focus at a different location than the sagittal rays, and neither of these focus at the paraxial focus. The net effect is to make an image that vaguely looks like a comet, hence the name coma. Coma goes as \( \theta \), so the direction of the comet flips sign for opposite field angles. Coma is characterized by either the tangential or sagittal transverse/angular coma (TTC, TSC, ATC, ASC) which describe the height/angle of either the tangential or sagittal marginal rays at the paraxial focus: \( TTC = 3TSC \).

Figure 21 shows the nature of coma.

Figure 22 shows the behavior of coma as one passes through paraxial focus.
Figure 20: Astigmatism image sequence as function of focal position
In fact, there are two more third-order aberrations: distortion and field curvature. Neither affects image quality, only location (unless you are forced to use a flat image plane!). Field curvature gives a curved focal plane: if imaging onto a flat detector, this will lead to focus deviations as one goes off-axis. Distortion affects the location of images in the focal plane, and goes as $\theta^3$. The amount of field curvature and distortion can be derived from the aberration coefficients and the mirror parameters.

We can also determine the relevant coefficients for a surface with a displaced stop (Schroeder p 77), or for a surface with a decentered pupil (Schroeder p89-90); it’s just more geometry and algebra. With all these relations, we can determine the optical path differences for an entire system: for a multi-surface system, we just add the OPD’s as we go from surface to surface. The final aberrations can be determined from the system OPD.

Understand the basic concepts of aberration. Know what the five third-order aberrations are (spherical, coma, astigmatism, field curvature, distortion) and have a basic idea about how they affect image quality and/or location.

6.3.3 Aberration compensation and different telescope types

Using the techniques above, we can write expressions for the system aberrations as a function of the surface figures (and field angles). If we give ourselves the freedom to choose surface figures, we can eliminate one (or more) aberrations.

For example, given a conic constant of the primary mirror, we can use the aberration relations to determine $K_2$ such that spherical aberration is zero; this will give us perfect
Figure 22: Coma image sequence as function of focal position
images on-axis. We find that:

\[ K_2 = -\left(\frac{(m + 1)}{(m - 1)}\right)^2 + \frac{m^3}{k(m - 1)^3}(K_1 + 1) \]

satisfies this criterion. If we set the primary to be a parabola \((K_1 = -1)\), this gives the conic constant of the secondary we must use to avoid spherical aberration. This type of telescope is called a classical telescope. Using the aberration relations, we can determine the amount of astigmatism and coma for such telescopes, and we find that coma gives significantly larger aberrations than astigmatism, until one gets to very large field angles.

If we allow ourselves the freedom to choose both \(K_1\) and \(K_2\), we can eliminate both spherical aberration and coma. Designs of this sort are called aplanatic. The relevant expression, in terms of the magnification and back focal distance (we could use the relations discussed earlier to present these in terms of other paraxial parameters), is:

\[ K_1 = -1 - \frac{2(1 + \beta)}{m^2(m - \beta)} \]

We can only eliminate two aberrations with two mirrors, so even this telescope will be left with astigmatism.

There are two different classes of two-mirror telescopes that allow for freedom in the shape of both mirrors: Cassegrain telescopes and Gregorian telescopes (Newtonians have a flat secondary). For the classical telescope with a parabolic primary, the Cassegrain secondary is hyperbolic, whereas for a Gregorian it is ellipsoidal (because of the appropriate conic sections derived above for convex and concave mirrors with finite conjugates). For the aplanatic design, the Cassegrain telescope has two hyperbolic mirrors, while the Gregorian telescope has two ellipsoidal mirrors. An aplanatic Cassegrain telescope is called a Ritchey-Chretien telescope.

The following table gives some characteristics of “typical” telescopes. Aberrations are given at a field angle of 18 arc-min in units of arc-seconds. Coma is given in terms of tangential coma.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>CC</th>
<th>CG</th>
<th>RC</th>
<th>AG</th>
</tr>
</thead>
<tbody>
<tr>
<td>(m)</td>
<td>4.00</td>
<td>-4.00</td>
<td>4.00</td>
<td>-4.00</td>
</tr>
<tr>
<td>(k)</td>
<td>0.25</td>
<td>-0.417</td>
<td>0.25</td>
<td>-0.417</td>
</tr>
<tr>
<td>1 - k</td>
<td>0.75</td>
<td>1.417</td>
<td>0.75</td>
<td>1.417</td>
</tr>
<tr>
<td>(mk)</td>
<td>1.000</td>
<td>1.667</td>
<td>1.000</td>
<td>1.667</td>
</tr>
<tr>
<td>ATC</td>
<td>2.03</td>
<td>2.03</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>AAS</td>
<td>0.92</td>
<td>0.92</td>
<td>1.03</td>
<td>0.80</td>
</tr>
<tr>
<td>ADI</td>
<td>0.079</td>
<td>0.061</td>
<td>0.075</td>
<td>0.056</td>
</tr>
<tr>
<td>(\kappa_m R_1)</td>
<td>7.25</td>
<td>-4.75</td>
<td>7.625</td>
<td>-5.175</td>
</tr>
<tr>
<td>(\kappa_p R_1)</td>
<td>4.00</td>
<td>-8.00</td>
<td>4.00</td>
<td>-8.00</td>
</tr>
</tbody>
</table>
The image quality is clearly better for the aplanatic designs than for the classical designs, as expected because coma dominates off-axis in the classical design. In the aplanatic design, the Gregorian is slightly better. However, when considerations other than just optical quality are considered, the Cassegrain usually is favored: for the same primary mirror, the Cassegrain is considerably shorter and thus it is less costly to build an enclosure and telescope structure. To keep the physical length the same, the Gregorian would have to have a faster primary mirror, which are more difficult (i.e. costly) to fabricate, and which will result in a greater sensitivity to alignment errors. Both types of telescopes have a curved focal plane.

Understand how multi-surface systems can be used to reduce or remove aberrations. Know the terminology: aplanatic telescope, Ritchey-Chretien telescope.

6.4 Sources of aberrations

So far, we have been discussing aberrations which arise from the optical design of a system when we have a limited number of elements. However, it is important to realize that aberrations can arise from other sources as well. These other sources can give additional third-order aberrations, as well as higher order aberrations. Some possible sources include:

- design: as we have seen, it may not be possible to remove all aberrations with a limited number of surfaces
- misfigured or imperfectly figured optics: rarely is an element made exactly to specification!
- misalignments. If the mirrors in a multiple-element system are not perfectly aligned, aberrations will result. These can be derived (third-order) from the aberration expressions for decentered elements. For two mirror systems, one finds that decentering or tilting the secondary introduces a constant amount of coma over the field. Coma dominates astigmatism for a misaligned telescope.
- mechanical/support problems. When the mirrors are mounted in mirror cells the weight of the mirror is distributed over some support structures. Because the mirrors are not infinitely stiff, some distortion of the mirror shape will occur. Generally, such distortion will probably change as a function of which way the telescope is pointing. Separate from this, because the telescope structure itself is not perfectly stiff, one expects some flexure which gives a different secondary (mis)alignment as a function of where one is pointing. Finally, one might expect the spacing between the primary and secondary to vary with temperature, if the telescope structure is made of materials which have non-zero coefficients of expansion.
chromatic aberration. Generally, we’ve only been discussing mirrors since this is what is used in telescopes. However, astronomers often put additional optics (e.g., cameras or spectrographs) behind telescopes which may use refractive elements rather than mirrors. There are aberration relations for refractive elements just as we’ve discussed, but these have additional dependences on the indices of refraction of the optical elements. For most refractive elements, the index of refraction varies with wavelength, so one will get wavelength-dependent aberrations, called chromatic aberrations. These can be minimized by good choices of materials or by using combinations of different materials for different elements; however, it is an additional source of aberration.

seeing. The earth’s atmosphere introduces optical path differences between the rays across the aperture of the telescope. This is generally the dominant source of image degradation from a ground-based telescope. Consequently, one builds telescopes in good sites, and as far as design and other sources of image degradation are concerned, one is generally only interested in getting these errors small when compared with the smallest expected seeing errors.

6.5 Ray tracing

For a fully general calculation of image quality, one does not wish to be limited to third-order aberrations, nor does one often wish to work out all of the relations for the complex set of aberrations which result from all of the sources of aberration mentioned above. Real world situations also have to deal with vignetting in optical systems, in which certain rays may be blocked by something and never reach the image plane (e.g., in a two-mirror telescope, the central rays are blocked by the secondary).

Because of these and other considerations, analysis of optical systems is usually done using ray tracing, in which the parameters of an optical system are entered into a computer, and the computer calculates the expected images on the basis of geometric optics. Many programs exist with many features: one can produce spot diagrams which show the location of rays from across the aperture at an image plane (or any other location), plots of transverse aberrations, plots of optical path differences, etc., etc.

6.6 Physical (diffraction) optics

Up until now, we have avoided considering the wave nature of light which introduces diffraction from interference of light coming from different parts of the aperture. Because of diffraction, images of a point source will be slightly blurred. From simple geometric arguments, we can estimate the size of the blur introduced from diffraction: as shown in Figure 23
Diffraction is expected to be important when $\Delta \sim \lambda$, i.e.,

$$\theta \sim \frac{\lambda}{D}$$

Using this, we find that the diffraction blur is smaller than the blur introduced by seeing for $D > 0.2$ meters at 5500 Å, even for the excellent seeing conditions of 0.5 arcsecond images. However, the study of diffraction is relevant because of several reasons: 1) the existence of the Hubble Space Telescope (and other space telescopes), which is diffraction limited (no seeing), 2) the increasing use of infrared observations, where diffraction is more important than in the optical, and 3) the development of adaptive optics, which attempts to remove some of the distortions caused by seeing. Consequently, it’s now worthwhile to understand some details about diffraction.

To work out in detail the shape of the images formed from diffraction involves understanding wave propagation. Basically, one integrates over all of the source points in the aperture (or exit pupil for an optical system), determining the contribution of each point at each place in the image plane. The contributions are all summed taking into account phase differences at each image point, which causes reinforcement at some points and cancellation at others. The expression which sums all of the individual source points is called the _diffraction integral_. When the details are worked out, one finds that the intensity in the image plane is related to the intensity and phase at the exit pupil. In fact the wavefront is described at any plane by the _optical transfer function_, which gives the intensity and phase of the wave at all locations in that plane. The OTF at the pupil plane and at the image plane are a Fourier transform pair. Consequently, we can determine the light distribution in the image plane by taking the Fourier transform of the pupil plane; the light distribution, or point spread function, is just the modulus-squared of the OTF.
at the image plane. Symbolically, we have

\[ PSF = \left| \int (OTF(pupil)) \exp ikx \right|^2 \]

where

\[ OTF(pupil) = P(x, y) \exp ik\phi(x, y) \]

\( P(x, y) \) is the pupil function, which gives the transmission properties of the pupil, and usually consists of ones and zeros for locations where light is either transmitted or blocked (e.g., for a circular lens, the pupil function is unity within the radius of lens, and zero outside; for a typical telescope the pupil function includes obscuration by the secondary and secondary support structure). \( \phi \) is the phase in the pupil. More relevantly, \( \phi \) can be taken to be the optical path difference in the pupil with some fiducial phase, since only OPDs matter, not the absolute phase. Finally the wavenumber \( k \) is just \( \frac{2\pi}{\lambda} \).

For the simple case of a plane wave with no phase errors, the diffraction integral can be solved analytically. The result for a circular aperture with a central obscuration, when the fractional radius of the obscuration is given by \( \epsilon \), the expression for the PSF is:

\[ PSF \propto \left[ \frac{2J_1(v)}{v} - \epsilon^2 \frac{2J_1(\epsilon v)}{\epsilon v} \right]^2 \]

\[ v = \frac{\pi r}{\lambda F} \]

where \( J_1 \) is a first order Bessel function, \( r \) is the distance in the image plane, \( \lambda \) is the wavelength, and \( F \) is the focal ratio \( (F = f/D) \).

This expression gives the so-called Airy pattern which has a central disk surrounded by concentric dark and bright rings. One finds that the radius of the first dark ring is at the physical distance \( r = 1.22\lambda F \), or alternatively, the angular distance \( \alpha = 1.22\lambda/D \). This gives the size of the Airy disk.

For more complex cases, the diffraction integral is solved numerically by doing a Fourier transform. The pupil function is often more complex than a simple circle, because there are often additional items which block light in the pupil, such as the support structures for the secondary mirror.

Figure 24 shows the Airy pattern, both without obscurations, and with a central obscuration and spiders in a setup typical of a telescope.

In addition, there may be phase errors in the exit pupil, because of the existence of any one of the sources of aberration discussed above. For general use, \( \phi \) is often expressed as an series, where the expansion is over a set of orthogonal polynomials for the aperture which is being used. For circular apertures with (or without) a central obscuration (the case most often found in astronomy), the appropriate polynomials are
Figure 24: The Airy pattern, with and without obscurations
called Zernike polynomials. The lowest order terms are just uniform slopes of phase across the pupil, called tilt, and simply correspond to motion in the image plane. The next terms correspond to the expressions for the OPD which we found above for focus, astigmatism, coma, and spherical aberration, generalized to allow any orientation of the phase errors in the pupil. Higher order terms correspond to higher order aberrations.

Figure 25 shows the form of some of the low order Zernike terms: the first corresponds to focus aberration, the next two to astigmatism, the next two to coma, the next two to trefoil aberration, and the last to spherical aberration.

A wonderful example of the application of all of this stuff was in the diagnosis of spherical aberration in the Hubble Space Telescope, which has been corrected in subsequent instruments in the telescope, which introduce spherical aberration of the opposite sign. To perform this correction, however, required and accurate understanding of the amplitude of the aberration. This was derived from analysis of on-orbit images, as shown in Figure 26.

Note that it is possible in some cases to try to recover the phase errors from analysis of images. This is called phase retrieval. There are several ways of trying to do this, some of which are complex, so we won’t go into them, but it’s good to know that it is possible. But an accurate amplitude of spherical aberration was derived from these images. This derived value was later found to correspond almost exactly to the error expected from an error which was made in the testing facility for the HST primary mirror, and the agreement of these two values allowed the construction of new corrective optics to proceed...

Some figures from HST Optical Systems Failure Report.

Understand the principles of diffraction optics and, in particular, how diffraction scales with wavelength and aperture diameter. Know the terminology: optical transfer function, pupil function, and how phase errors across the pupil can be decomposed into a series of Zernike polynomials.

6.7 Adaptive Optics

The goal of adaptive optics is to partially or entirely remove the effects of atmospheric seeing. Note that these day, this is to be distinguished form active optics, which works at lower frequency, and whose main goal is to remove aberrations coming from the change in telescope configuration as the telescope moves (e.g., small changes in alignment from flexure or sag of the primary mirror surface as the telescope moves). Active optics generally works as frequencies less than (usually significantly) 1 Hz, whereas adaptive optics must work at 10 to 1000 Hz. At low frequencies, the active optics can be done with actuators on the primary and secondary mirrors themselves. At the high frequencies required
Figure 25: Images of Zernike terms, orders 4-11.
Figure 26: HST images and corresponding models
for adaptive optics, however, these large mirrors cannot respond fast enough, so one is required to form a pupil on a smaller mirror which can be rapidly adjusted; hence adaptive optics systems are really separate astronomical instruments.

Many adaptive optics systems are functioning and/or under development: see ESO/VLT adaptive optics, CFHT adaptive optics, Keck adaptive optics, Gemini adaptive optics, http://www.cfht.hawaii.edu/Instruments/Imaging/AOB/other-aosystems.html

The basic idea of an adaptive optics system is to rapidly sense the wavefront errors and then to correct for them on timescales faster than those at which the atmosphere changes. Consequently, there are really three parts to an adaptive optics system:

1. a component which senses wavefront errors,
2. a control system which figures out how to correct these errors, and
3. an optical element which receives the signals from the control system and implements wavefront corrections.

There are several methods used for wavefront sensing. Two ones in fairly common use among today’s adaptive optics system are Shack-Hartmann sensors and wavefront curvature sensing devices. In a Shack Hartman sensor, as shown in Figure 27,
an array of lenslets is put in a pupil plane and each lenslet images a small part of the pupil. Measuring image shifts between each of the images gives a measure of the local wavefront tilts. Wavefront curvature devices look at the intensity distribution in out-of-focus images. Other wavefront sensing techniques include pyramid wavefront sensors and phase diversity techniques. Usually, a star is used as the source, but this is not required for some wavefront sensors (i.e., extended source can be used).

To correct wavefront errors, some sort of deformable mirror is used. These can be generically split into two categories: segmented and continuous faceplate mirrors, where the latter are more common. A deformable mirror is characterized by the number of adjustable elements: the more elements, the more correction can be done. LCD arrays have also been used for wavefront correction.

In general, it is very difficult to achieve complete correction even for ideal performance, and one needs to consider the effectiveness of different adaptive optics systems. This effectiveness depends on the size of the aperture, the wavelength, the number of resolution elements on the deformable mirror, and the quality of the site. Clearly, more resolution elements are needed for larger apertures. Equivalently, the effectiveness of a system will decrease as the aperture is increased for a fixed number of resolution elements. One can consider the return as a function of Zernike order corrected and aperture size. For large telescopes, you’ll only get partial correction unless a very large number of resolution elements on the deformable mirror are available. The following table gives the mean square amplitude, $\Delta_j$, for Kolmogorov turbulence after removal of the first $j$ terms; the rms phase variation is just $\sqrt{\Delta_j}/2\pi$. For small apertures, you can make significant gains with removal of just low order terms, but for large apertures you need very high order terms. Note various criteria for quality of imaging, e.g., $\lambda/4$, etc.

<table>
<thead>
<tr>
<th>$Z_j$</th>
<th>n</th>
<th>m</th>
<th>Expression</th>
<th>Description</th>
<th>$\Delta_j$</th>
<th>$\Delta_j - \Delta_{j-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z1</td>
<td>0</td>
<td>0</td>
<td>constant</td>
<td>constant</td>
<td>1.030</td>
<td>S</td>
</tr>
<tr>
<td>Z2</td>
<td>1</td>
<td>1</td>
<td>$2r \cos \phi$</td>
<td>tilt</td>
<td>0.582 S</td>
<td>0.448 S</td>
</tr>
<tr>
<td>Z3</td>
<td>1</td>
<td>1</td>
<td>$2r \sin \phi$</td>
<td>tilt</td>
<td>0.134 S</td>
<td>0.448 S</td>
</tr>
<tr>
<td>Z4</td>
<td>2</td>
<td>1</td>
<td>$\sqrt{3(2r^2 - 1)}$</td>
<td>defocus</td>
<td>0.111 S</td>
<td>0.023 S</td>
</tr>
<tr>
<td>Z5</td>
<td>2</td>
<td>2</td>
<td>$\sqrt{6r^4 \sin 2\phi}$</td>
<td>astigmatism</td>
<td>0.088 S</td>
<td>0.023 S</td>
</tr>
<tr>
<td>Z6</td>
<td>2</td>
<td>2</td>
<td>$\sqrt{6r^4 \cos 2\phi}$</td>
<td>astigmatism</td>
<td>0.0648 S</td>
<td>0.023 S</td>
</tr>
<tr>
<td>Z7</td>
<td>3</td>
<td>1</td>
<td>$\sqrt{8(3r^3 - 2r) \sin \phi}$</td>
<td>coma</td>
<td>0.0587 S</td>
<td>0.0062 S</td>
</tr>
<tr>
<td>Z8</td>
<td>3</td>
<td>1</td>
<td>$\sqrt{8(3r^3 - 2r) \cos \phi}$</td>
<td>coma</td>
<td>0.0525 S</td>
<td>0.0062 S</td>
</tr>
<tr>
<td>Z9</td>
<td>3</td>
<td>3</td>
<td>$\sqrt{8r^3 \sin 3\phi}$</td>
<td>trifoil</td>
<td>0.0463 S</td>
<td>0.0062 S</td>
</tr>
<tr>
<td>Z10</td>
<td>3</td>
<td>3</td>
<td>$\sqrt{8r^3 \cos 3\phi}$</td>
<td>trifoil</td>
<td>0.0401 S</td>
<td>0.0062 S</td>
</tr>
<tr>
<td>Z11</td>
<td>4</td>
<td>0</td>
<td>$\sqrt{5(6r^4 - 6r^2 + 1)}$</td>
<td>spherical</td>
<td>0.0377 S</td>
<td>0.0024 S</td>
</tr>
</tbody>
</table>

$r = \text{distance from center circle}; \phi = \text{azimuth angle}; S = (D/r_0)^{5/3}$.

Another important limitation is that one needs an object on which you can derive the
Measurements of wavefront are subject to noise just like any other photon detection so bright sources may be required. This is even more evident when one considers that you need a source which is within the same isoplanatic patch as your desired object, and when you recall that the wavefront changes on time scales of milliseconds. These requirements place limitations on the amount of sky over which it is possible to get good correction. It also places limitations on the sorts of detectors which are needed in the wavefront sensors (fast readout and low or zero readout noise).

\[ \text{Conditions are: } \theta_0 = 0.3 \text{ arcsec seeing at } \mu \text{; } \tau_{\text{det}} \sim 0.3 \tau_0 = 0.3r/V\text{wind}; V\text{wind} = 10 \text{ mIsec; } H = 5000; \text{ photon detection efficiency (includes transmission and QE) = 20\%; spectral bandwidth = 300 nm; } SNR = 100 \text{ per Hartmann-Shack image; detector noise = } 5e^- . \]

The isoplanatic patch limitation is severe. In many cases, we might expect non-optimal performance if the reference object is not as close as it should be ideally.

In most cases, both because of lack of higher order correction and because of reference star vs. target wavefront differences, adaptive optics works in the partially correcting regime. This typically gives PSFs with a sharp core, but still with extended wings.

The problem of sky coverage can be avoided if one uses so-called laser guide stars. The idea is to create a star by shining a laser up into the atmosphere. To date, two generic classes of lasers have been used, Rayleigh and sodium beacons. The Rayleigh beacons work by scattering off a layer roughly 30 km above the Earth’s surface; the sodium beacons work by scattering off a layer roughly 90 km above the Earth’s surface. Laser guide stars still have some limitations. For one, the path through the atmosphere which the laser traverses does not exactly correspond to the path that light from a star traverses, because the latter comes from an essentially infinite distance; this leads to the effect called focal anoisoplanatism. In addition, laser guide stars cannot generally be used to track image motion since the laser passes up and down through the same atmosphere and image motion is cancelled out. To correct for image motion, separate tip-tilt tracking is required.

Note that even with perfect correction, one is still limited by the isoplanatic patch.
size. As one moves further and further away from the reference object, the correction will gradually degrade, because a different path through the atmosphere is being probed.

To get around this, one can consider the use of multiple laser guides stars (laser guide star constellation) to characterize the atmosphere over a broader column. However, if this is done, one cannot correct all field angles simultaneously at the telescope pupil, because the aberrations are different for different field angles. Instead, one could choose to correct them in a plane conjugate to the location of the dominant source of atmospheric aberration. This is the basis of a ground layer adaptive optics (GLAO) system, where a correction is made for aberration in the lower atmosphere.

In principle, even better correction over a wider field of view is possible with multiple deformable mirrors, giving rise to the concept of multi-conjugate adaptive optics (MCAO) systems. In such systems, each adaptive optic would correct at a different location in the atmosphere.

Systems with single laser guide stars have certainly been tested and appear to work; but remember, only over an isoplanatic patch, and often with partially corrected images. Several implementations of system with multiple guide stars actually exist (at VLT and Keck?) to allow sampling of a larger cylinder/cone through the atmosphere; some of these are designed to correct at particular layers to maximize FOV, e.g. ground layer adaptive optics (GLAO). The bulk of adaptive optics work has been done in the near-IR.

Extreme (high-contrast) AO.

A variant on adaptive optics: lucky imaging.

Science with adaptive optics. Typical AO PSFs. Morphology vs. photometry.

### 6.7.1 AO Examples

- Gemini AO animation video
  - [http://www.alpao.com/Applications/Adaptive_optics_for_Astronomy.htm](http://www.alpao.com/Applications/Adaptive_optics_for_Astronomy.htm)
  - Galactic center AO (see bottom of page, note scale)
- Neptune
- Neptune movie
- on/off sun image
- Simulated seeing on bench
- Young star video

*Understand the principles of how adaptive optics systems work. Understand the challenges of adaptive optics: getting sufficient light for wavefront sensing, isoplanatic patch size, partial correction, and how they can be at least partially addressed, e.g., with laser guide stars and multiconjugate adaptive optics.*