

29

Aberrations

29.1	Introduction	2
29.2	Power Series Expansions	8
29.3	Chromatic Aberrations	13
29.4	Primary Aberrations	16
29.4.1	Aperture and Field Dependence	16
29.4.2	Symmetry and Periodicity Properties	18
29.4.3	Presentation of Aberrations and their Impact on Image Quality	20
29.4.4	Calculation of the Seidel Sums	29
29.4.5	Stop Shift Formulae	36
29.4.6	Several Aberration Expressions from the Seidel Sums	38
29.4.7	Thin Lens Aberrations	41
29.5	Pupil Aberrations	45
29.6	High-order Aberrations	50
29.6.1	Fifth-order Aberrations	50
29.6.2	Seventh and Higher-order Aberrations	53
29.7	Zernike Polynomials	55
29.8	Special Aberration Formulae	56
29.8.1	Sine Condition and the Offence against the Sine Condition	57
29.8.2	Herschel Condition	60
29.8.3	Aplanatism and Isoplanatism	61
29.8.4	Aldis Theorem	61
29.8.5	Spherical Aberration, a Surface Contribution Formula	64
29.8.6	Aplanatic Surface and Aplanatic Lens	68
29.9	Literature	70

29.1

Introduction

In this and in the following section (29.2) we will deal with monochromatic aberrations only. In section 29.3 chromatic aberrations will be introduced and will then be included in our further discussions. As we have already explained in chapter 11, in general, a real optical imaging system does not perform ideal imaging. So, rays emerging from one object point O will not all meet at a single image point O' . An example with three meridional rays is shown in figure 29-1.

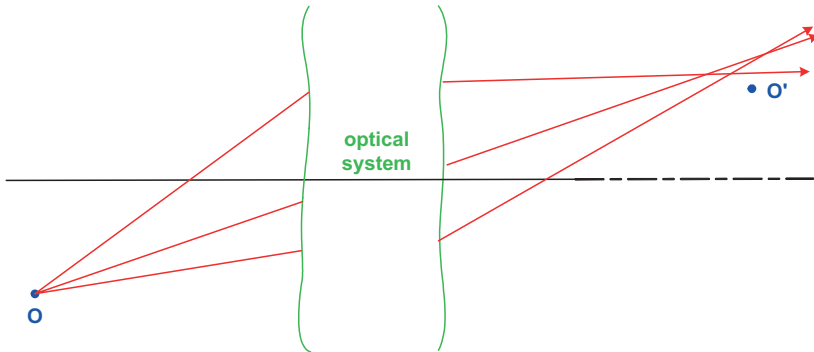


Figure 29-1: Aberrations of an imaging system.

There are several methods used to describe how the rays miss the image point, see section 11.2, Description of Aberrations. Very often it is convenient to think in terms of transverse aberrations. Consider a single object point and a given position of the image plane and on this plane consider a given reference point, which is the point where the image should be. Usually the assumed image point will be the Gaussian image point, but this is in fact not necessary. The position of the image plane as well as the choice of the image reference point may differ from the Gaussian values if required. For instance, one reason for choosing the Gaussian image plane but not the Gaussian image height may be some distortion if distortion is not of interest. In this case the intersection of the real chief ray with the image plane may be adequate as the image reference point. A very common reason for choosing image planes, which are different from the Gaussian image plane, is to study the behavior of the aberrations with a change in the focus.

There are several ways to present transverse aberrations in graphical form. We will outline some examples for a very simple system, a single biconvex lens as shown in figure 29-2. Two field points are considered, one field point is on the axis, while the second field point is off the axis.

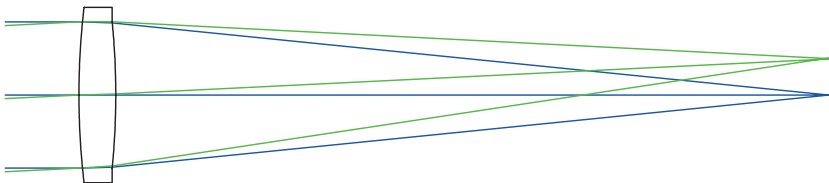


Figure 29-2: Biconvex lens, $f' = 100$ mm, $F/5$, image height 5 mm.

Spot diagrams are a very popular way of presenting transverse aberrations, see figures 29-3 and 29-4. The through-focus spot diagrams show the characteristics of a lot of aberrations and suggest the size of the image blur. However, even with this simple example, it can be seen that the impression of the ray intersection spots strongly depends on the chosen ray distribution in the pupil. For figure 29-3 a rectangular pupil grid is used and for figure 29-4 a polar pupil grid is used.

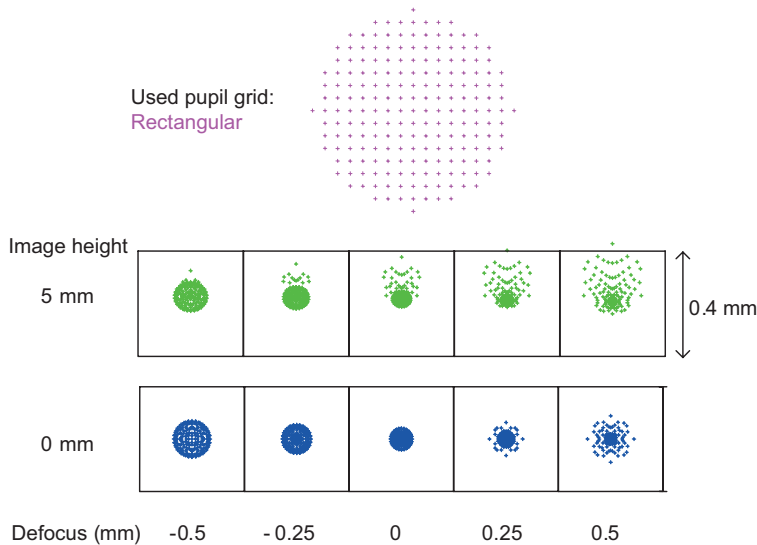


Figure 29-3: Through-focus spot diagram with rectangular pupil grid.

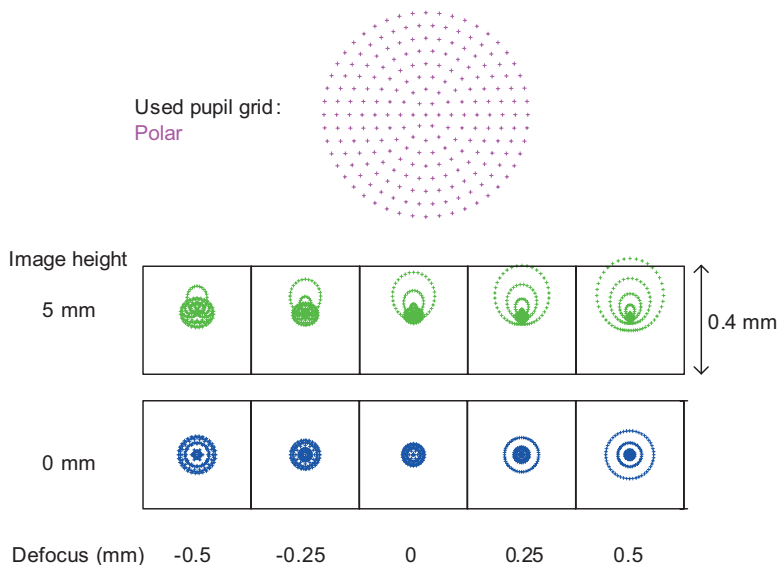


Figure 29-4: Through-focus spot diagram with polar pupil grid.

In general it is difficult to distinguish typical aberrations from spot diagrams. However, for this purpose the transverse aberration fans are an adequate means, see figure 29-5. Here the transverse aberrations of two pupil sections are represented: For the on-axis image the ray bundle is symmetric with respect to the optical axis, so all information is contained in the presentation of the meridional section only. The meridional pupil section is also called the tangential pupil section. For an image height of 5 mm, in the tangential pupil section, due to the rotational symmetry of the system, all aberrations are in the image y -direction. From the sagittal pupil section the rays have aberration components y' in image y -direction as well as components x' in the image x -direction and these components exhibit characteristic symmetries (point symmetry and mirror symmetry, respectively) as shown in figure 29-5. Usually, for the sagittal section, only the more important x -component of the aberration is shown. As will be explained later, the transverse aberration fans in figure 29-5 are clearly dominated by spherical aberration and also by coma.

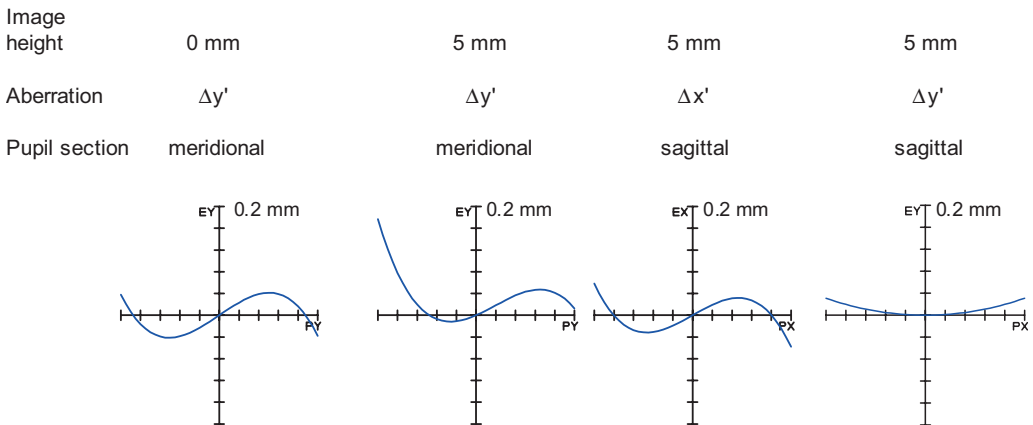


Figure 29-5: Transverse ray aberrations.

The use of transverse aberrations is a powerful method and all types of aberration, such as spherical aberration, coma, astigmatism, field curvature and distortion, as well as axial chromatic aberration and lateral chromatic aberration can be understood and represented as transverse aberrations.

Nevertheless, there are situations where it is desirable to use longitudinal aberrations. For instance, astigmatism and field curvature are easily understood as longitudinal phenomena. But spherical aberration, coma, and axial chromatic aberration can also be understood as longitudinal aberrations. Figure 29-6 shows the spherical aberration and the astigmatic field curves of the system considered in figure 29-2 as longitudinal aberrations. The only aberrations, which cannot be described as longitudinal aberrations, are distortion and lateral chromatic aberration.

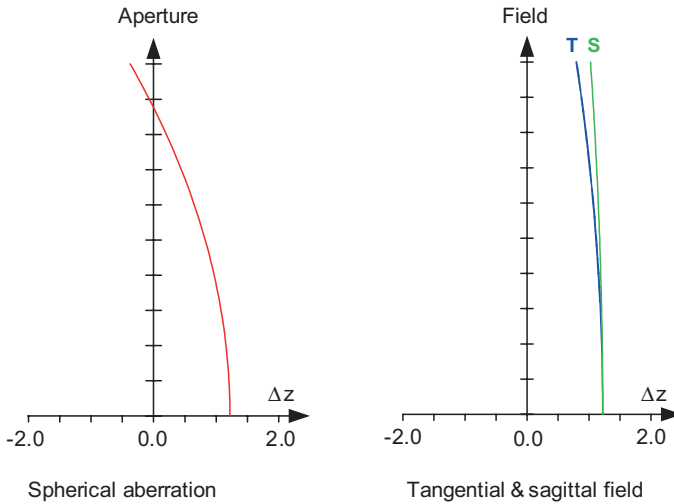


Figure 29-6: Longitudinal ray aberration curves.

For optical systems which image to infinity, the transverse as well as the longitudinal aberrations do not make sense, as they both become infinite. In this case, for the image at infinity, instead of transverse aberrations measured in length units, angular aberrations measured in angle units will be adequate. So angle aberrations, which were introduced in section 11.2, Description of Aberrations, correspond to transverse aberrations and exhibit the respective properties. Also, for longitudinal aberrations there is an adequate representation when the image is at infinity: Instead of the longitudinal aberration itself, the reciprocal value of the intersection lengths is used. As the unit of dimension diopters are normally used. To express an intersection length S' in diopters, S' should be measured in m. Then the corresponding value \tilde{S}' in diopters is defined as

$$\tilde{S}' [\text{diopter}] = \frac{1}{S'[\text{m}]} \quad (29-1)$$

So aberrations measured in diopters correspond to the longitudinal aberrations. For instance, in vision optics, diopters are the preferred aberration description.

Transverse and longitudinal ray aberrations are easy to understand, and they represent a complete and powerful method of describing the aberrational behavior of an optical imaging system. So, for what reason do we need wave aberrations? There are, in fact, some specific advantages connected with the understanding and the use of wave aberrations. The most important benefits of a description based on wave aberrations are as follows.

- Wavefront aberrations can be measured very easily and very accurately by means of interferometric methods. This is a big advantage because the measurement of ray aberrations with any comparable completeness and accuracy is almost impossible.

- In spite of the fact that wavefront aberrations are strictly a geometrical optics method there is a strong relationship to physical image formation. As pointed out in chapter 12, the wave aberration is the essential input for calculating the diffraction image. As an aberration-free system would image an object point not as a single image point but as the Airy pattern, it is understandable that very small aberrations would not change the Airy pattern significantly. In this case the image quality is determined more by diffraction effects than by geometrical aberrations. When working with wave aberrations there is a very useful rule of thumb, the so-called Rayleigh limit: If the wave aberration is less than one quarter of the wave length, the system can be regarded as diffraction limited.
- When considering surface contributions, wave aberrations are convenient in that the single surface contributions sum up directly to the total aberration. It is possible to obtain a surface contribution formula for transverse aberrations, see section 29.8.4, Aldis Theorem, but these formulae are not as plausible as are the wave aberrations.
- Finally, aberration theory informs us that there are distinct types of aberrations such as spherical aberration, coma, astigmatism, etc. For the proof of this statement, see section 29.2, where we show that wave aberrations are much more convenient than ray aberrations.

Consider an imaging system and a single object point O . Instead of investigating the rays itself emerging from the object point, we now look into the behavior of the wavefronts. In principle, there is no new information about the aberrations, as the wavefronts are strictly connected to the rays. The wavefront is defined as the locus of a constant optical path measured from the object point. As a reminder: the optical path is the geometrical distance multiplied by the local refractive index. It can be shown that the wavefronts are usually perpendicular to the rays. So in object space, where all rays pass through the object point, the wavefronts have the shape of concentric spheres, with the object point as center. In image space, when there are no aberrations at all, the wavefronts are again spheres, centered on the image point. In the case of aberrations, the wavefronts in the image space are no longer spheres, and the deviation to a suitable reference sphere is the wave aberration. There is some arbitrariness in choosing the reference sphere, but in practice this is not a problem. According to figure 29-7 the reference sphere is determined by its center O' , which is the assumed image point, and by its radius R . Usually the radius R is chosen so that the reference sphere contains the intersection point of the chief ray with the optical axis, point Q in figure 29-7, that is the location of the exit pupil. In the case of an infinite image, as the image point is at infinity, the radius R becomes infinite and the reference "sphere" becomes a plane, perpendicular to the chief ray.

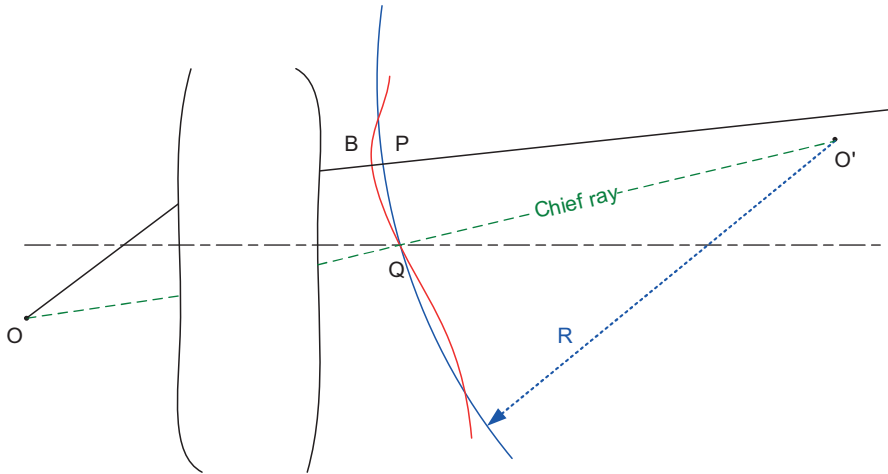


Figure 29-7: Wave aberrations.

We denote the optical path length from the object point O to a point B by $[OB]$. From figure 29-7 it can be seen that

$$[OB] = [OQ] \quad (29-2)$$

as B and Q are points on the same wavefront. So, the wave aberration W is defined as the optical path along the ray from the reference sphere to the wavefront:

$$W = [OB] - [OP] = [OQ] - [OP]. \quad (29-3)$$

An alternative designation for the wave aberration W is the optical path difference $OPD = W$.

The given procedure used to determine the reference sphere radius R , using the position of the exit pupil, breaks down when the imaging system is telecentric in the image space. In this case the chief rays are parallel to the optical axis and the exit pupil is at infinity. Nevertheless for telecentric systems it also makes sense to define wave aberrations relative to a reference sphere. Since for practical use the actual size of the reference sphere radius R is unimportant, the radius can be chosen to have any plausible size. The only thing that must be avoided is a reference sphere radius that is too small, which means that the reference sphere should be far enough from the focal region where the rays from the imaging ray bundle intersect each other. An example is shown in figure 29-8.

The graphical presentation of the wave aberration for the same example as shown in figure 29-2 is given in figure 29-9.

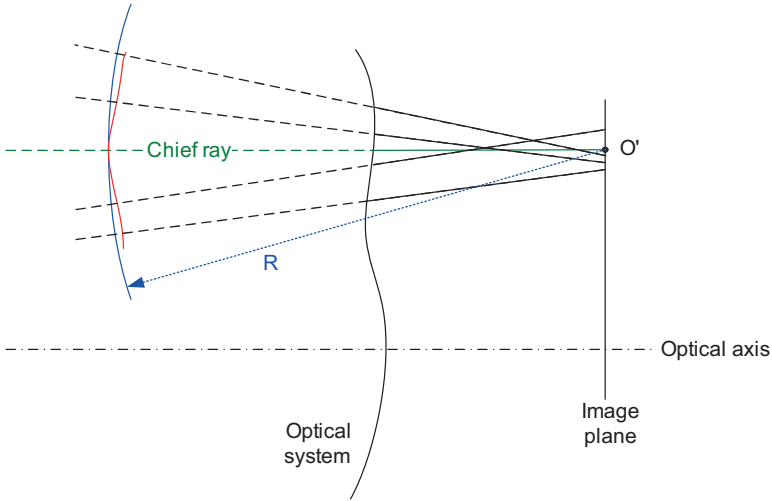


Figure 29-8: Wave aberration and reference sphere for a system which is telecentric in the image space. The aberrations are strongly exaggerated for clarity.

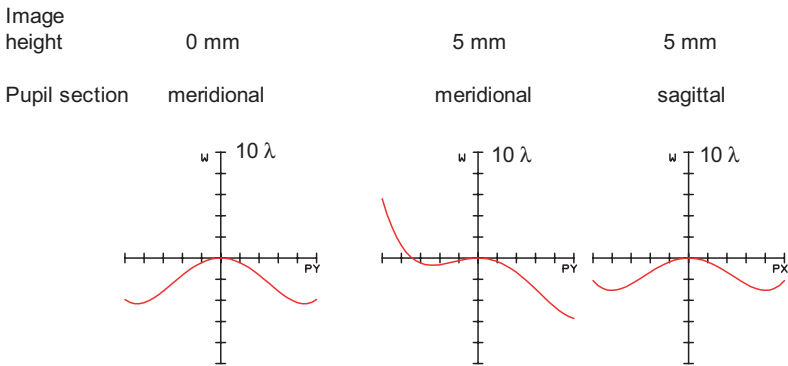


Figure 29-9: Wave aberrations for the system in figure 29-2.

29.2 Power Series Expansions

The wave aberration clearly depends on the chosen ray. When the wave aberration for all rays emerging from the object point O and passing through the exit pupil is to be described, the rays concerned are identified by their pupil coordinates x_p and y_p and the wave aberration $W(x_p, y_p)$ is a function of two variables. It must be pointed out that this function $W(x_p, y_p)$ only describes the aberrational behavior for the fixed chosen object point. To obtain complete information about the system aberrations

one has to consider not only one object point but the whole object field. If we characterize an object point by its object plane coordinates x and y , then the wave aberration becomes a function of four variables $W(x, y, x_p, y_p)$. This function of four variables is really necessary when optical systems without rotational symmetry are to be investigated [29-1]. In the more usual case, when the rotational symmetry of the optical system is given, there is redundancy in the four variables. In figure 29-10 we will see that, in this case, three variables are actually needed to describe the systems aberrations completely.

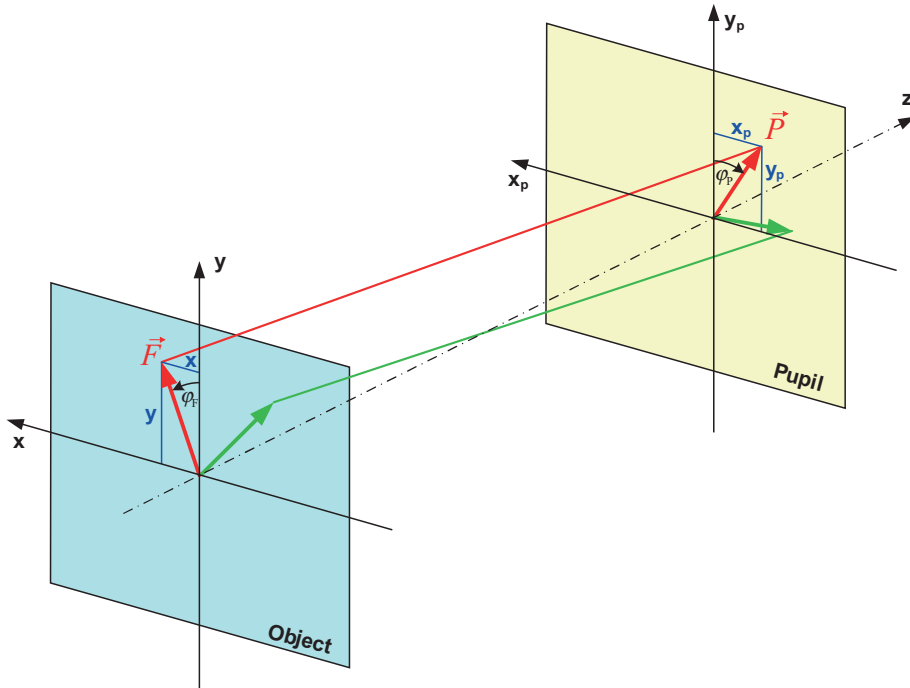


Figure 29-10: Equivalent rays by rotational symmetry.

In figure 29-10, from the rotationally symmetric optical system, only the object plane and the pupil plane are shown. We regard a single ray with object coordinates x and y and pupil coordinates x_p and y_p (the red ray in figure 29-10). For this ray the wave aberration is denoted by $W(x, y, x_p, y_p)$. Now, when this ray is rotated about the optical axis by an arbitrary amount (the green ray in figure 29-10), due to the rotational symmetry of the system, the wave aberration will not change. So, instead of describing the ray by the four variables x , y , x_p , and y_p , the Cartesian object and pupil coordinates, we have to find adapted variables which are invariant with respect to rotation about the optical axis. Let \vec{F} indicate the object vector from the object plane origin to the object point (x, y) and \vec{P} the pupil vector from the pupil plane origin to the pupil point (x_p, y_p) . Let the lengths of these vectors be F and P respectively, then such rotationally invariant variables are:

- the square of the object or field vector length $\vec{F} \cdot \vec{F} = F^2 = x^2 + y^2$
- the square of the pupil vector length $\vec{P} \cdot \vec{P} = P^2 = x_p^2 + y_p^2$
- the scalar product of the field vector and the pupil vector

$$\vec{P} \cdot \vec{F} = P \cdot F \cdot \cos(\varphi_F - \varphi_P) = x_p \cdot x + y_p \cdot y.$$

The last quantity $\vec{P} \cdot \vec{F}$ contains not only the lengths of the field and pupil vectors, but also the angle $\varphi_F - \varphi_P$ between these two vectors.

With these new variables the wave aberration for an arbitrary ray can be written as $W = W(\vec{P} \cdot \vec{P}, \vec{P} \cdot \vec{F}, \vec{F} \cdot \vec{F})$.

As the wave aberration is now dependent only on the length of the object vector, on the length of the pupil vector, and on the angle $\varphi_F - \varphi_P$ between the object and pupil vector, without loss of generality we can choose the object point on the y -axis. So, we set $x = 0$ and for the wave aberration we obtain

$$W = W(x_p^2 + y_p^2, y_p y, y^2). \quad (29-4)$$

In eq. (29-4) y represents the object field coordinate. In the case of an object at infinity y should be understood as the object field angle. The function (29-4) can now be expanded as a power series in three variables, arranging the terms in the proper order we get:

$$\begin{aligned} W &= W(x_p^2 + y_p^2, y_p y, y^2) \\ W &= a_0 + b_1(x_p^2 + y_p^2) + b_2 y y_p + b_3 y^2 \\ &+ c_1(x_p^2 + y_p^2)^2 + c_2 y y_p(x_p^2 + y_p^2) + c_3 y^2 y_p^2 + c_4 y^2(x_p^2 + y_p^2) + c_5 y^3 y_p + c_6 y^4 \\ &+ d_1(x_p^2 + y_p^2)^3 + d_2 y y_p(x_p^2 + y_p^2)^2 + d_3 y^2 y_p^2(x_p^2 + y_p^2) + d_4 y^2(x_p^2 + y_p^2)^2 + d_5 y^3 y_p(x_p^2 + y_p^2) \\ &+ d_6 y^3 y_p^3 + d_7 y^4 y_p^2 + d_8 y^4(x_p^2 + y_p^2) + d_9 y^5 y_p + d_{10} y^6 \\ &+ \text{terms with higher order.} \end{aligned} \quad (29-5)$$

The designation of the expansion coefficients $a_i, b_i, c_i, d_i, \dots$ corresponds to the sum of the powers in the variables. Going back to the original Cartesian coordinates y, x_p , and y_p , the total powers are zero for a_0 , two for b_i , four for c_i , and six for d_i . It will be shown that the b_i terms represent the primary aberrations and the c_i represent the secondary aberrations.

This power expansion is strictly mathematical. As the wave aberration is not an arbitrary function to be expanded, but is defined in such a way that it vanishes at the pupil center ($x_p = y_p = 0$), all coefficients in expressions with no dependence on the pupil coordinates must be zero. So, $a_0, b_3, c_6, d_{10}, \dots$ are set to zero.

This procedure of ignoring the terms which only depend on the field coordinate, when setting the coefficients b_3, c_6, d_{10}, \dots to zero, may appear a little strange. Usually mathematical results have some physical meaning. As we will see in section 29.5, these terms will actually have meaning when the pupil aberrations are addressed.

For some further discussion it is advantageous to express the pupil dependence of the wave aberration not only in Cartesian coordinates as in (29-5) but also in polar coordinates. As shown in figure 29-11 we define the length of the pupil vector r and the azimuth angle φ as

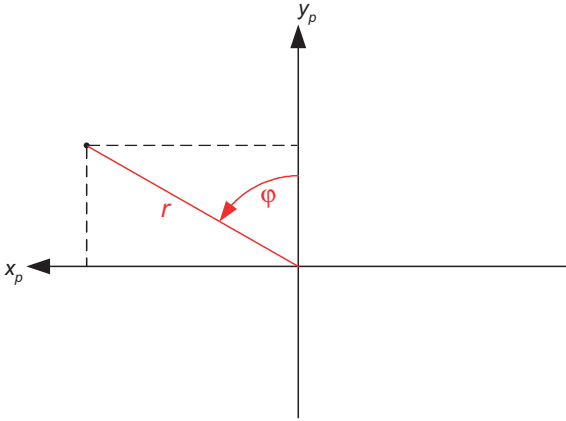


Figure 29-11: Polar coordinates for the pupil.

$$r = \sqrt{x_p^2 + y_p^2} \quad (29-6)$$

$$\begin{aligned} x_p &= r \cdot \sin \varphi \\ y_p &= r \cdot \cos \varphi. \end{aligned} \quad (29-7)$$

In terms of polar coordinates in the pupil the wave aberration W reads:

$$\begin{aligned} W &= b_1 r^2 + b_2 y r \cos \varphi \\ &+ c_1 r^4 + c_2 y r^3 \cos \varphi + c_3 y^2 r^2 \cos^2 \varphi + c_4 y^2 r^2 + c_5 y^3 r \cos \varphi \\ &+ d_1 r^6 + d_2 y r^5 \cos \varphi + d_3 y^2 r^4 \cos^2 \varphi + d_4 y^2 r^4 + d_5 y^3 r^3 \cos \varphi \\ &+ d_6 y^3 r^3 \cos^3 \varphi + d_7 y^4 r^2 \cos^2 \varphi + d_8 y^4 r^2 + d_9 y^5 r \cos \varphi \\ &+ \text{terms with higher order.} \end{aligned} \quad (29-8)$$

According to eqs (11-12) and (11-13) the transverse ray aberrations ${}^o x'$ and ${}^o y'$ can be calculated to a good approximation by differentiating the wave aberration with respect to x_p and y_p , respectively:

$${}^o x' = -\frac{R}{n'} \cdot \frac{\partial W}{\partial x_p}, \quad (29-9)$$

$${}^o y' = -\frac{R}{n'} \cdot \frac{\partial W}{\partial y_p}, \quad (29-10)$$

where R is the radius of the reference sphere and n' is the refractive index in image space. The approximation inherent in eqs (29-9) and (29-10) is adequate for deriving the primary transverse aberrations. In the case of vanishing primary aberrations, eqs (29-9) and (29-10) can also be used to derive the secondary transverse aberrations. So, applying these equations to (29-5) the power series expansion of the transverse ray aberrations can be derived:

$${}^{\circ} x' = -\frac{R}{n'} \left\{ 2b_1 x_p + 4c_1 x_p (x_p^2 + y_p^2) + 2c_2 y x_p y_p + 2c_4 y^2 x_p + \dots \right\} \quad (29-11)$$

$${}^{\circ} y' = -\frac{R}{n'} \left\{ 2b_1 y_p + b_2 y + 4c_1 y_p (x_p^2 + y_p^2) + c_2 y (x_p^2 + 3y_p^2) + 2(c_3 + c_4) y^2 y_p + c_5 y^3 + \dots \right\}. \quad (29-12)$$

With the help of eqs (29-6) and (29-7) and using the identities (29-13) and (29-14)

$$2 \cos^2 \varphi = 1 + \cos 2\varphi, \quad (29-13)$$

$$2 \sin \varphi \cos \varphi = \sin 2\varphi, \quad (29-14)$$

the transverse aberrations ${}^{\circ} x'$ and ${}^{\circ} y'$ can be expressed in polar coordinates in the pupil:

$${}^{\circ} x' = -\frac{R}{n'} \left\{ 2b_1 r \sin \varphi + 4c_1 r^3 \sin \varphi + c_2 y r^2 \sin 2\varphi + 2c_4 y^2 r \sin \varphi + \dots \right\} \quad (29-15)$$

$${}^{\circ} y' = -\frac{R}{n'} \left\{ 2b_1 r \cos \varphi + b_2 y + 4c_1 r^3 \cos \varphi + c_2 y r^2 (2 + \cos 2\varphi) + 2(c_3 + c_4) y^2 r \cos \varphi + c_5 y^3 + \dots \right\}. \quad (29-16)$$

In the power series expansions eqs (29-5), (29-8), (29-11), (29-12), (29-15), (29-16) the single summands are clearly distinguished by the different powers of the field variable y and of the pupil variables x_p and y_p for Cartesian coordinates in the pupil and r for polar coordinates in the pupil. The sum of the powers for the field and aperture (pupil) variables give the order of the single aberration, while the distribution of the power sum between the field and aperture variables determines the type of different aberrations.

According to eq. (29-5) the lowest power sum is two in the terms with the coefficients b_1 and b_2 (a_0 and b_3 are identified to be zero). So, b_1 and b_2 seem to represent the lowest order and therefore usually seem to be the most important aberrations. However, this is not the case, as usually these two terms are not regarded as aberrations at all. The terms originate from the choice of the reference image point, which is the center of the reference sphere. An axial displacement of the reference image point results in a term $b_1(x_p^2 + y_p^2) = b_1 r^2$. The related wave aberration, as discussed in section 11.5.3, is that of a defocus and can be canceled by proper focusing. A displacement of the reference image point perpendicular to the axis, which means the choice of the image height, results in a term $b_2 y y_p = b_2 y r \cos \varphi$. This corresponds to a tilt of the reference sphere (see section 11.5.2). As this term is linear in y , the field variable, it can be canceled by the proper choice of image scale, or magnification. As soon as the chromatic aberrations are also addressed (next paragraph), the coefficients b_1 and b_2 will become a new significance.

The terms with the coefficients c_1 to c_6 all have the power sum four. As c_6 is identified to be zero there are five terms which represent the five monochromatic primary aberrations. These are: spherical aberration, coma, astigmatism, (sagittal) field curvature, and distortion. These aberrations, consisting of the terms with the lowest powers which are regarded as aberrations, are also called third-order aberrations and also Seidel aberrations. The designation “third-order” relates to the power sum in the expressions (29-11) to (29-16) for the transverse aberrations. As the transverse aberrations are derived from the wave aberrations by differentiation with respect to the pupil coordinate, see eqs (29-9) and (29-10), the power sum of a transverse aberration term is always one less than the power sum in the corresponding wave aberration term. The transverse aberrations have been the basis of Seidel’s investigations, and so, for historical reasons, these monochromatic primary aberrations are called third-order aberrations although considered as wave aberrations, the order is four.

29.3

Chromatic Aberrations

Now we will have a preliminary look at chromatic aberrations, which result from a change in the refractive indices when the wavelength is changed. Clearly, all the monochromatic aberrations will have their chromatic variations. Informally, one speaks about colored coma, colored astigmatism, etc. For the chromatic variation in the spherical aberration there is a separate designation, known as spherochromatism or in German “Gaussfehler”. But when varying the refractive indices the paraxial or Gaussian quantities, such as the axial image position and the image size, also have their chromatic variations, and these chromatic variations turn out to represent the primary chromatic aberrations. Axial chromatic aberration or in short axial color must be regarded as the chromatic variation of image position or chromatic focal shift. Lateral chromatic aberration or a chromatic difference in the magnification or short lateral color is the chromatic variation of the image size.

It is accepted procedure to use the wave aberration description not only for monochromatic but also for chromatic aberrations. In order to do this, some conditions must be met. For monochromatic aberrations the wavefront, which passes through the center of the exit pupil, is compared with a reference sphere which represents the ideal wavefront. When defining chromatic aberrations as wave aberrations two wavefronts, which are based on different wavelengths, are compared. This seems slightly artificial, but it turns out to be very helpful to have the same basic understanding of monochromatic as well as chromatic aberrations.

So, axial color as a wave aberration, see figure 29-12 and eq. (29-18), has the same dependence on the pupil and field coordinates as the term containing b_1 representing the monochromatic focal shift in eqs (29-5) and (29-8). For lateral color the wave aberration dependence on aperture and field corresponds to the term containing b_2 representing the monochromatic change in the image size.

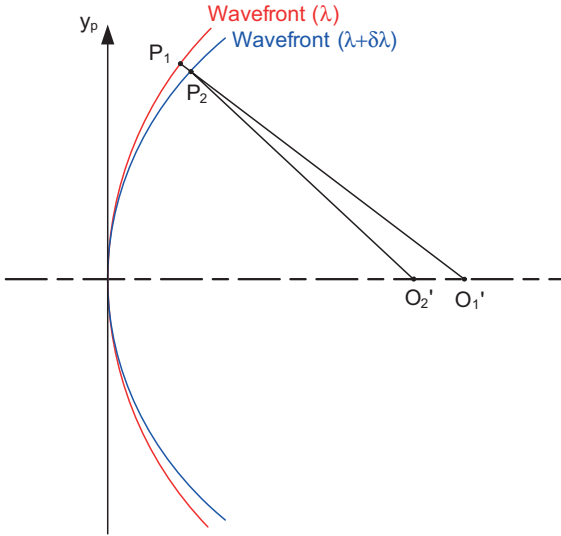


Figure 29-12: Axial color as a wave aberration.

In figure 29-12 the wavefront based on the wavelength $\check{\omega}$ can be regarded as the reference sphere and the chromatic change in the wave aberration $\check{\omega}W$ when the wavelength is changed to $\check{\omega} + \check{\omega}\check{\delta}$ is the optical path from P_1 to P_2

$$\check{\omega}W = [P_1, P_2]. \tag{29-17}$$

When calculating the optical path $[P_1, P_2]$, which is the geometrical path multiplied by the refractive index of the image space medium, it is not evident whether the index n' (for wavelength $\check{\omega}$) or the index $n' + \check{\omega}n'$ (for wavelength $\check{\omega} + \check{\omega}\check{\delta}$) should be applied. In this context the wavelength change $\check{\omega}\check{\delta}$ is assumed to be small, so that when calculating the optical path based on the geometrical path, n' can be used and $\check{\omega}n'$ can be neglected.

In figure 29-12 the wavefront ($\check{\omega}$) can be seen as a sphere with its center at O_1' then the wavefront ($\check{\omega} + \check{\omega}\check{\delta}$) can be assumed to be a sphere with a different radius and center at O_2' . The distance from O_1' to O_2' is the axial color as a longitudinal aberration. For the axial color as a wave aberration, from the geometry in figure 29-12, it can be derived, to a first approximation, that

$$\check{\omega}W = \check{b}_1 \cdot (x_p^2 + y_p^2) = \check{b}_1 \cdot r^2. \tag{29-18}$$

The coefficient \check{b}_1 is different from the coefficient b_1 in eqs (29-5) and (29-8), \check{b}_1 describes the amount of axial chromatic aberration.

Concerning lateral color there are similar considerations. When the wavefront for $\check{\omega}$ is regarded as the reference sphere then to a good approximation the wavefront for $\check{\omega} + \check{\omega}\check{\delta}$ is a sphere with almost the same radius and it is tilted with respect to the reference sphere, as shown in figure 29-13.

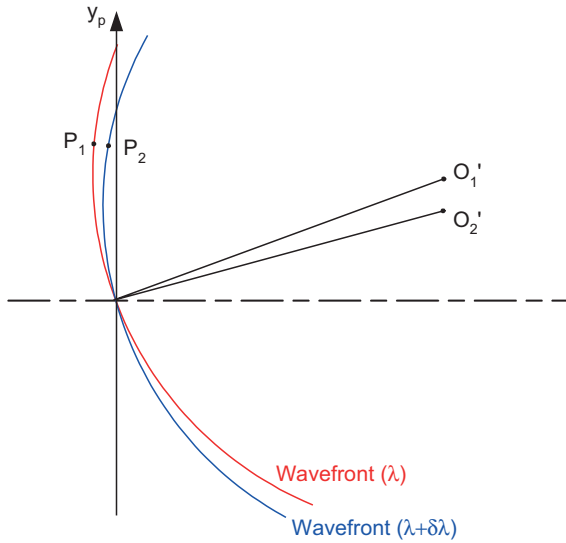


Figure 29-13: Lateral color as a wave aberration.

The distance from O_1' to O_2' represents the lateral color as a transverse aberration.

The dependence of the corresponding wave aberration on the aperture and field coordinates is given by

$$\tilde{\omega}W = \tilde{b}_2 \cdot \gamma\gamma_p = \tilde{b}_2 \cdot \gamma r \cos \varphi. \quad (29-19)$$

The coefficient \tilde{b}_2 is different from the coefficient b_2 in eqs (29-5) and (29-8), \tilde{b}_2 describes the amount of lateral chromatic aberration.

During these considerations of the primary chromatic aberrations we have assumed, that only the type of aberration which we have just considered, appears in the system and that all other aberrations are set to zero.

From eqs (29-18) and (29-19) it can be seen that the power sum of the pupil and field coordinates is two. So the primary chromatic aberrations are second-order wave aberrations (and first-order transverse aberrations).

29.4

Primary Aberrations

29.4.1

Aperture and Field Dependence

The field and the aperture dependence of the primary aberrations are listed in table 29-1. In this table the Seidel sums, which will be defined in (29-23) – (29-29), are already included.

Table 29-1: Powers of the primary wave aberrations. Aperture variables are the pupil coordinates x_p , y_p and r , respectively; the field variable is the field coordinate γ .

Aberration	Coefficient	Seidel sum	Wave aberration		Transverse aberration		Longitudinal aberration	
			Aperture	Field	Aperture	Field	Aperture	Field
Spherical aberr.	c_1	S_I	4	0	3	0	2	0
Coma	c_2	S_{II}	3	1	2	1	1	1
Astigmatism	c_3	S_{III}	2	2	1	2	0	2
Field curvatures (sagittal)	c_4	(Petzval) S_{IV}	2	2	1	2	0	2
Distortion	c_5	S_V	1	3	0	3	–	–
Axial color	\tilde{b}_1	C_I	2	0	1	0	0	0
Lateral color	\tilde{b}_2	C_{II}	1	1	0	1	–	–

Table 29-1 also includes the longitudinal aberrations. As can be seen from this table the powers for the aperture variables decrease by one when going from wave aberration to transverse aberration as well as from transverse aberration to longitudinal aberration, see section 11.5. On the other hand, the powers for the field variable do not depend on the type of aberration description: wave, transverse or longitudinal.

In the following sections we will discuss the single primary aberrations in more detail.

Based on eqs (29-8), (29-18) and (29-19) the typical shape of the primary wave aberrations can be calculated. Figure 29-14 shows the monochromatic and figure 29-15 the chromatic primary aberrations in a three-dimensional representation.

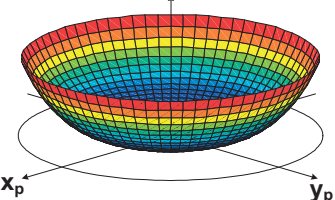
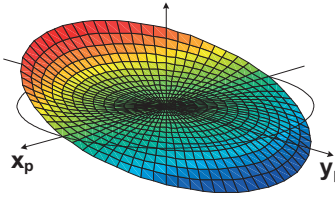
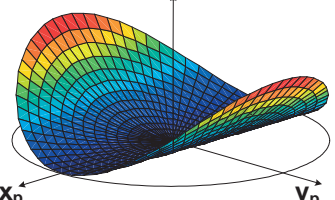
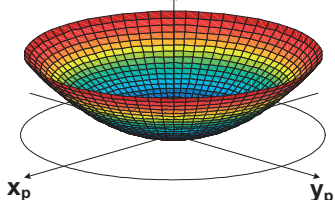
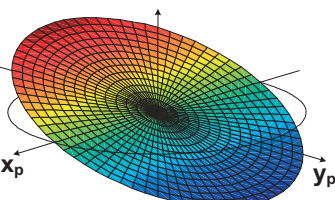
Primary monochromatic wave aberrations	
Spherical aberration	Coma
$W = c_1 \cdot (x_p^2 + y_p^2)^2 = c_1 \cdot r^4$	$W = c_2 \cdot yy_p \cdot (x_p^2 + y_p^2) = c_2 \cdot yr^3 \cos \varphi$
	
Astigmatism	Field curvature (sagittal)
$W = c_3 \cdot y^2 y_p^2 = c_3 \cdot y^2 r^2 \cos^2 \varphi$	$W = c_4 \cdot y^2 \cdot (x_p^2 + y_p^2) = c_4 \cdot y^2 r^2$
	
Distortion	
$W = c_5 \cdot y^3 y_p = c_5 \cdot y^3 r \cos \varphi$	
	

Figure 29-14: Primary monochromatic wave aberrations.

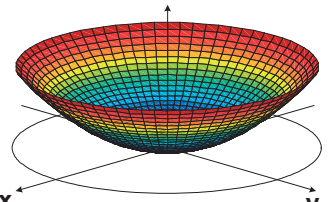
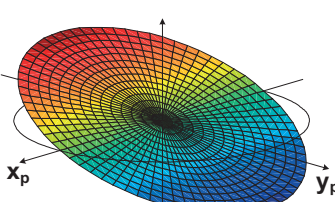
Primary chromatic wave aberrations	
Axial color	Lateral color
$\delta W = \tilde{b}_1 \cdot (x_p^2 + y_p^2) = \tilde{b}_1 \cdot r^2$	$\delta W = \tilde{b}_2 \cdot yy_p = \tilde{b}_2 \cdot yr \cos \varphi$
	

Figure 29-15: Primary chromatic wave aberrations.

29.4.2

Symmetry and Periodicity Properties

Often it is helpful to investigate symmetry and periodicity properties of the single aberration type when analyzing the total aberrations of an optical system. This may help to identify the dominating aberrations. For these considerations the field dependence of the aberrations does not play a role. In figure 29-16 (primary monochromatic aberrations) and figure 29-17 (primary chromatic aberrations) symmetry and periodicity is shown for the wave aberrations as well as for the transverse aberrations. Three different types of symmetry occur:

1. Rotational symmetry with reference to an axis and to a point, respectively. Spherical aberration, field curvature and axial chromatic aberration all exhibit this type of symmetry.
2. Mirror symmetry with reference to two planes and to two straight lines, respectively. This produces the appearance of astigmatism.
3. Mirror symmetry with reference to one plane and to one straight line, respectively. Coma and lateral chromatic aberration possess this type of symmetry.

The periodicity is defined as the period of the aberration for a single ray, while the pupil coordinates of this ray move through a concentric circle in the pupil. That means $r = \text{constant}$ and $\varphi = 0 - 2\pi$, keeping the field variable constant. It is important to note that wave aberrations and transverse aberrations show the same symmetry but different periodicity. For our treatment of the transverse aberrations in figures 29-16 and 29-17 we use two circles in the pupil, first with a radius r (for instance the maximum pupil radius) and second with radius $r/2$. In figures 29-16 and 29-17 the spot diagrams for these rays are shown. The proportional relationships given for the transverse aberrations and the corresponding figures are generated from eqs (29-15) and (29-16) by suppressing the constants.

It should be mentioned that the identification of aberrations by this method, which is based on symmetry properties, is not quite trouble-free. With higher order aberrations new types of aberration show up and exhibit symmetry properties which do not correspond completely to the above results for the primary aberrations. For example when regarding fifth-order aberrations (see section 29.6.1) the so-called oblique spherical aberration exhibits mirror symmetry with reference to two planes, see figure 29-46. So, the fifth-order oblique spherical aberration has the same type of symmetry as third-order astigmatism. But this inconsistency does not really cause a problem if kept in mind.

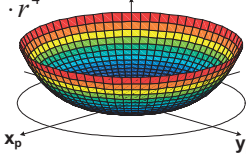
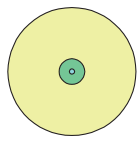
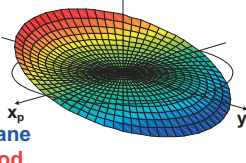
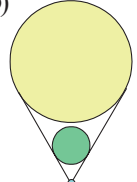
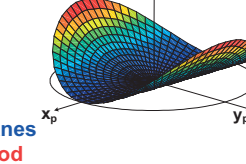

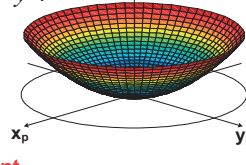
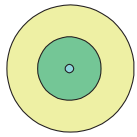
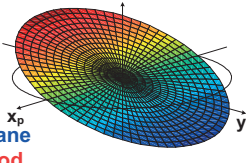

Type	Wave aberration	Geometrical spot
Spherical aberration Symmetry to Periodicity	$W = c_1 \cdot r^4$  axis constant	$\Delta x' \propto c_1 \cdot r^3 \sin \varphi$ $\Delta y' \propto c_1 \cdot r^3 \cos \varphi$ point 1 period 
Coma Symmetry to Periodicity	$W = c_2 \cdot y r^3 \cos \varphi$  one plane 1 period	$\Delta y' \propto c_2 \cdot y r^2 \cdot (2 + \cos 2\varphi)$ $\Delta x' \propto c_2 \cdot y r^2 \sin 2\varphi$ one straight line 2 periods 
Astigmatism Symmetry to Periodicity	$W = c_3 \cdot y^2 r^2 \cos^2 \varphi$  two planes 2 period	$\Delta x' = 0$ $\Delta y' \propto c_3 \cdot y^2 r \cos \varphi$ two straight lines 1 period 
Field curvature (sagittal) Symmetry to Periodicity	$W = c_4 \cdot y^2 r^2$  axis constant	$\Delta x' \propto c_4 \cdot y^2 r \sin \varphi$ $\Delta y' \propto c_4 \cdot y^2 r \cos \varphi$ point 1 period 
Distortion Symmetry to Periodicity	$W = c_5 \cdot y^3 r \cos \varphi$  one plane 1 period	$\Delta x' = 0$ $\Delta y' \propto c_5 \cdot y^3$ one straight line constant 

Figure 29-16: Symmetry and periodicity of monochromatic primary aberrations.

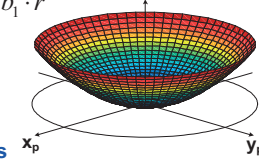
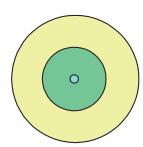
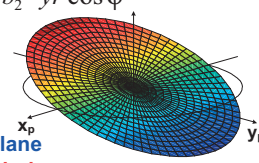

Type	Wave aberration	Geometrical spot
<p>Axial color</p> <p>Symmetry to Periodicity</p>	$\delta W = \tilde{b}_1 \cdot r^2$  <p>axis constant</p>	$\Delta x' \propto \tilde{b}_1 \cdot r \sin \varphi$ $\Delta y' \propto \tilde{b}_1 \cdot r \cos \varphi$  <p>point 1 period</p>
<p>Lateral color</p> <p>Symmetry to Periodicity</p>	$\delta W = \tilde{b}_2 \cdot y r \cos \varphi$  <p>one plane 1 period</p>	$\Delta x' = 0$ $\Delta y' \propto \tilde{b}_2 \cdot y$  <p>one straight line constant</p>

Figure 29-17: Symmetry and periodicity of chromatic primary aberrations.

29.4.3

Presentation of Aberrations and their Impact on Image Quality

Aberrations are represented in different ways. In the figures 29-18 – 29-21 and 29-27 – 29-29 for the primary aberrations we show the typical diagrams for wave aberrations, for transverse aberrations and, if it makes sense, for longitudinal aberrations. In addition, the corresponding through-focus spot diagrams are shown and the impact on image quality is illustrated by giving the modulation transfer function MTF versus frequency as well as versus the defocus (see section 30.9). To allow easy comparison for the different aberration types, constant preconditions are used.

- The numerical aperture is $NA = 0.25$.
- The amount of single aberrations is chosen to result in a y -direction spot diameter of 0.01 mm.
- The wavelength is 500 nm.
- The scale used in the diagrams is not changed except for the scale for the focus shift in spot and MTF diagrams for astigmatism, figure 29-27 and 29-28, and for field curvature, figure 29-29.

It should be noted that, in real optical systems, the different aberration types usually appear simultaneously as has already been shown in figure 29-5. In figures 29-18 – 29-21 and 29-27 – 29-29, the aberrations are displayed separately in order to show their characteristics. Only when the typical appearance of the different aberrations is known in an aberration diagram of a real system, can the dominating aberrations be detected.

The following discussion on the different types of aberration curves can be compared with eq. (29-5) for the wave aberrations and with eqs (29-11) and (29-12) for the transverse aberrations.

Spherical aberration

As shown in figure 29-18, for spherical aberration, the wave aberration curves are of fourth power, so the transverse aberration curves are third-power polynomials. Spherical aberration is the only aberration for which the aberrations from the tangential and sagittal pupil section are identical. In the given example the spherical aberration is under-corrected, which means that the upper marginal ray hits the axis in front of the paraxial image and so the intersection height in the paraxial image plane is negative. Spherical aberration with the opposite sign is called over-corrected. From the through-focus analysis in the spot diagrams, as well as in the MTF, it can be seen that the best image quality is obtained not at the paraxial image plane but at an inward defocused position. Figure 29-19 represents the same spherical aberration but focused to get the maximum MTF. Because of the defocusing in the

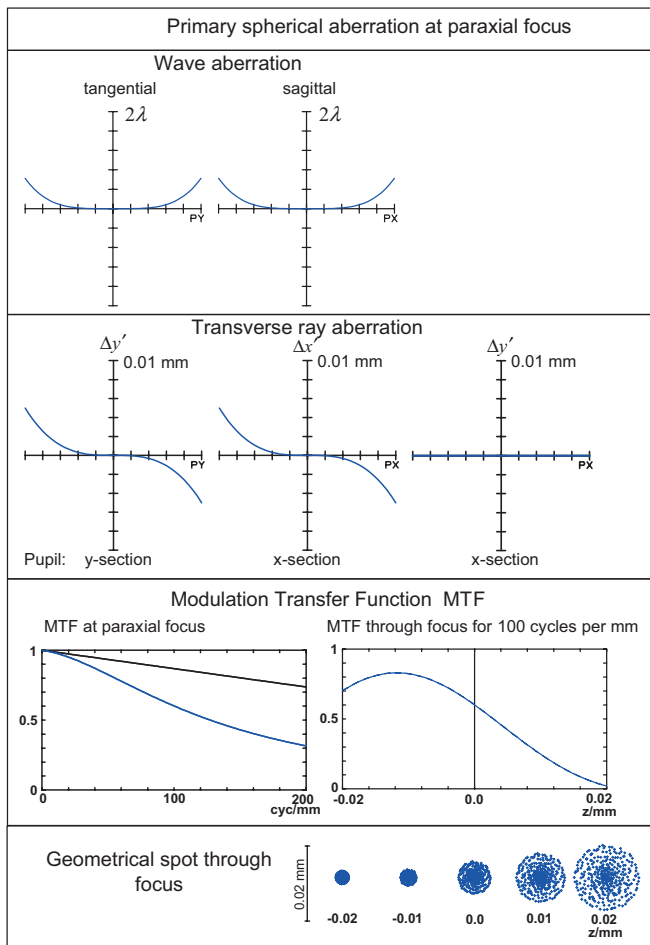


Figure 29-18: Spherical aberration at paraxial focus.

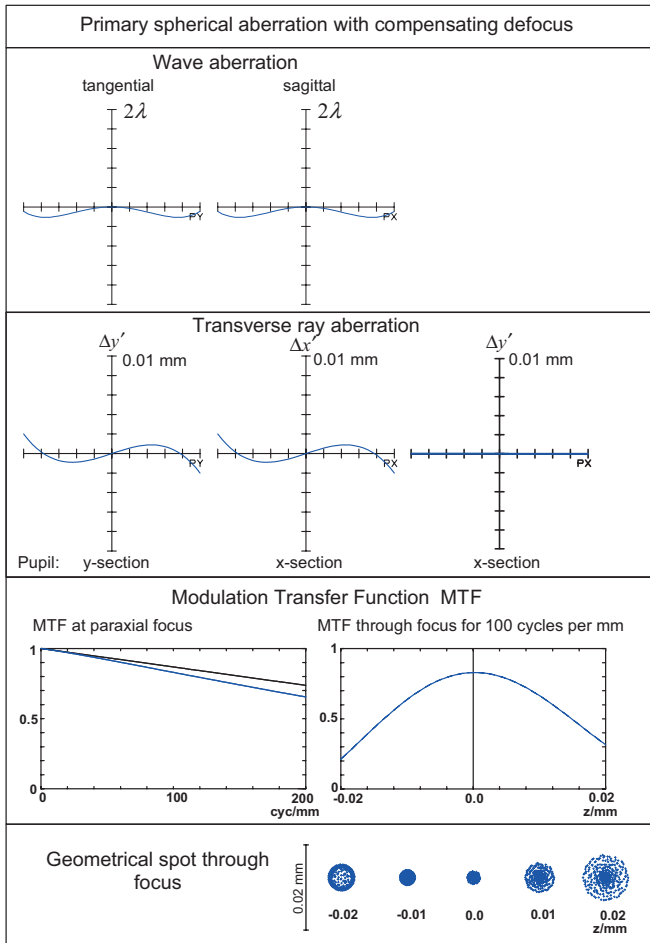


Figure 29-19: Spherical aberration of figure 29-18 with compensating defocus.

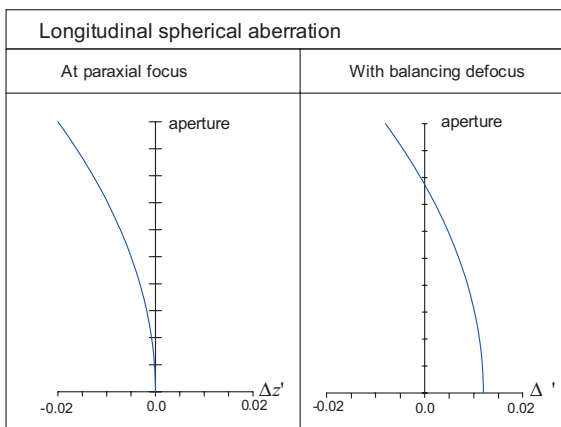


Figure 29-20: Longitudinal aberration according to figures 29-18 and 29-19.

wave aberration curves, a second-power term is added and in the transverse aberration curves this is a linear term. These additional terms balance the aberrations as well as possible. In figure 29-20 the corresponding longitudinal aberrations are presented. In this diagram the amount of balancing defocus can be clearly seen.

Coma

For primary coma it is typical that the wave aberration is zero for the sagittal section, see figure 29-21. In the tangential section the wave aberration is a third-power polynomial, so the transverse aberration is a second-power polynomial. A characteristic

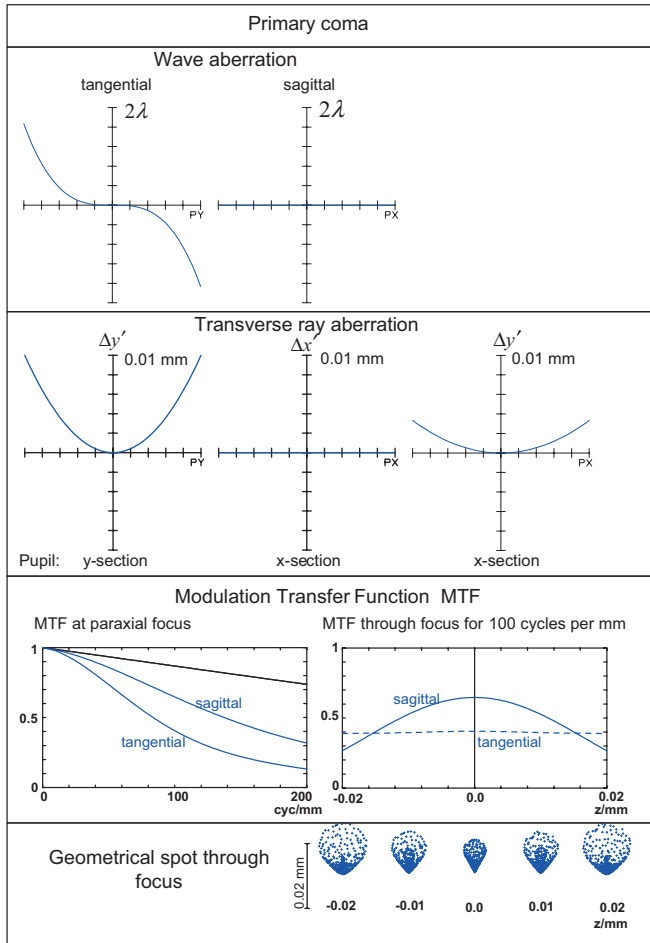


Figure 29-21: Coma.

property of the coma can be seen in the transverse aberration of rays in the sagittal pupil section where the x -component is zero while the y -component, representing the sagittal coma (the coma from the sagittal pupil section), is similar to the y -component from the tangential pupil section, representing the tangential coma. Actually the sagittal coma is one-third of the tangential coma. It is interesting to note that coma is the only aberration which possesses contributions in the y -component from the sagittal section. This is an important feature when the transverse aberrations of a real system which incorporates different aberration types are analyzed. The y -component from the sagittal section always shows the pure sagittal coma and no other aberrations. On the other hand, in the usual representation of the transverse aberrations from the sagittal pupil section, only the x -component is shown while the y -component is disabled. From through-focus analysis with help of spot diagrams and also the through-focus MTF, it is obvious that any defocus cannot improve the imaging. A further interesting property of coma can be seen from the through-focus MTF. In spite of the different sizes of the through-focus spot diagrams, the tangential MTF (resolution in the y -direction) is almost constant over a certain defocus range, while the sagittal MTF (resolution in the x -direction) shows a clear maximum at the focus with the smallest spot diagram.

As can be seen in figures 29-16 and 29-21, coma has the most complex spot diagram. Figure 29-22 demonstrates the geometrical coma spot in a three-dimensional arrangement. The quantities “tangential coma” (from the chief ray intersection point C to the tangential ray intersection point T) and “sagittal coma” (from the chief ray intersection point C to the sagittal ray intersection point S) are illustrated in this figure.

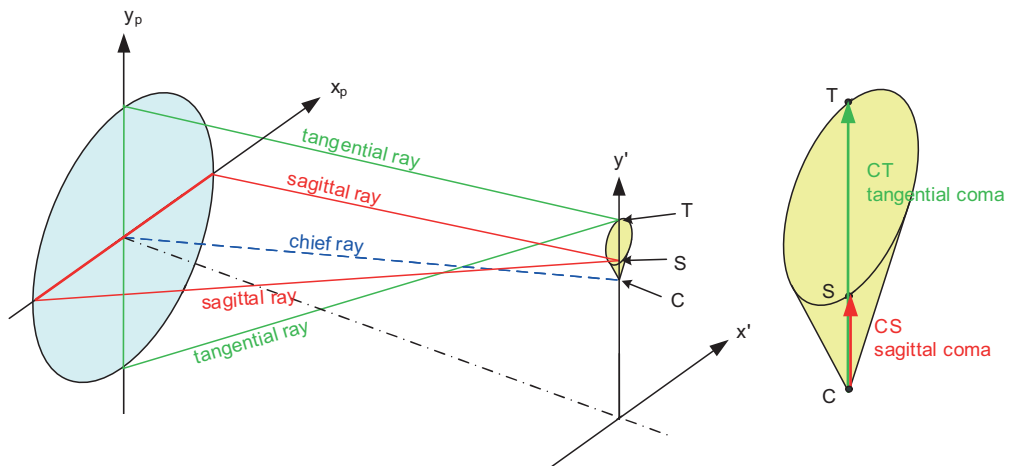


Figure 29-22: Geometric spot for primary coma.

Due to the term 2ϕ in eqs (29-15) and (29-16) for the transverse aberrations x' and y' rays running through a concentric circle with radius r in the pupil, form a circle in the image with a radius proportional to r^2 and the image circle is decentered with respect to the chief ray by a factor which is also proportional to r^2 . Figure 29-23 is a schematic illustration of the ray path in a simple two-dimensional configuration as calculated from eqs (29-15) and (29-16).

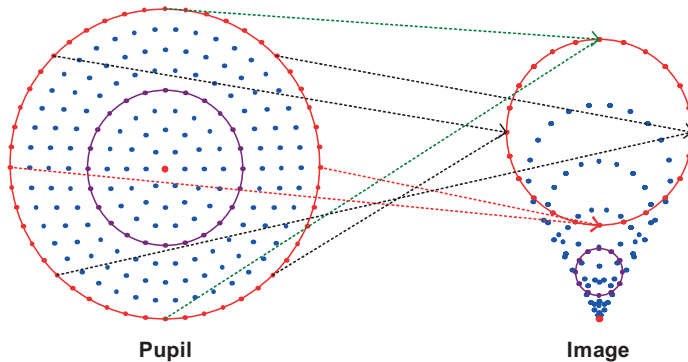


Figure 29-23: Schematic ray path for primary coma.

Depending on the sign of the coefficient c_2 in eqs (29-5), (29-8), and (29-11) – (29-16), the coma is designated either as an outer or as an inner coma. In figure 29-22 an outer coma is shown. All rays have image intersection points which have a larger distance to the image center than does the intersection point of the chief ray. Loosely speaking, the coma is driven outwards. In the opposite case when the aberration is driven inwards, it is designated as inner coma. In figure 29-24 outer and inner coma are shown schematically in one diagram.

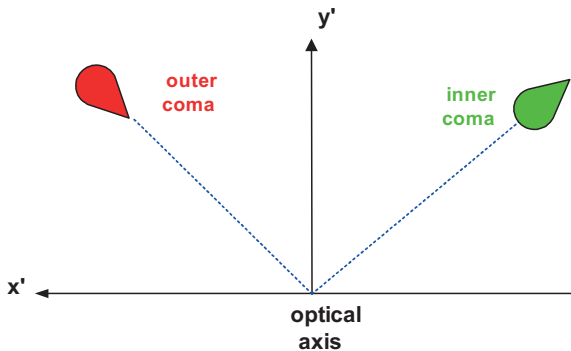


Figure 29-24: Outer and inner coma.

Figure 29-25 represents an example of outer coma but in this figure the purely geometrical spot is not shown but instead the physical intensity distribution is given. The rotational symmetry with respect to the optical axis, the image center, is evident.

Coma as ray aberration is usually understood as transverse aberration. Nevertheless, sometimes it may be convenient to interpret coma as a longitudinal aberration:

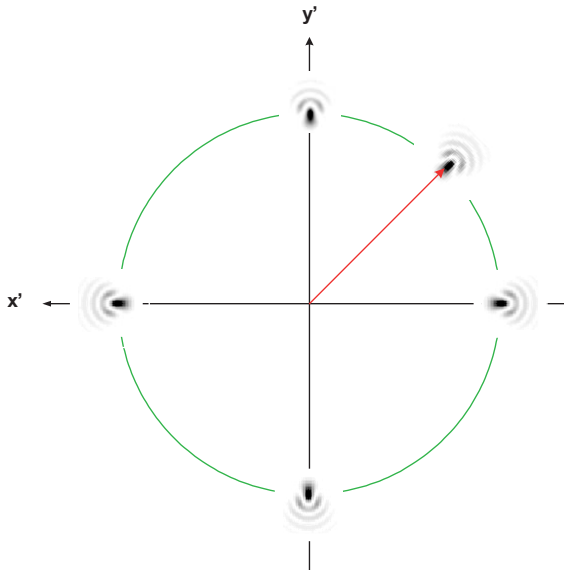


Figure 29-25: Symmetry and intensity of coma.

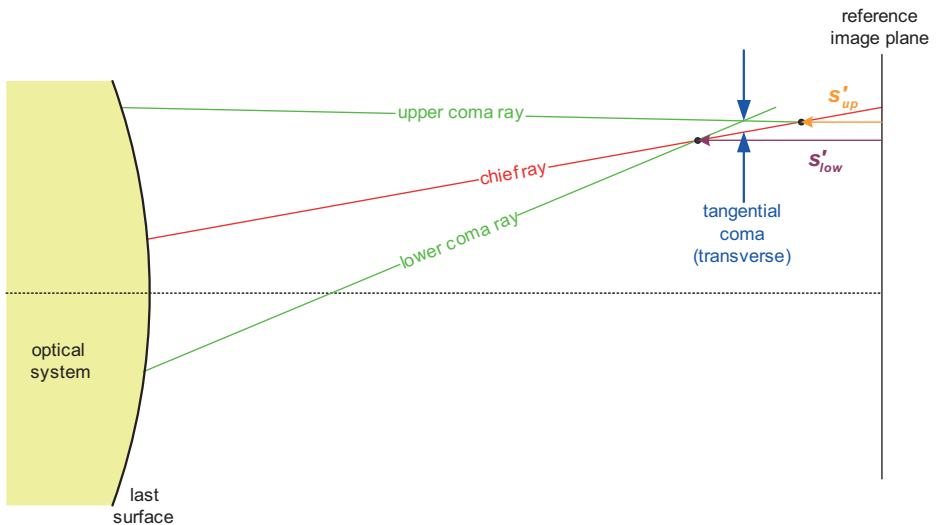


Figure 29-26: Definition of longitudinal coma.

$$s'_{coma} = \frac{s'_{up} - s'_{low}}{2} \tag{29-20}$$

Figure 29-26 is a sketch illustrating the definition for longitudinal coma, which is determined by the intersection points of the upper and lower coma ray with the chief ray. The upper and lower coma rays are also called meridional or tangential coma rays.

Astigmatism and Field Curvature

As shown in figures 29-27 and 29-28, astigmatism corresponds to a difference in the focus position for the sagittal and tangential pupil section. In figure 29-27 the focus

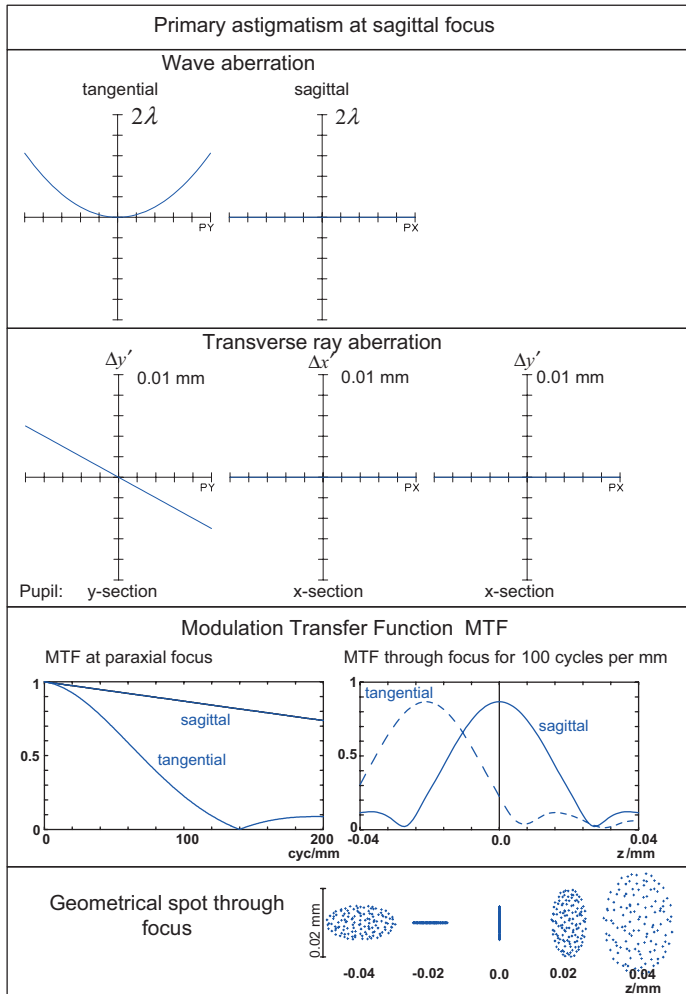


Figure 29-27: Astigmatism at sagittal focus.

is set to the sagittal image, as it is calculated in eqs (29-15) and (29-16). In figure 29-28 the mid-position between the sagittal and tangential focus is chosen as the receiving image plane. In general this is of course only possible for a chosen image height. Figure 29-29 summarizes three cases: the paraxial image plane, the image plane at the sagittal focus for maximum field, and the image plane at medial focus for maximum field.

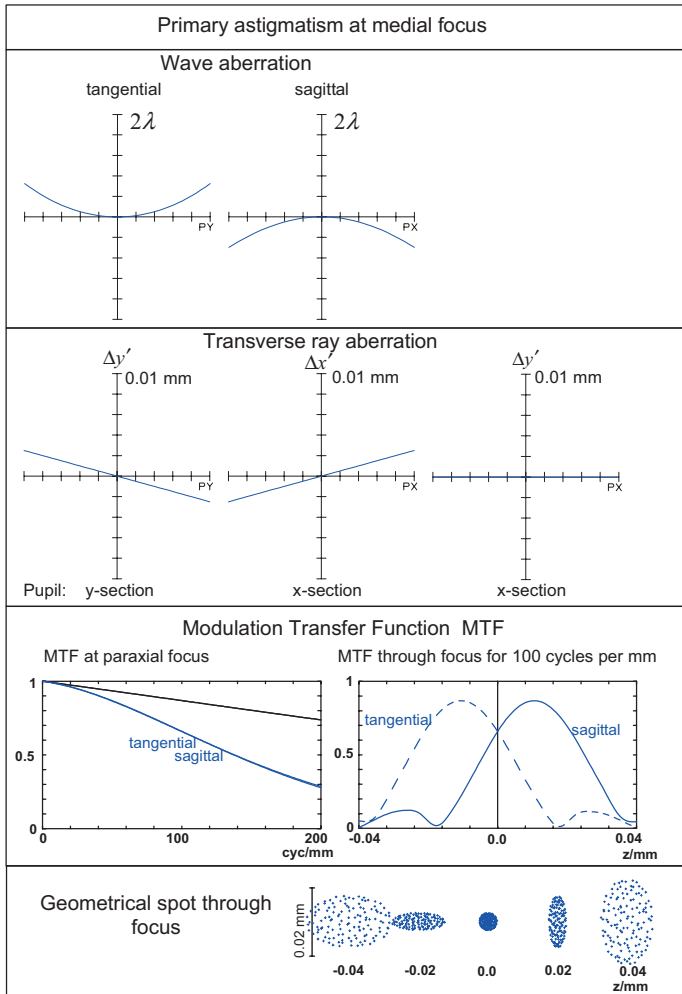


Figure 29-28: Astigmatism of figure 29-27 with compensating defocus.

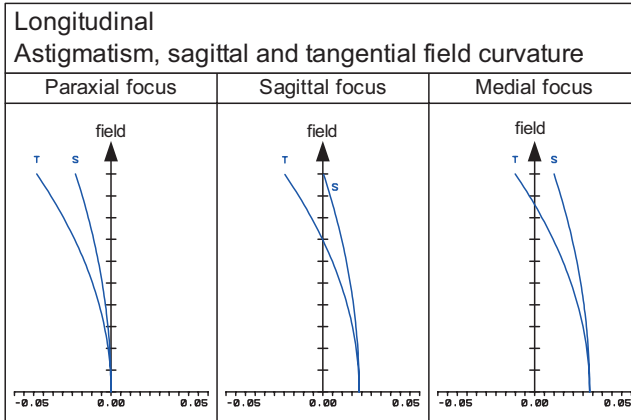


Figure 29-29: Sagittal and tangential field curvature as longitudinal aberration.

As shown in figure 29-29, the field dependence of the longitudinal sagittal and tangential field curvature is of second power. The astigmatism is defined as the longitudinal distance (parallel to the z -axis of the optical system) of the tangential field measured from the sagittal field. A further discussion of astigmatism, sagittal and tangential field curvatures, and Petzval curvature follows in the next paragraph, when the Seidel sums are introduced.

29.4.4

Calculation of the Seidel Sums

According to table 29-1 the coefficients c_1 to c_5 represent the amount of primary aberrations. These coefficients depend on the system parameters (radii, separations, materials) and in addition they may depend on the position of the object as well as the position of the aperture stop. There are different sets of formulae to calculate the primary aberration coefficients for a given system. We will use the simplest formula set which also provide a clear insight into how the single aberrations depend on the different parameters. But instead of calculating the coefficients c_1 to c_5 used in the mathematical derivation of the power series, we will define a new set of coefficients, the so-called Seidel sums: S_I , S_{II} , S_{III} , S_{IV} , and S_V which also represent the amount of primary aberrations. The relation between the c_i -coefficients and the S -coefficients will be given later. To make the equations as simple as possible the following paraxial quantities are defined for each surface of the system:

$$A = n(hc + u) = n \cdot i = n' \cdot i' \quad (29-21)$$

$$\bar{A} = n(\bar{h}c + \bar{u}) = n \cdot \bar{i} = n' \cdot \bar{i}' \quad (29-22)$$

All quantities with a bar refer to the chief ray, the corresponding quantities without the bar refer to the marginal ray. As usual n is the refractive index, i is the paraxial

incidence angle, u is the paraxial convergence angle, h is the paraxial incidence height at the surface, and c is the surface curvature. In figure 29-30 these paraxial quantities are shown for the marginal ray. Actually all these quantities have an index which indicates the surface number, but here the index is suppressed. So A and \bar{A} are the refraction invariants at the chosen surface.

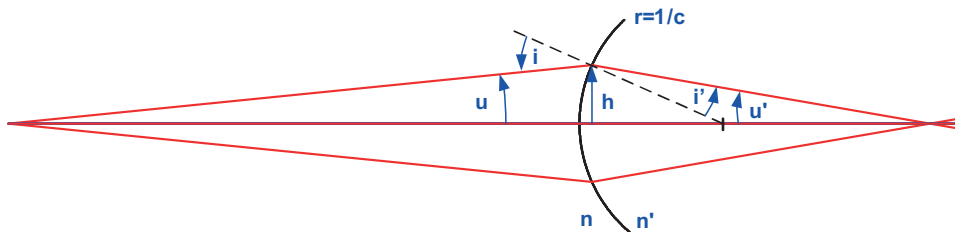


Figure 29-30: Quantities used to calculate the refraction invariant.

The increment on refraction is designated by ${}^{\sigma}(x) = x' - x$. H is the Lagrange invariant containing the aperture and the maximum field, see section 2.5, Invariants. For the chromatic aberrations we have the refractive index n for the reference wavelength $\check{\omega}$ and the refractive index $n + \check{\omega}n$ for the wavelength $\check{\omega} + \check{\omega}\check{\omega}$.

Based on these quantities, which are to be calculated by tracing two paraxial rays, namely the marginal ray for full aperture and the chief ray for the maximum field, the following expressions define the Seidel sums [29-2].

$$\text{Spherical aberration} \quad S_{\text{I}} = - \sum A^2 \cdot h \cdot {}^{\sigma} \left(\frac{u}{n} \right) \quad (29-23)$$

$$\text{Coma} \quad S_{\text{II}} = - \sum \bar{A} A \cdot h \cdot {}^{\sigma} \left(\frac{u}{n} \right) \quad (29-24)$$

$$\text{Astigmatism} \quad S_{\text{III}} = - \sum \bar{A}^2 \cdot h \cdot {}^{\sigma} \left(\frac{u}{n} \right) \quad (29-25)$$

$$\text{Petzval} \quad S_{\text{IV}} = - \sum H^2 \cdot c \cdot {}^{\sigma} \left(\frac{1}{n} \right) \quad (29-26)$$

$$\text{Distortion} \quad S_{\text{V}} = - \sum \left\{ \frac{\bar{A}^3}{A} \cdot h \cdot {}^{\sigma} \left(\frac{u}{n} \right) + \frac{\bar{A}}{A} \cdot H^2 \cdot c \cdot {}^{\sigma} \left(\frac{1}{n} \right) \right\} \quad (29-27)$$

$$\text{Axial color} \quad C_{\text{I}} = \sum A \cdot h \cdot {}^{\sigma} \left(\frac{\check{\omega}n}{n} \right) \quad (29-28)$$

$$\text{Lateral color} \quad C_{\text{II}} = \sum \bar{A} \cdot h \cdot {}^{\sigma} \left(\frac{\check{\omega}n}{n} \right) \quad (29-29)$$

where the summation is taken over all surfaces of the system.

Concerning the chromatic aberrations given in eqs (29-28) and (29-29) it must be mentioned that the height h and the refraction invariants A and \bar{A} actually depend on the wavelength. As for most imaging systems the influence of this dependence is small, h , A and \bar{A} are calculated only for the primary wavelength and the wavelength

dependence is neglected, as it is done in eqs (29-28) and (29-29). A more comprehensive treatment is given by Wynne [29-3].

It should be mentioned that also for reflecting mirror surfaces the formulae (29-23) to (29-29) should be applied. As usual for a reflective surface the refractive indices are set to $n' = -n$. Looking at eqs (29-28) and (29-29) it is evident that reflecting surfaces possess zero color aberrations. From eqs (29-23) – (29-27) for monochromatic aberrations, eq. (29-26) for the Petzval curvature becomes a special case due to the term ν ($1/n$). In fact, mirror surfaces will play a notable role in the correction of the Petzval curvature, see chapter 31, Correction of Aberrations.

The Seidel sums $S_I - S_V$ and $C_I - C_{II}$ are calculated for the maximum aperture and the maximum field. So, to express the wave aberration (29-8) in terms of these Seidel sums, the pupil coordinate r and the field coordinate y must be redefined as relative and dimensionless coordinates:

$$\begin{aligned} \rho &= \frac{r}{r_{\max}}, \\ \check{\eta} &= \frac{y}{y_{\max}}. \end{aligned} \tag{29-30}$$

With these quantities the total primary monochromatic wave aberration reads

$$\begin{aligned} W(\rho, \varphi, \check{\eta}) = \\ \frac{1}{8} S_I \rho^4 + \frac{1}{2} S_{II} \check{\eta} \rho^3 \cos \varphi + \frac{1}{2} S_{III} \check{\eta}^2 \rho^2 \cos^2 \varphi + \frac{1}{4} (S_{III} + S_{IV}) \check{\eta}^2 \rho^2 + \frac{1}{2} S_V \check{\eta}^3 \rho \cos \varphi. \end{aligned} \tag{29-31}$$

The constants $1/8$, $1/2$, etc. in (29-31) are a consequence of the common definition of the Seidel sums (29-23) – (29-29). As can be seen by a comparison of (29-8) and (29-31) there is a correspondence between the c -coefficients in the power series expansion and the Seidel coefficients. Table 29-2 gives an overview.

Table 29-2: Wave aberration expansions: power series and corresponding Seidel coefficients.

Aberration	Power series coefficient Dimension unit: (length) ⁻³	Seidel coefficient Dimension unit: (length)
Spherical aberration	c_1	S_I
Coma	c_2	S_{II}
Astigmatism (half astigmatic difference)	c_3	S_{III}
Sagittal field	c_4	$S_{III} + S_{IV}$
Tangential field	$c_4 + 2c_3$	$3S_{III} + S_{IV}$
Petzval field	$c_4 - c_3$	S_{IV}
Distortion	c_5	S_V

It should be noted that there is an important difference in the dimension units of the c -coefficients and the S -coefficients. While the S -coefficients have the dimension of length (due to h in eqs (29-23) – (29-29)) the c -coefficients have the dimension of $(\text{length})^{-3}$.

From table 29-2 the relations between the aberrations of astigmatism, sagittal field curvature, tangential field curvature, and Petzval field curvature can clearly be seen. These four aberrations are calculated from two coefficients, either c_3 and c_4 or S_{III} and S_{IV} . On the other hand, when any two of these four aberrations are known, the other two can be determined. This can be seen from table 29-2 as well as from eqs (29-32) and (29-33). When the astigmatism is zero, $S_{III} = 0$, the tangential and the sagittal field coincide with the Petzval field and S_{IV} represents the curved image plane. That is the only situation when there is a typical image formation at the location of the Petzval field. When the astigmatism is not zero, the Petzval field just serves as a basis to determine the sagittal and tangential field and $S_{IV} = 0$ is a precondition for a flat field. As seen from eqs (29-23) – (29-29) the Petzval field S_{IV} is the only Seidel coefficient which is completely independent of the ray quantities u , h , and A from the marginal as well as \bar{h} and \bar{A} from the chief ray, so S_{IV} is independent of object, image, and stop position. It is only the surface curvatures of the system and the refractive indices which determine the Petzval field S_{IV} . The Lagrange invariant H appears as a factor because the given formulation of the Seidel coefficients are related to the full aperture and the maximum field.

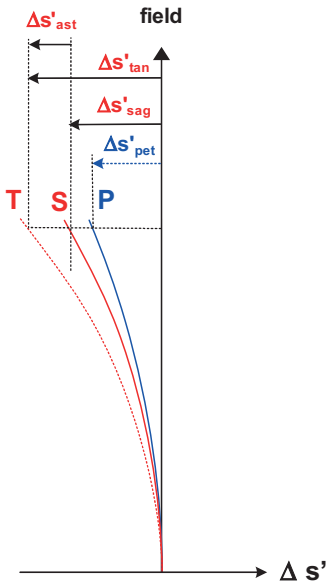


Figure 29-31: Longitudinal aberrations related to the Petzval, sagittal, and tangential field curvatures.

The longitudinal aberrations ${}^{\circ} s'_{\text{pet}}$, ${}^{\circ} s'_{\text{sag}}$, ${}^{\circ} s'_{\text{tan}}$, and ${}^{\circ} s'_{\text{ast}} = {}^{\circ} s'_{\text{tan}} - {}^{\circ} s'_{\text{sag}}$ which are measured parallel to the optical axis from the paraxial image plane to the related field are shown in figure 29-31. Eqs (29-32) and (29-33), which were already given in chapter 11, are in accordance with table 29-2,

$${}^{\circ} s'_{\text{ast}} = {}^{\circ} s'_{\text{tan}} - {}^{\circ} s'_{\text{sag}}, \quad (29-32)$$

$${}^{\circ} s'_{\text{pet}} = \frac{3 \cdot {}^{\circ} s'_{\text{sag}} - {}^{\circ} s'_{\text{tan}}}{2}. \quad (29-33)$$

The relation between the Seidel coefficients and the related longitudinal aberrations, are given in (29-78) – (29-81).

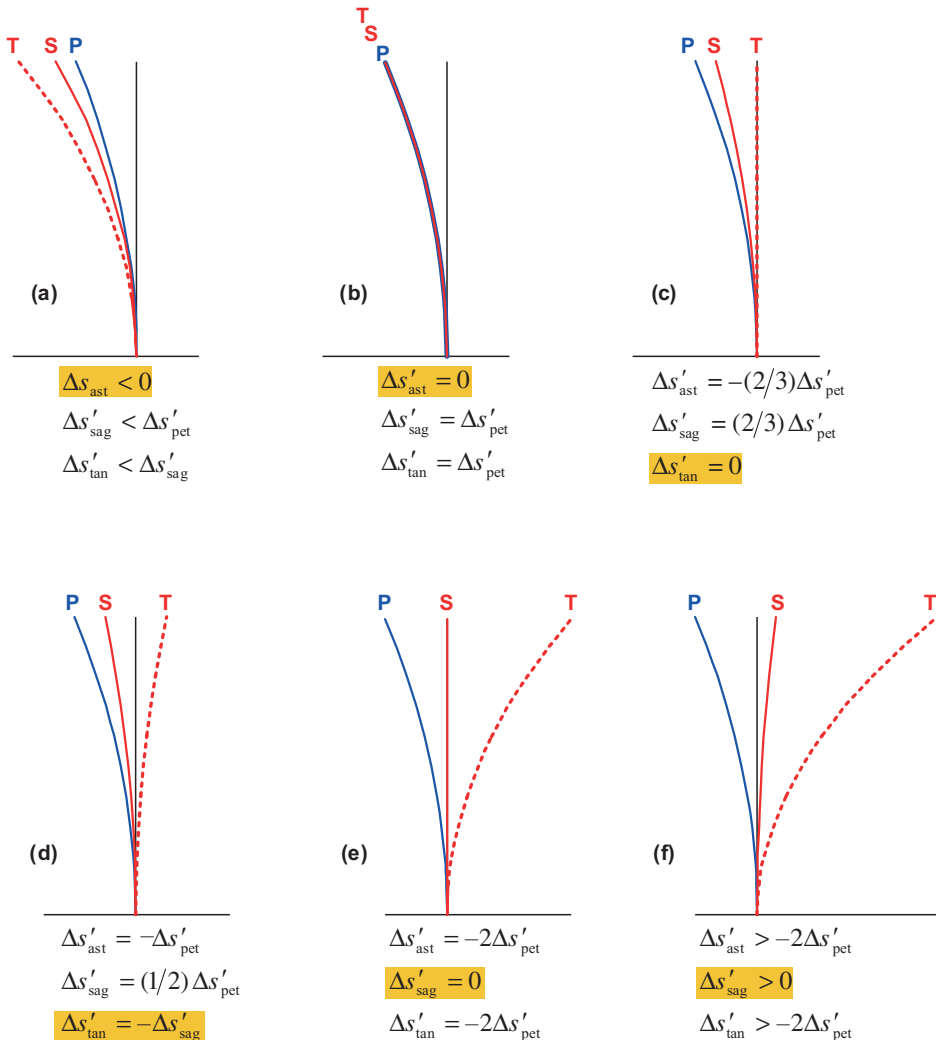


Figure 29-32: Different amounts of astigmatism with a constant Petzval field.

In figure 29-32 a constant, inwardly curved Petzval field is assumed and the sagittal and tangential fields for different amounts of astigmatism are shown. In partition (a) a typical uncorrected situation with some negative astigmatism is shown. In partition (b) the astigmatism is corrected to zero, but the image is not flat, but it is curved according to the Petzval field. Partition (c) represents a special situation, the tangential field is flat, but due to the negative Petzval field, which is held constant for all partitions, the sagittal field is negative and so there is some positive astigmatism. Partition (d) shows another specific situation when the sagittal field is negative and the tangential field is positive with the same absolute value. So, the medium field between the sagittal and tangential field is flat. This can be seen as a possible compromise when optimizing a system, if the Petzval field cannot be further reduced. In fact the astigmatism is larger than in partition (c), but in (c) the intermediate image between the sagittal and tangential is curved. The next partition (e) shows a flat sagittal field. The disadvantage of this situation, compared with the flat tangential field in (c), is the relatively large astigmatism. Partition (f) shows an even larger astigmatism with the result that both the sagittal and tangential fields are positive.

There is an interesting systematic in the formulae for the Seidel sums (29-23) – (29-29) which can easily be seen when the single summands, the so-called surface contributions, are designated by square brackets in the following way:

$$S_I = \sum_{\bar{k}} [s_I]_{\bar{k}} \quad (29-34)$$

$$S_{II} = \sum_{\bar{k}} [s_{II}]_{\bar{k}} = \sum_{\bar{k}} \frac{\bar{A}}{A} [s_I]_{\bar{k}} \quad (29-35)$$

$$S_{III} = \sum_{\bar{k}} [s_{III}]_{\bar{k}} = \sum_{\bar{k}} \frac{\bar{A}}{A} [s_{II}]_{\bar{k}} \quad (29-36)$$

$$S_{IV} = \sum_{\bar{k}} [s_{IV}]_{\bar{k}} \quad (29-37)$$

$$S_V = \sum_{\bar{k}} [s_V]_{\bar{k}} = \sum_{\bar{k}} \frac{\bar{A}}{A_{\bar{k}}} \{ [s_{III}]_{\bar{k}} + [s_{IV}]_{\bar{k}} \} \quad (29-38)$$

$$C_I = \sum_{\bar{k}} [c_I]_{\bar{k}} \quad (29-39)$$

$$C_{II} = \sum_{\bar{k}} [c_{II}]_{\bar{k}} = \sum_{\bar{k}} \frac{\bar{A}}{A} [c_I]_{\bar{k}} \quad (29-40)$$

A graphical presentation of the single surface contributions can be very helpful in order to get a specific insight as to which surfaces contribute which amount of the various aberrations. As an example, figure 29-33 shows a cross-section of a retrofocus-type photographic lens and figure 29-34 gives the associated surface contributions for the Seidel aberrations in a bar diagram [29-4]. It should be noticed that the scale for the various aberrations is different.

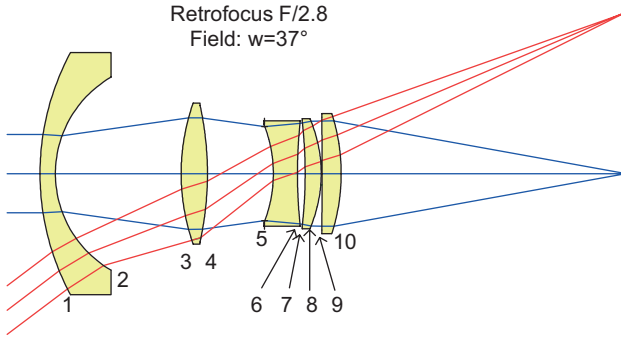


Figure 29-33: Retrofocus-type photographic lens, modified from [29-5].

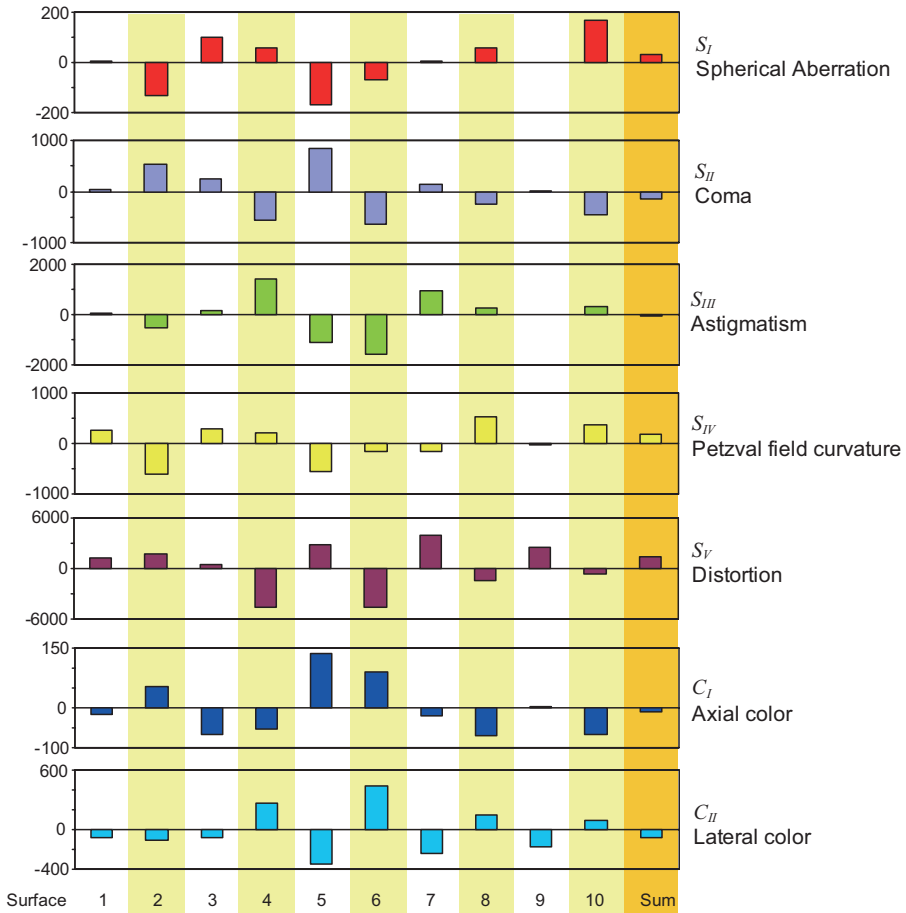


Figure 29-34: Bar diagram showing Seidel sums and related surface contributions for the lens in figure 29-33.

It can be seen in figure 29-34 that, for all aberrations, the Seidel sums in the last column are essentially smaller than the main surface contributions, which partly cancel each other. Further, in combination with figure 29-33 it can be seen that relatively large surface contributions arise at surfaces where the associated rays have large incidence angles. For instance: regarding spherical aberration, the marginal ray exhibits relatively large incidence angles at surfaces 5 and 10, while for distortion the chief ray exhibits large incidence angles at surfaces 4, 6, and 7.

When correcting the aberrations of an optical system it is usually not enough to obtain small aberrations for the whole system, but the single surface contributions should also be small, so that the system is not too sensitive to manufacturing tolerances. See chapter 35 for a further discussion.

29.4.5

Stop Shift Formulae

As can be seen from eqs (29-23) – (29-29) S_I is completely independent of the chief ray, which is expected, because S_I represents spherical aberration. S_{IV} is dependent on H , the Lagrange invariant and so S_{IV} depends on the maximum aperture and the maximum field, but S_{IV} is independent of the marginal ray quantities h , u , A , and also independent of the chief ray quantities \bar{A} . The Seidel sums which depend on the chief ray are: S_{II} , S_{III} , and S_V . As the chief ray is determined by the stop position, S_{II} , S_{III} , and S_V must change when the stop position in the system is changed. There are so-called stop shift formulae which are very simple and provide an interesting insight into the impact of shifting the stop. These stop shifts are specified in a way that the marginal ray does not change, that means that while shifting the stop position, the size of the stop is adapted to the marginal ray as shown in figure 29-35.

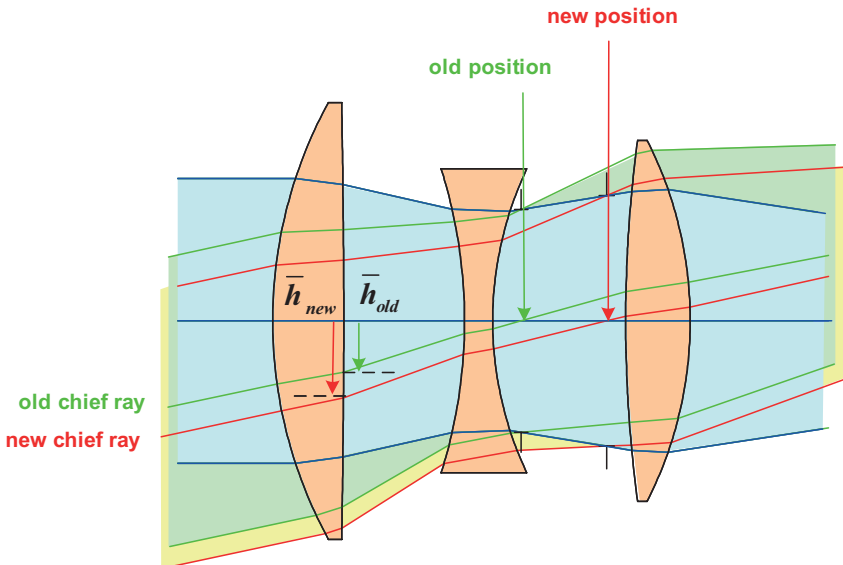


Figure 29-35: Stop shift.

The stop position can be characterized at each surface by the so-called Seidel eccentricity

$$E = \frac{\bar{h}}{h}, \quad (29-41)$$

which indicates the eccentricity of the oblique ray bundle relative to the axial bundle. The main advantage in describing the stop position at each surface with this parameter E is the handling of a stop shift. When the stop is shifted a certain distance, as shown in figure 29-35, at a chosen surface the chief ray intersection heights change from \bar{h}_{old} to \bar{h}_{new} and the Seidel eccentricity is changed by the amount

$$\omega E = \frac{\bar{h}_{\text{new}} - \bar{h}_{\text{old}}}{h}. \quad (29-42)$$

It transpires that, for a given stop shift, this quantity ωE has the same value for all surfaces of the system. Thus ωE indicates a certain stop shift.

Let S_1^* to S_V^* and C_1^* and C_{II}^* indicate the Seidel sums after a stop shift represented by ωE , then there is a very simple and instructive set of formulae describing the transformation of the Seidel sums, the so-called stop shift formulae:

$$S_1^* = S_I \quad (29-43)$$

$$S_{II}^* = S_{II} + \omega E \cdot S_I \quad (29-44)$$

$$S_{III}^* = S_{III} + \omega E \cdot S_{II} + \omega E^2 \cdot S_I \quad (29-45)$$

$$S_{IV}^* = S_{IV} \quad (29-46)$$

$$S_V^* = S_V + \omega E \cdot (3S_{III} + S_{IV}) + 3\omega E^2 \cdot S_{II} + \omega E^3 \cdot S_I \quad (29-47)$$

$$C_1^* = C_I \quad (29-48)$$

$$C_{II}^* = C_{II} + \omega E \cdot C_I \quad (29-49)$$

From these formulae, for instance, it can be seen that with a stop shift the coma (S_{II}) changes in the same rate as the spherical aberration (S_I) occurs in the system. So, more or less theoretically, if S_I is unequal to zero, then the coma can be made zero by the appropriate localization of the stop position. On the other hand the coma will be unaffected by a stop shift if the spherical aberration is zero. Such types of statements based on the stop shift formulae (29-43) – (29-49) are valid of course only for primary aberrations. With higher order aberrations equivalent statements would be much more complicated. Corresponding formulae are therefore not in use.

29.4.6

Several Aberration Expressions from the Seidel Sums

Based on the Seidel sums $S_I - S_V$ and $C_I - C_{II}$ as given in eqs (29-23) – (29-29) all primary aberrations of a given optical system can be calculated.

There are in particular: wave aberrations, transverse aberrations, angular aberrations, fractional expressions, and all field curvatures.

In several of these expressions u' designates the image side convergence angle (maximum aperture) and n' is the refractive index in the image space. The tables 29-3 – 29-8 summarize these results.

Table 29-3: Wave aberrations based on Seidel sums.

Primary wave aberrations		
Spherical aberration	$W_{\text{spherical}} = \frac{S_I}{8} \rho^4$	(29-50)
Coma	$W_{\text{coma}} = \frac{S_{II}}{2} \tilde{\eta} \rho^3 \cos \varphi$	(29-51)
Astigmatism	$W_{\text{ast}} = \frac{S_{III}}{2} \tilde{\eta}^2 \rho^2 \cos^2 \varphi$	(29-52)
Petzval field	$W_{\text{petzval}} = \frac{S_{IV}}{4} \tilde{\eta}^2 \rho^2$	(29-53)
Sagittal field	$W_{\text{sag}} = \frac{S_{III} + S_{IV}}{4} \tilde{\eta}^2 \rho^2$	(29-54)
Tangential field	$W_{\text{tan}} = \frac{3S_{III} + S_{IV}}{4} \tilde{\eta}^2 \rho^2$	(29-55)
Distortion	$W_{\text{dist}} = \frac{S_V}{2} \tilde{\eta}^3 \rho \cos \varphi$	(29-56)
Axial color	$W_{\text{ac}} = \frac{C_I}{2} \rho^2$	(29-57)
Lateral color	$W_{\text{lc}} = C_{II} \tilde{\eta} \rho \cos \varphi$	(29-58)

Table 29-4: Transverse aberrations based on Seidel sums.

Primary transverse aberrations		
Spherical aberration (meridional ray)	$\epsilon \gamma' = \frac{S_I}{2n'u'} \left(\frac{Y_p}{r_{\max}} \right)^3$	(29-59)
Sagittal coma (sagittal in the pupil)	$\epsilon \gamma' = \frac{S_{II}}{2n'u'} \check{\eta} \left(\frac{x_p}{r_{\max}} \right)^2$	(29-60)
Tangential coma (tangential in the pupil)	$\epsilon \gamma' = \frac{3S_{II}}{2n'u'} \check{\eta} \left(\frac{Y_p}{r_{\max}} \right)^2$	(29-61)
Sagittal field (sagittal ray)	$\epsilon x' = \frac{S_{III} + S_{IV}}{2n'u'} \check{\eta}^2 \left(\frac{x_p}{r_{\max}} \right)$	(29-62)
Tangential field (meridional ray)	$\epsilon \gamma' = \frac{3S_{III} + S_{IV}}{2n'u'} \check{\eta}^2 \left(\frac{Y_p}{r_{\max}} \right)$	(29-63)
Length of astigmatic focal line (full aperture)	$\epsilon \gamma' = \left \frac{2S_{III}}{n'u'} \right \check{\eta}^2$	(29-64)
Distortion	$\epsilon \gamma' = \frac{S_V}{2n'u'} \check{\eta}^3$	(29-65)
Axial color (transverse)	$\epsilon \gamma' = \frac{C_I}{nu} \left(\frac{Y_p}{r_{\max}} \right)$	(29-66)
Lateral color	$\epsilon \gamma' = \frac{C_{II}}{nu} \check{\eta}$	(29-67)

Table 29-5: Angular aberrations based on Seidel sums.

Primary angular aberrations		
$\epsilon w'_y$ is the angle difference (radians) in the tangential plane		
$\epsilon w'_x$ is the angle difference (radians) in the sagittal plane		
Spherical aberration (meridional ray)	$\epsilon w'_y = \frac{S_I}{2n'r_{\max}} \left(\frac{Y_p}{r_{\max}} \right)^3$	(29-68)
Sagittal coma (sagittal in the pupil)	$\epsilon w'_y = \frac{S_{II}}{2n'r_{\max}} \check{\eta} \left(\frac{x_p}{r_{\max}} \right)^2$	(29-69)
Tangential coma (tangential in the pupil)	$\epsilon w'_y = \frac{3S_{II}}{2n'r_{\max}} \check{\eta} \left(\frac{Y_p}{r_{\max}} \right)^2$	(29-70)
Sagittal field (sagittal ray)	$\epsilon w'_x = \frac{S_{III} + S_{IV}}{2n'r_{\max}} \check{\eta}^2 \left(\frac{x_p}{r_{\max}} \right)$	(29-71)
Tangential field (meridional ray)	$\epsilon w'_y = \frac{3S_{III} + S_{IV}}{2n'r_{\max}} \check{\eta}^2 \left(\frac{Y_p}{r_{\max}} \right)$	(29-72)
Distortion	$\epsilon w'_y = \frac{S_V}{2n'r_{\max}} \check{\eta}^3$	(29-73)
Axial color	$\epsilon w'_y = \frac{C_I}{n'r_{\max}} \left(\frac{Y_p}{r_{\max}} \right)$	(29-74)
Lateral color	$\epsilon w'_y = \frac{C_{II}}{n'r_{\max}} \check{\eta}$	(29-75)

Table 29-6: Longitudinal aberrations based on Seidel sums.

Primary longitudinal aberrations	
Spherical aberration (meridional ray)	${}^{\circ} s' = -\frac{S_I}{2n'u'^2} \left(\frac{y_p}{r_{\max}} \right)^2$ (29-76)
Coma	${}^{\circ} s'_{\text{coma}} = -\frac{3S_{II}}{2n'u'^2} \tilde{\eta} \left(\frac{y_p}{r_{\max}} \right)$ (29-77)
Sagittal field (Gaussian image to s-focus)	${}^{\circ} s'_{\text{sag}} = -\frac{S_{III} + S_{IV}}{2n'u'^2} \tilde{\eta}^2$ (29-78)
Tangential field (Gaussian image to t-focus)	${}^{\circ} s'_{\text{tan}} = -\frac{3S_{III} + S_{IV}}{2n'u'^2} \tilde{\eta}^2$ (29-79)
Astigmatism (from s- to t-focus)	${}^{\circ} s'_{\text{ast}} = {}^{\circ} s'_{\text{tan}} - {}^{\circ} s'_{\text{sag}} = -\frac{S_{III}}{n'u'^2} \tilde{\eta}^2$ (29-80)
Petzval field (Gaussian image to Petzval)	${}^{\circ} s'_{\text{pet}} = -\frac{S_{IV}}{2n'u'^2} \tilde{\eta}^2$ (29-81)
Axial color	${}^{\circ} s' = -\frac{C_1}{n'u'^2}$ (29-82)

Table 29-7: Fractional aberrations based on Seidel sums.

Primary fractional aberrations	
Fractional distortion	$\frac{S_V}{2n'u'y'_{\max}} \tilde{\eta}^2 = \frac{S_V}{2H} \tilde{\eta}^2$ (29-83)
Fractional lateral color	$\frac{C_{II}}{2n'u'y'_{\max}} = \frac{C_{II}}{H}$ (29-84)

Table 29-8: Field curvature aberrations based on Seidel sums.

Primary image surface curvatures	
Sagittal field curvature	$c_{\text{sag}} = -\frac{S_{III} + S_{IV}}{n'u'^2 y_{\max}^2} = -\frac{n'(S_{III} + S_{IV})}{H^2}$ (29-85)
Tangential field curvature	$c_{\text{tan}} = -\frac{3S_{III} + S_{IV}}{n'u'^2 y_{\max}^2} = -\frac{n'(3S_{III} + S_{IV})}{H^2}$ (29-86)
Petzval field	$c_{\text{pet}} = -\frac{S_{IV}}{n'u'^2 y_{\max}^2} = -\frac{n'S_{IV}}{H^2}$ (29-87)

29.4.7

Thin Lens Aberrations

In eqs (29-23) – (29-29) the surface contributions of the primary aberrations are given. Often it is advantageous to work with the contributions of a thin lens. The idealization of zero lens thickness leads to a considerable simplification of the formulae. Although there are no thin lenses in the real world, the results from the thin lens theory hold in many cases to a good approximation. To get the thin lens contributions, two surface summands of eqs (29-23) – (29-29) must be considered and should be expressed in terms of the parameters of a thin lens. These thin lens parameters are the bending parameter X and the position or conjugate or magnification parameter M (see sections 10.1.1 Bending of Lenses, and 10.1.2 Position Parameter),

$$X = \frac{c_1 + c_2}{c_1 - c_2} \quad (29-88)$$

where c_1 and c_2 are the curvatures of the two surfaces, and

$$M = \frac{u_2' + u_1}{u_2' - u_1} \quad (29-89)$$

where u_1 and u_2' are the angles of the paraxial marginal ray before and after the lens, respectively. Let n be the refractive index of the lens, then the refractive power reads

$$\check{\sigma} = (n - 1)(c_1 - c_2). \quad (29-90)$$

For the chromatic aberrations we will use the Abbe number, which in this context is defined as

$$\check{\epsilon} = \frac{n - 1}{\frac{\partial n}{\partial \lambda}}. \quad (29-91)$$

To get the essential information from the thin lens contribution formulae, it is sufficient to assume the stop at the lens, so that the incidence height of the chief ray is zero. If the influence of a remote stop position is to be incorporated, it is easy to apply the stop shift formulae (29-43) – (29-49).

With these parameters the primary aberrations of the thin lens (stop at the lens) are:

$$S_I = \frac{\check{\sigma}^3 h^4}{4} \left(\frac{3n + 2}{n} M^2 - \frac{4n + 4}{(n - 1)n} XM + \frac{n + 2}{(n - 1)^2 n} X^2 + \frac{n^2}{(n - 1)^2} \right) \quad (29-92)$$

$$S_{II} = \frac{H \check{\sigma}^2 h^2}{2} \left(\frac{2n + 1}{n} M - \frac{n + 1}{(n - 1)n} X \right) \quad (29-93)$$

$$S_{III} = H^2 \check{\sigma} \quad (29-94)$$

$$S_{IV} = \frac{H^2 \delta}{n} \quad (29-95)$$

$$S_V = 0 \quad (29-96)$$

$$C_I = \frac{\delta h^2}{\bar{\epsilon}} \quad (29-97)$$

$$C_{II} = 0. \quad (29-98)$$

Again, to calculate the Seidel sums, the paraxial rays for both maximum aperture and maximum field must be used. As the Lagrange invariant H contains both the aperture and the field size as linear factors, the dependence of the Seidel contributions on the aperture and field can easily be checked.

From eq. (29-92) it can be seen that the spherical aberration S_I depends quadratically on both M , the conjugate parameter, and on X , the bending parameter. In figures 29-36, 29-37 and 29-38 this behavior is illustrated. In figure 29-36 the spherical aberration is shown as a function of the bending parameter X for several conjugate parameters M and for a refractive index $n = 1.5$. In figures 29-37 and 29-38 contour plots and a section for $M = 3$ are shown for two different refractive indices, for $n = 1.5$ and for $n = 1.9$. With the help of the lines for zero spherical aberration it can be seen that, in the range of real imaging (neither virtual object nor image) represented by $-1 \leq M \leq +1$, the spherical aberration is positive, which implies under-correction. Only for extremely large conjugate parameters M and the appropriate bending parameters X can the spherical aberration vanish or even exhibit over-correction. Figure 29-38 shows that for the high refractive index $n = 1.9$ the smallest M which allows approximately zero spherical aberration is $M = 3$, the appropriate bending parameter is then $X = 4$. Because of the symmetry of the eq. (29-92) the spherical aberration is also very close to zero when changing the sign of both M and X , i. e. for $M = -3$ and $X = -4$.

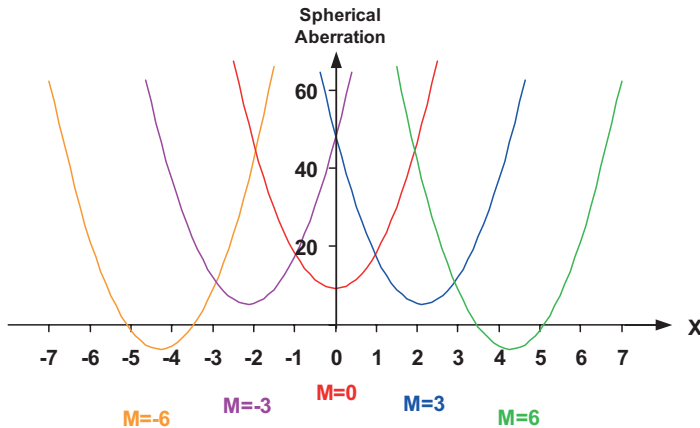


Figure 29-36: Spherical aberration, $\frac{4}{\delta^3 h^4} S_I$ of a thin lens as a function of the bending parameter X and the conjugate parameter M . The refractive index is $n = 1.5$.

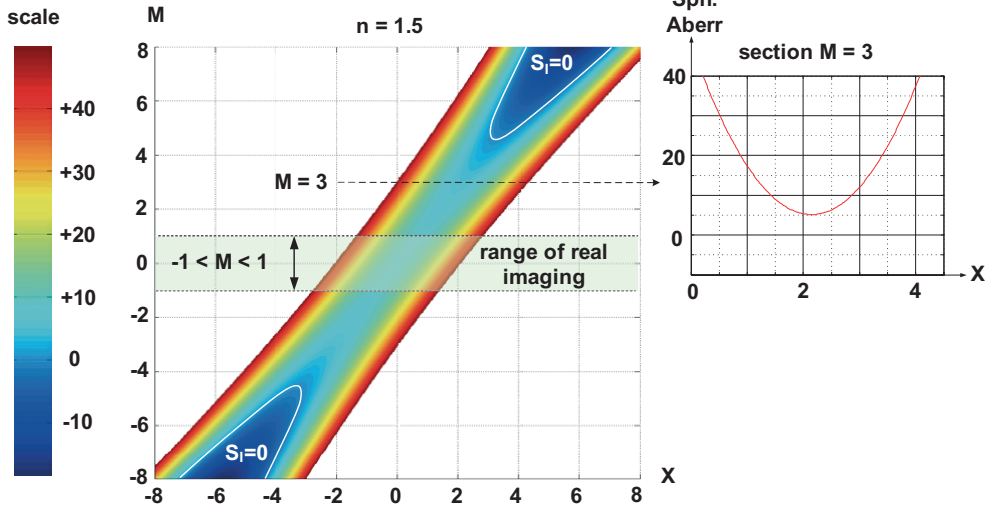


Figure 29-37: Spherical aberration, $\frac{4}{\sigma^3 h^4} S_1$ of a thin lens as a function of the bending parameter X and the conjugate parameter M . The refractive index is $n = 1.5$.

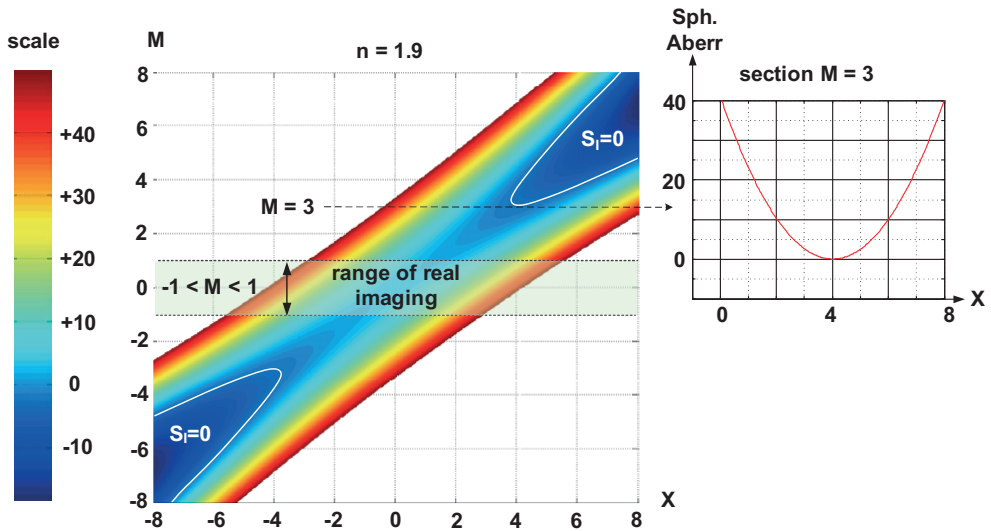


Figure 29-38: Spherical aberration, $\frac{4}{\sigma^3 h^4} S_1$ of a thin lens as a function of the bending parameter X and the conjugate parameter M . The refractive index is $n = 1.9$.

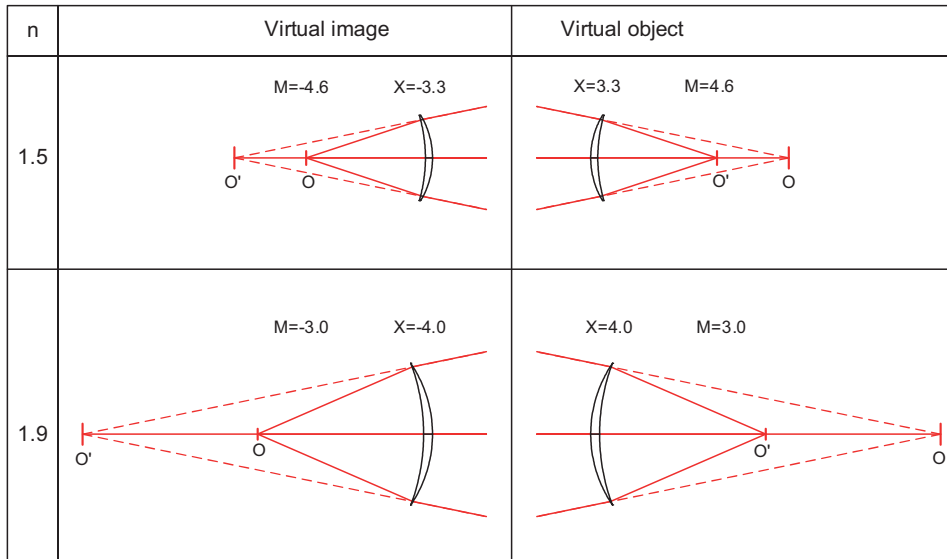


Figure 29-39: Single lens with zero spherical aberration with minimum $|M|$ for $n = 1.5$ and for $n = 1.9$.

In figure 29-39 it is demonstrated again that the condition for zero primary spherical aberration for a single lens is possible only in virtual imaging and for a strongly meniscus shaped lens. For the refractive indices $n = 1.5$ and $n = 1.9$ the minimum values for the absolute conjugate parameter $|M|$ which allow zero spherical aberration are used, together with the appropriate lens bending X . In fact eq. (29-92) is based on thin lenses but in figure 29-39 finite lens thicknesses are introduced for more clarity. On the other hand these thicknesses have low influence and can often be neglected.

For a given conjugate parameter M the bending parameter X , which delivers the minimum spherical aberration can be calculated from eq. (29-92):

$$X_{\text{sph min}} = \frac{2(n^2 - 1)}{n + 2} M. \tag{29-99}$$

A special application of this formula yields the optimal bending of a single thin lens, for the object at infinity, which means the incoming light is collimated and the conjugate parameter is $M = +1$. A lens with this bending parameter,

$$X_{\text{min}} = \frac{2(n^2 - 1)}{n + 2}, \tag{29-100}$$

is called the lens with the optimal bending with respect to spherical aberration. According to eq. (29-100) the optimal bending depends on the refractive index n . For $n = 1.686$ one gets $X = 1$, the plano-convex lens with the convex surface on the object side. For indices $n < 1.686$ the optimal bending shape is biconvex and for $n > 1.686$ the best form is a meniscus shape.

Eq. (29-93) for S_{II} shows that coma depends linearly on the conjugate parameter M as well as on the bending parameter X . So, for a lens, whatever the conjugate parameter may be, there is always a suitable bending

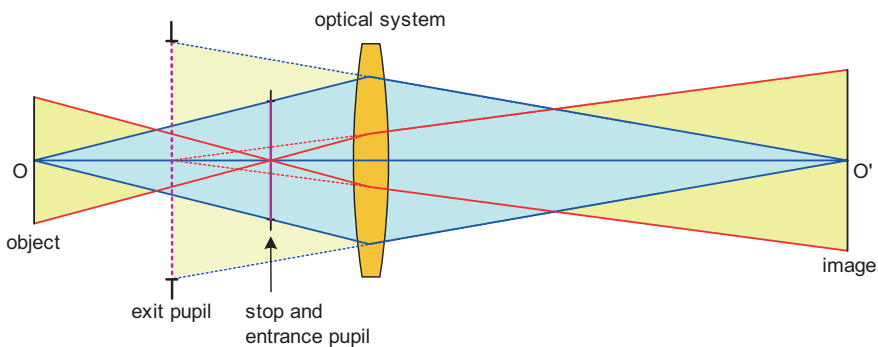
$$X = \frac{(2n + 1)(n - 1)}{(n + 1)} M \quad (29-101)$$

which makes the primary coma zero. If the stop is not at the lens, the contribution to coma due to the spherical aberration according to the stop shift formula (29-44) must also be taken in account.

From eqs (29-94)–(29-96) we see that astigmatism, Petzval curvature and distortion are independent of both the lens bending (X) and the conjugate position (M). But with a remote stop due to the contributions from spherical aberration and also from coma, which take effect through the stop shift formulae (29-45) and (29-47), astigmatism and distortion are no longer independent of the lens bending (X) and of the conjugate position (M). The primary chromatic aberrations C_I and C_{II} are always independent of both X and M .

29.5 Pupil Aberrations

After the discussion of the influence of a stop shift on the primary aberrations in paragraph 29.4.5, it sounds reasonable to also investigate an object shift. Of course in most optical designs an object shift does not have the relevance of a pupil shift but the object shift is studied in order to obtain a better theoretical insight, [29-6], [29-7]. The best way of doing this is to exchange the function of the object and the pupil. More precisely: rather than considering the imaging of the object onto the image through the pupil, the imaging of the entrance pupil onto the exit pupil through the object and image are examined instead. In this process, as shown in figure 29-40 the



	Object imaging	Pupil imaging
Blue rays	Marginal rays	Chief rays
Red rays	Chief rays	Marginal rays

Figure 29-40: Object and pupil imaging.

marginal rays of the normal imaging become the chief rays of the pupil imaging and the chief rays of the normal imaging act as marginal rays in the pupil imaging.

In fact there is a complete dualism in object and pupil imaging. So all the given formulae can be used for the pupil imaging to calculate pupil aberrations, provided that the quantities based on the marginal ray such as u , A , h and so on are replaced by their corresponding quantities \bar{u} , \bar{A} , \bar{h} based on the chief ray, and vice versa. These considerations are not really difficult but lengthy and – much more important – the results are usually not in use. Nevertheless there are some interesting conclusions from this pupil aberration theory which should be quoted here:

- **Wave aberration power series expansion**
Primary spherical pupil aberration only depends on the field coordinate. As a wave aberration it depends on the 4th power in the field coordinate, which corresponds to the term $c_6 y^4$ in the power series expansion (29-5) for the wave aberration. This term was set to zero, because at that time only image aberrations were considered. So, all terms in the power series expansion have indeed meaning, if it is not only the image aberrations but also the pupil aberrations that are taken into account.
- **Distortion**
The spherical pupil aberration plays a part when, for normal imaging, the distortion is investigated independently of the object and corresponding image shift. In fact, the change in distortion is proportional to the pupil spherical aberration in the exit pupil. So, if the distortion is to be independent of the object position, then the spherical pupil aberration must be zero. Under the precondition of zero spherical pupil aberration the distortion is zero if the so-called tangent condition [29-8] is fulfilled. The tangent condition reads

$$\frac{\tan \bar{U}'}{\tan \bar{U}} = \text{constant} \quad \text{for all field angles } \bar{U} \text{ and } \bar{U}'. \quad (29-102)$$

- **Obscuration**
When two parts of an interlinked optical system, for example, an objective and an eyepiece, are put together, both the field positions and the pupil positions must be matched. In greater detail: the object plane of the eyepiece has to match the image plane of the objective and the entrance pupil of the eyepiece has to match the exit pupil of the objective. When the parts to be combined are designed simultaneously this usually doesn't represent a problem, but when they are designed independently, then pupil aberrations, in particular, may frustrate the combination. A similar situation occurs with the combination of an eyepiece and the eye of the observer. This case is actually more critical, since the pupil of the eye is not really fixed. Particularly in wide-angle eyepieces, where the observer has to rotate the eye to look from one to the other end of the field of view. Due to this rotation the pupil of the eye takes a decentered position and, dependent on the size of the exit pupil of the eyepiece and the size of the eye pupil, this may result in a decreased image

illumination. If the eyepiece exhibits a strong spherical pupil aberration, then the decentered eye pupil may result in a partial obscuration of the image field. An example of a significant pupil aberration is given in figures 29-41 to 29-43. Figure 29-41 shows an eyepiece with a rather large pupil aberration, together with an idealized eye. The curve of the transverse spherical pupil aberration is given in figure 29-42. Note the maximum values of about ± 4 mm in the zone of the pupil. This large aberration is the reason for the obscuration increasing with the decentering of the eye pupil, as displayed in figure 29-43. When the pupil of the eye is decentered, the rays in the zone of the pupil are truncated first, as can be understood from the ray path in the insert of figure 29-41. Therefore, the intensity first drops in a zonal region of the retina image. A certain area in the image even shows zero intensity. If high-quality imaging of the pupil of the instrument is required, then the decentering would cause a uniform decrease in the intensity.

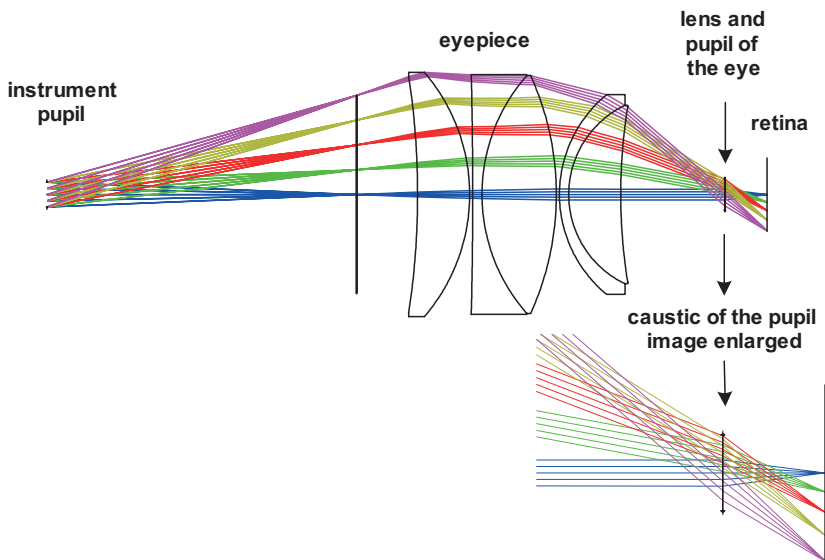


Figure 29-41: A ray path in the eyepiece with an idealized eye. The system suffers from a large spherical pupil aberration, see figure 29-42. The ray caustic is also shown in an enlarged illustration.

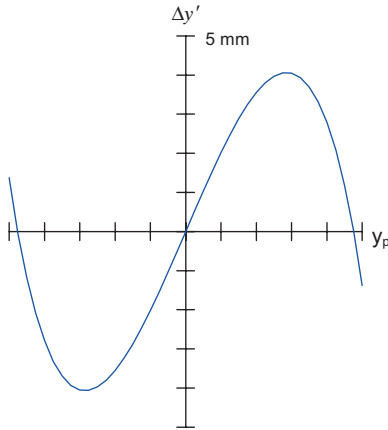


Figure 29-42: Spherical pupil aberration as a transverse aberration of the eyepiece in figure 29-41.

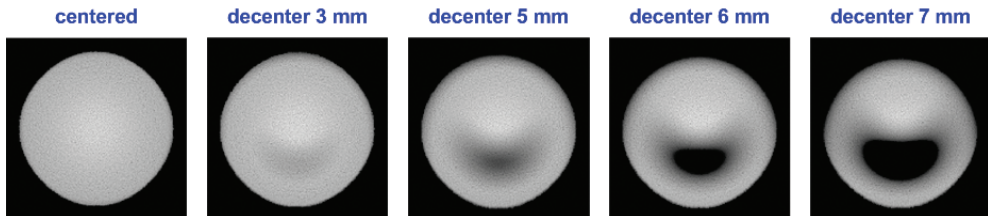


Figure 29-43: Illumination of the retina image for the system shown in figure 29-41 with different values of the decentered eye pupil. The offset of the pupil causes a truncation of the ray bundles and therefore the intensity is locally decreased.

- Ray aiming

Pupil aberrations play a role in another field, which is more closely related to practical ray tracing than to image aberrations. In order to trace a selected ray, such as a marginal ray or a chief ray, through an optical system, then for fast ray tracing, the entrance pupil is very often used to aim the ray. As the entrance pupil is defined as the paraxial image of the aperture stop through that part of the system which is in front of the stop, it may occur that rays aimed towards the entrance pupil cannot pass through the system. This is due to large pupil aberrations. An example is given in figure 29-44 (a), which shows a wide-angle lens modified from [29-9] to demonstrate this effect. None of the rays from the maximum oblique bundle, when aimed towards the entrance pupil can pass through the lens. In such systems with large pupil aberrations the rays to be traced must be aimed towards the real aperture stop as shown in figure 29-44 (b). This is more time consuming but does solve the problem.

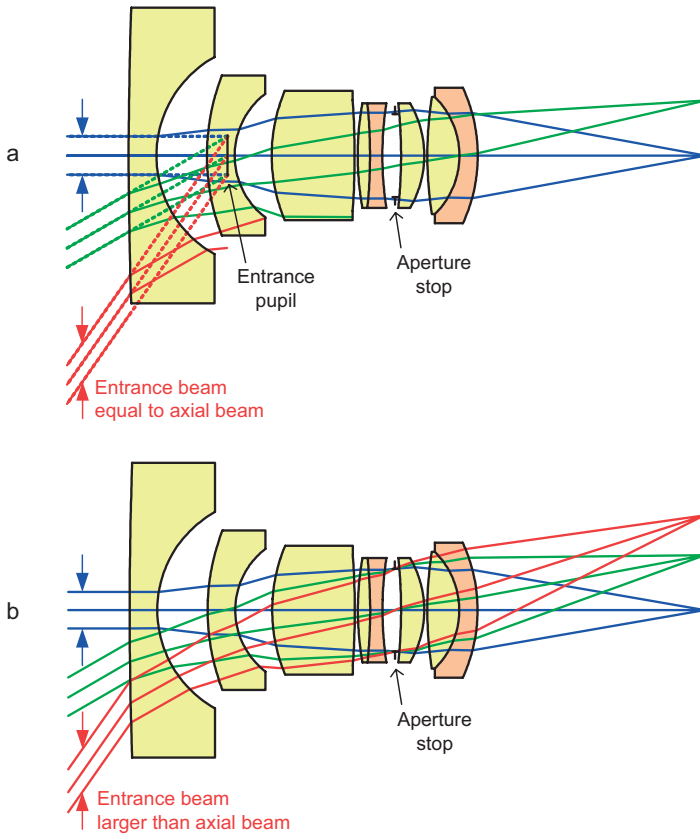


Figure 29-44: Pupil aberration and ray aiming. (a) Ray aiming towards the paraxial entrance pupil; (b) ray aiming towards the aperture stop.

- **Illumination**

As can be seen in figure 29-44 (b), when the rays are traced to the aperture stop, there is also an effect on the meridional bundle diameters. With increasing field angle the meridional bundle diameters also increase, which reduces the illumination fall-off in the image. By further studying this effect it can be shown that the increased entrance beam diameter in the oblique bundles is a consequence of pupil coma aberration: when imaging the aperture stop on the entrance pupil there is a considerable amount of outer coma.

- **Petzval curvature**

The Seidel Petzval sum for usual and pupil imaging is identical. So the Petzval curvature in the exit pupil, when assuming a flat entrance pupil, is the same as in the image.

29.6 High-order Aberrations

The primary or third-order aberrations discussed in section 29.4 represent the basic principles of aberration theory. In practice, already the secondary or fifth-order aberrations are not usually used in complete detail to analyse the aberrations of an optical system. As we will see in the next paragraph, the set of nine fifth-order aberrations contain new types of aberration, particularly with regard to the pupil or aperture dependence of the aberrations. Nevertheless, the main problem in lens design is to prevent or to correct all, including the higher-order aberrations. It is well known by the experienced optical designer that, during the optimization of an optical system, the third-order aberrations are much more movable than the higher-order aberrations. So, if it is not possible to correct the aberrations to zero, the result of the optimization is a balancing of third-order against higher-order aberrations of the same, or at least of a similar, type. In this view the amount of third-order aberrations is chosen to balance the uncorrectable higher-order aberrations. For this reason, in the next paragraph, the fifth-order aberrations are grouped together with the related third-order aberrations.

29.6.1

Fifth-order Aberrations

The terms containing the coefficients d_1 to d_9 in eqs (29-5) and (29-8) represent the nine secondary or fifth-order wave aberrations. In the following figures (29-45 – 29-52) the pupil dependence of these fifth-order aberrations are arranged with the related third-order types.

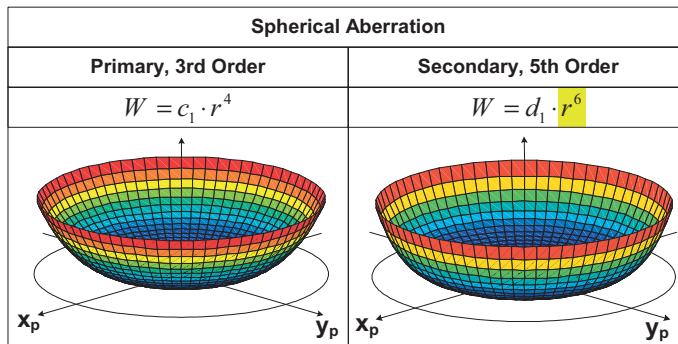


Figure 29-45: Primary and secondary spherical aberration.

The primary spherical aberration is of 4th power in the aperture. The secondary spherical aberration is of 6th power in the aperture. Higher power spherical aberration is easy to detect. As an example, in figure 29-53, a microscope objective with at least ninth-order spherical aberration is shown.

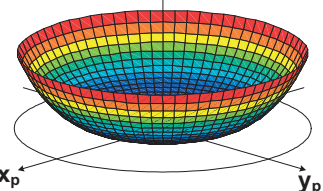
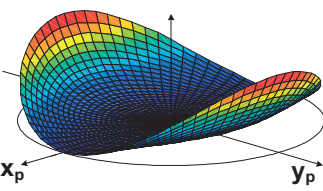
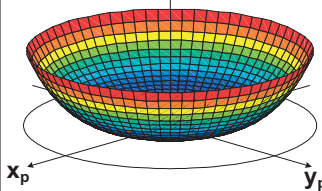
Spherical Aberration		
Primary, 3rd Order	Secondary, 5th Order	
Oblique Spherical Aberration		
$W = c_1 \cdot r^4$	$W = d_3 \cdot y^2 r^4 \cos^2 \varphi$	$W = d_4 \cdot y^2 r^4$
		

Figure 29-46: Primary and oblique spherical aberration.

The oblique spherical aberration (the sum of the d_3 and d_4 terms) is of 4th power in aperture as is the primary spherical aberration. But the field dependence of the oblique spherical aberration is of 2nd power and due to the d_3 term the amount of wave aberration is different in the tangential and sagittal directions. In practical lens design the resulting component in the sagittal direction is often dominant and is called: the “oblique spherical sagittal aberration”. As mentioned in the discussion about the symmetry properties of aberrations in section 29.4.2, the oblique spherical aberration does not have rotational symmetry as does the third-order spherical aberration.

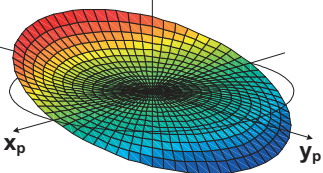
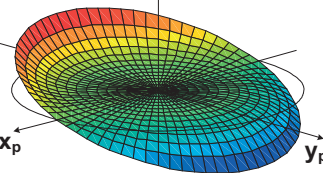
Coma	
Primary, 3rd Order	Secondary, 5th Order
Linear coma	
$W = c_2 \cdot y r^3 \cos \varphi$	$W = d_2 \cdot y r^3 \cos \varphi$
	

Figure 29-47: Primary and secondary linear coma.

The designation “linear” coma describes the linear field dependence of this aberration. The higher power appears only in the pupil coordinate. It should be noted that this higher-order aberration is included in the scope of the sine condition and the “offence against the sine condition” just because it is linear in the field coordinate as discussed in section 29.8.1.

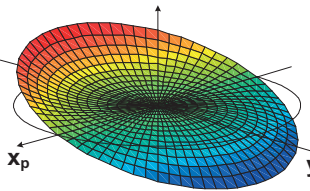
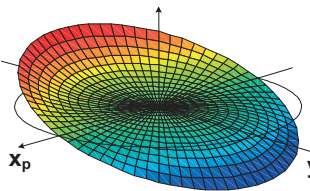
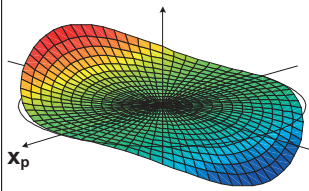
Coma		
Primary, 3rd Order	Secondary, 5th Order	
	Elliptical coma	
$W = c_2 \cdot y r^3 \cos \varphi$	$W = d_5 \cdot y^3 r^3 \cos \varphi$	$W = d_6 \cdot y^3 r^3 \cos^3 \varphi$
		

Figure 29-48: Primary and elliptical coma.

The designation “elliptical” coma results from the appearance of the associated transverse aberration. The circles in the image spot shown in figure 29-23 for third-order coma, originating from concentric circles in the pupil, will change to ellipses in the case of elliptical coma. The coefficient d_6 controls the shape of the ellipses. An example analog to figure 29-23 is shown in figure 29-49.

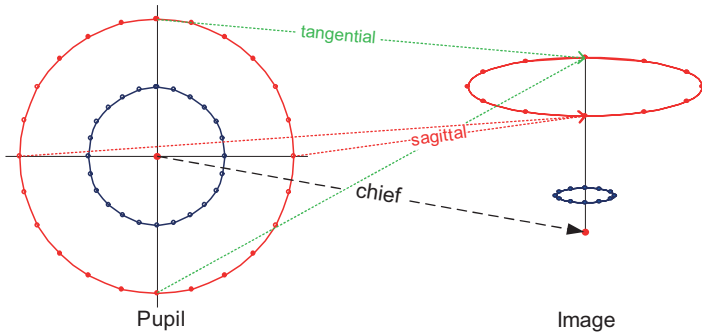


Figure 29-49: Schematic ray path for elliptical coma.

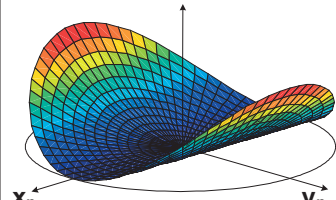
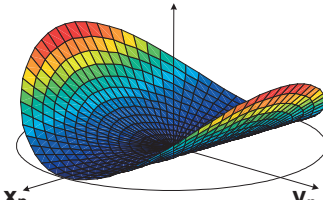
Astigmatism	
Primary, 3rd Order	Secondary, 5th Order
$W = c_3 \cdot y^2 r^2 \cos^2 \varphi$	$W = d_7 \cdot y^4 r^2 \cos^2 \varphi$
	

Figure 29-50: Primary and secondary astigmatism.

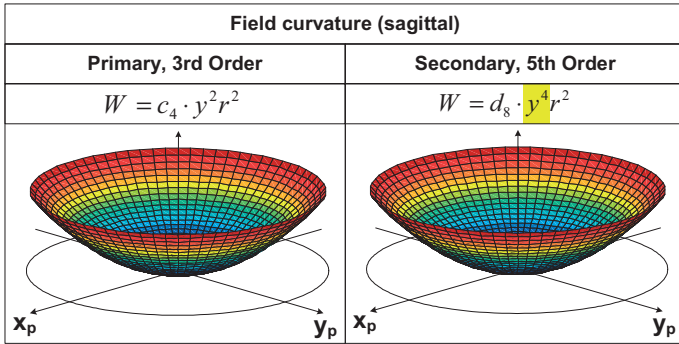


Figure 29-51: Primary and secondary sagittal field curvature.

Fifth-order astigmatism and fifth-order field curvature as shown in figures 29-50 and 29-51 are both easy to understand: it is just the field dependence which is of fourth power instead of second power for the primary aberrations. As the field variable is not a parameter for the wave surfaces shown, the figures for primary and secondary aberrations appear identical in these cases.

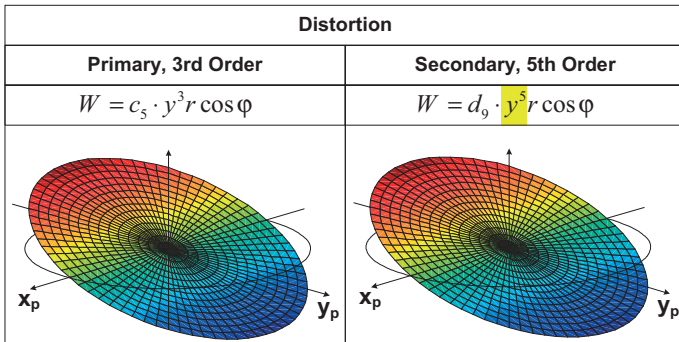


Figure 29-52: Primary and secondary distortion.

For distortion, in a similar way to astigmatism and field curvature, the fifth-order aberration is only influenced by the field variable which is of fifth power for fifth order and of third power for third order. So the figures for fifth and third order appear identical.

29.6.2

Seventh and Higher-order Aberrations

We have seen that there are five third-order and nine fifth-order aberration types.

The number N of aberration types increases with the order n according to

$$N = \frac{(n + 3)(n + 5)}{8} - 1. \tag{29-103}$$

So, there are 14 seventh-order and 20 ninth-order aberrations. From these higher-order aberrations, only those which are easy to understand and easy to recognize (especially spherical aberration and distortion) are used occasionally by the lens designers. As an example, Figure 29-53 shows the wave aberration of a microscope objective. The order of spherical aberration can be estimated by the number of maxima in the curve representing the wave aberration: the central maximum may be due to defocus, so the next minimum indicates at least third order, and the following three extreme values denote at least fifth, seventh and ninth order.

One analytical method of identifying the different orders of an aberration is to calculate the Zernike coefficients, which will be discussed in the next section. In order to analyze this example, let us use this powerful Zernike method. The result is shown in figure 29-54, and from this figure it can be seen that there are even higher orders of spherical aberration but there is an intense contribution of the ninth order. The indicated bar of the first-order spherical aberration of course represents the amount of balancing defocus.

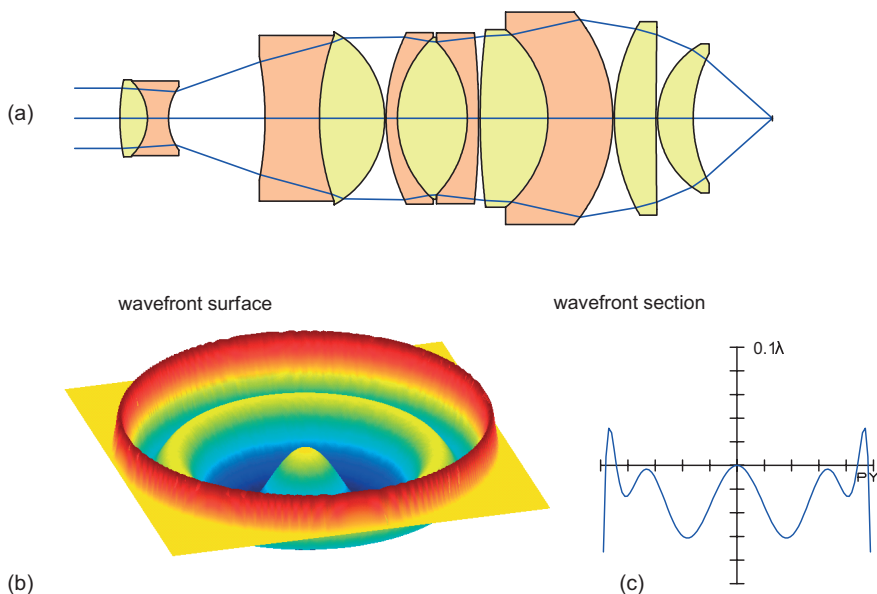


Figure 29-53: Microscope objective with up to ninth-order spherical aberration. (a) Microscope objective (modified version of USP 4588264) [29-10]; (b) wave aberration; (c) OPD-plot (cross-section through (b)).

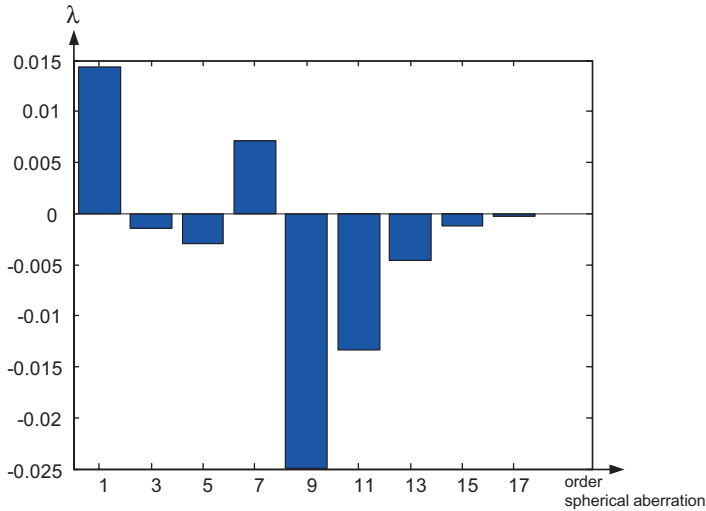


Figure 29-54: Spherical Zernike coefficients of the microscope objective of figure 29-53.

29.7

Zernike Polynomials

The Zernike polynomials are covered in detail in section 11.5.4. As stated there the Zernike polynomials can be used to describe the wave aberration of an optical system for a chosen field. One of the main advantages of the Zernike polynomials is their ability to describe arbitrary wavefronts or wave aberrations without any constraints to symmetry. This property is necessary, for instance, to describe the results of interferometric measurements. With rotationally symmetric optical systems the wave aberration is always symmetric about the meridional plane. Nevertheless, due to current computational power, Zernike polynomials are increasingly used in lens design. The first nine low-order Zernike polynomials are related to the Seidel aberrations as can be seen from table 29-9.

It should be noted that, in general, there is no strict correspondence between Seidel aberrations and the appropriate Zernike polynomial. This means that the Seidel coefficients cannot be calculated from the corresponding Zernike coefficients and vice versa. Only in the case when there are no higher aberrations than third order (Seidel aberrations) does a strict correlation between Seidel and Zernike coefficients exist. This is because the high-order Zernike polynomials also contain lower-order aberrations. The advantage of building the Zernike polynomials with several orders is that the highest order is balanced by the lower orders. For spherical aberration (fifth-order and seventh-order) this can easily be seen from table 29-9.

Table 29-9: Zernike polynomials and wave aberrations based on the power series expansion.

Zernike polynomials		Power series	
$Z_1 = 1$	Piston	a_0	Constant (= 0 per definition)
$Z_2 = r \cos \varphi$	Tilt y	$b_2 \gamma r \cos \varphi$ $c_5 \gamma^3 r \cos \varphi$	Image scale Distortion
$Z_3 = r \sin \varphi$	Tilt x	$\equiv 0$ (due to symmetry)	
$Z_4 = 2r^2 - 1$	Defocus + Piston	$b_1 r^2$ $c_4 \gamma^2 r^2$	Defocus Field curvature
$Z_5 = r^2 \cos 2\varphi$	Astigmatism y	$c_3 \gamma^2 r^2 \cos^2 \varphi$	Astigmatism
$Z_6 = r^2 \sin 2\varphi$	Astigmatism x	$\equiv 0$ (due to symmetry)	
$Z_7 = (3r^3 - 2r) \cos \varphi$	Coma y	$c_2 \gamma r^3 \cos \varphi$	Coma
$Z_8 = (3r^3 - 2r) \sin \varphi$	Coma x	$\equiv 0$ (due to symmetry)	
$Z_9 = 6r^4 - 6r^2 + 1$	Third-order spherical + Defocus + Piston	$c_1 r^4$	Spherical aberration
$Z_{16} = 20r^6 - 30r^4 + 12r^2 - 1$	Fifth-order spherical + Third-order spherical + Defocus + Piston	$d_1 r^6$	Fifth-order spherical
$Z_{25} = 70r^8 - 140r^6 + 90r^4 - 20r^2 + 1$	Seventh-order spherical + Fifth-order spherical + Third-order spherical + Defocus + Piston	$e_1 r^8$	Seventh-order spherical

29.8

Special Aberration Formulae

The power series expansion of the wave aberration as given in eq. (29-8) is the basis for the definition of the aberration types. There are further fundamental aberration formulae, which make statements not limited to just the primary aberrations. The most remarkable results in this area are Abbes sine condition together with its expansions (see 29.8.1) and the Aldis theorem (see section 29.8.4).

29.8.1

Sine Condition and the Offence against the Sine Condition

The Abbe sine condition is discussed in section 11.4 and the different formulations and interpretations are repeated here. As usual let u and U designate the paraxial and finite slope angle of an aperture ray, respectively.

The Abbe presentation reads [29-11]:

$$\frac{\sin U}{u} = \frac{\sin U'}{u'} \quad (29-104)$$

for all rays in the axial aperture pencil. With magnification m (and finite conjugates) the sine condition takes the form

$$m = \frac{nu}{n'u'} = \frac{n \sin U}{n' \sin U'} \quad (29-105)$$

for all rays in the axial aperture pencil, and for an object at infinity with the incoming marginal ray height h (where h is used for both, the paraxial and the finite marginal ray) and focal length f' becomes

$$f' = -\frac{h}{u'} = -\frac{h}{\sin U'} \quad (29-106)$$

for all rays in the axial aperture pencil. In the form of the Lagrange invariant, with paraxial object and image heights y and y' , respectively, the sine condition reads

$$ny \sin U = n'y' \sin U' \quad (29-107)$$

for all rays in the axial aperture pencil.

All the formulae for the sine condition given so far break down for afocal systems. In this case the sine condition is based on the intersection heights of the rays in the axial bundle on the first and last surface of the afocal system. H_1 and H_k are the intersection heights of the finite rays and h_1 and h_k are the corresponding paraxial heights. Then the sine condition reads

$$\frac{H_1}{H_k} = \frac{h_1}{h_k} \quad (29-108)$$

for all rays in the axial aperture pencil.

Under the precondition of zero spherical aberration (for the entire aperture) the fulfillment of the sine condition is a necessary and sufficient requirement for zero linear coma. Linear coma is not only the primary coma but includes all higher orders of coma, that are linear in the field variable. So the linear coma as a wave aberration is given by the sum

$$W_{\text{linComa}} = \sum_m C_m y y_p (x_p^2 + y_p^2)^m = \sum_m C_m y r^{2m+1} \cos \varphi \quad (29-109)$$

where the term with $m = 1$ represents the primary coma and all other terms are higher order linear coma.

The sine condition can be easily derived from the optical sine theorem which is a generalization of the Lagrange invariant and gives the paraxial image height y'_s

made by finite, sagittal rays of arbitrary aperture U in object space and U' in image space, as shown in figure 29-55. The optical sine theorem states

$$n\gamma \sin U = n'\gamma'_s \sin U'. \quad (29-110)$$

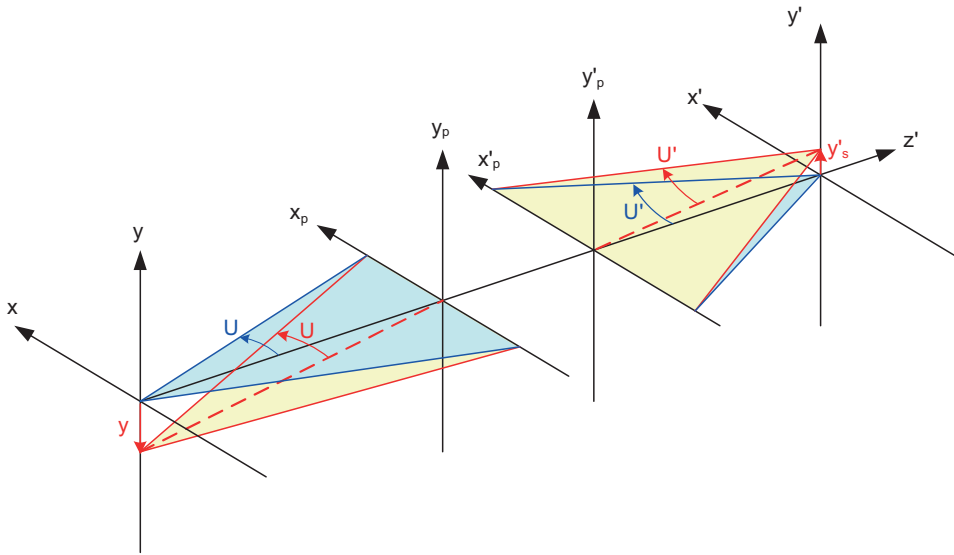


Figure 29-55: Optical sine theorem

It is essential to understand that in the optical sine theorem the image height γ'_s indicates the intersection point of the sagittal rays with apertures U and U' , respectively, and no statement is made about other rays. Because of the restriction to a paraxial field, the aperture angles U and U' are also found in the axial ray pencil.

Now, the “sagittal magnification” m_s is calculated from the optical sine theorem (29-110) as

$$m_s = \frac{\gamma'_s}{\gamma} = \frac{n \sin U}{n' \sin U'}. \quad (29-111)$$

Abbe found that, under the condition of zero spherical aberration, linear coma must be zero, if m_s is constant for all sagittal rays. For constant m_s , when U tends to zero, m_s is

$$m_s = \frac{n \sin U}{n' \sin U'} = \frac{n \sin u}{n' \sin u'} = m \quad (29-112)$$

which is the sine condition as formulated in (29-105). A system with zero spherical aberration and zero linear coma is called aplanatic. The aplanatic surface will be discussed in section 29.8.6 in detail.

In practice, the situation that the spherical aberration is exactly zero and the sine condition is strictly fulfilled, seldom occurs. But based on continuity arguments it is obvious that when the spherical aberration is small and the sine condition is “almost” fulfilled, the linear coma must also be small. A quantitative predication for this situation is given by the “Offense against the Sine Condition”, OSC. The OSC gives the transverse sagittal linear coma relative to the height of the paraxial chief ray. Figure 29-56 shows the quantities used in the OSC definition. M is the intersection point of the marginal ray with the optical axis, so “ s' ” is the longitudinal spherical aberration. Q is the intersection point of the chief ray in the plane defined by M and B'_s is the intersection point of the sagittal rays of the oblique bundle. As the field is assumed to be small, B'_s is also in the plane defined by M . The transverse distance QB'_s is the sagittal coma.

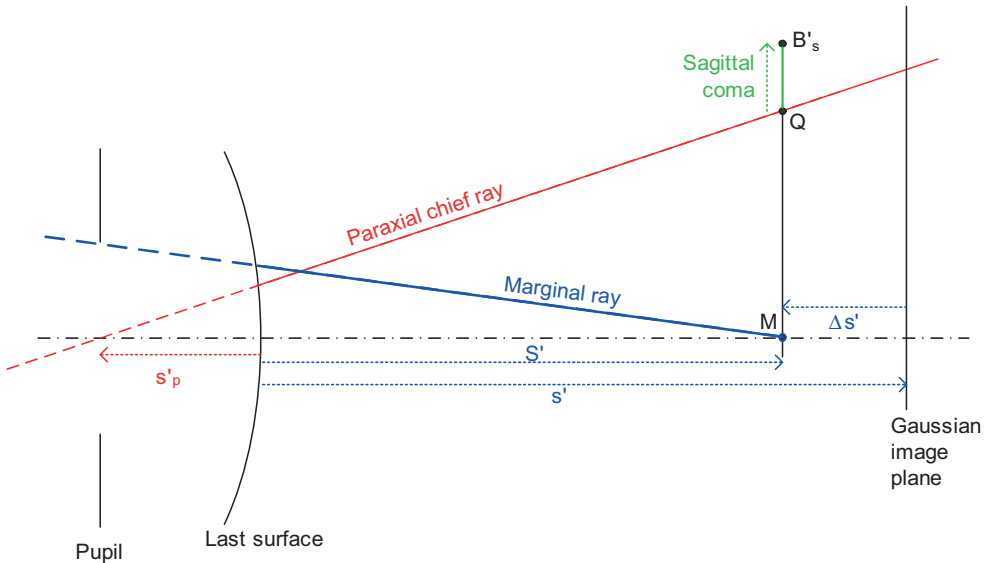


Figure 29-56: Quantities to define the OSC.

$$OSC = \frac{QB'_s}{MQ} \quad (29-113)$$

with the help of the optical sine theorem (29-110) this can easily be shown to be [29-2], [29-12]

$$OSC = \frac{u' \sin U}{u \sin U'} \left(1 - \frac{\sigma s'}{S' - s'_p} \right) - 1. \quad (29-114)$$

To determine OSC (i. e. the relative transverse sagittal coma) only the finite marginal ray, the paraxial marginal ray, and the paraxial chief ray must be traced. For the object at infinity the paraxial as well as the finite marginal ray have to be calculated with the same incidence height on the object side. Then the OSC reads

$$OSC = \frac{u'}{\sin U'} \left(1 - \frac{s'}{S' - s'_p} \right) - 1. \quad (29-115)$$

This form can also be used when the paraxial angle u is set equal to $\sin U$.

Whenever the OSC for an imaging optical system is known from some calculation the transverse coma for small fields (linear coma) can easily be calculated from (29-113). QB'_s represents the sagittal coma and, according to eqs (29-60) and (29-61) for the primary aberration, the tangential coma is three times this value. Eqs (29-114) and (29-115) clearly show the influence of both, the sine condition as well as the spherical aberration, on the OSC. Only for a system which is telecentric in image space does the spherical aberration have no influence, because the exit pupil is at infinity, $s'_p = \infty$ and that makes the bracketed term in (29-114) and (29-115) equal to one.

29.8.2

Herschel Condition

Let us consider an optical system which images an axial object point monochromatically, without aberration. The sine condition (29-105) can then be understood as the condition that the image remains free of aberrations when the object, and accordingly the image, is shifted a small distance perpendicular to the optical axis. "Small" means that aberrations, which depend quadratically or to an even higher power on the field coordinate, can be neglected. The Herschel condition [29-7], [29-8] is the corresponding condition for a small object (and image) shift along the axis. In the formulation analog to the sine condition (29-105) the Herschel condition reads

$$\bar{m} = \frac{n \sin^2(U/2)}{n' \sin^2(U'/2)} \quad (29-116)$$

for all rays in the axial aperture pencil.

\bar{m} designates the longitudinal magnification

$$\bar{m} = \frac{\partial z'}{\partial z} = \frac{n'}{n} m^2 \quad (29-117)$$

as shown in figure 29-57.

It can be shown that the Abbe sine condition and the Herschel condition can only be satisfied simultaneously in a trivial case which is when the system is a plane mirror with $m = 1$ and $\bar{m} = 1$. So, in general, an optical system cannot image a small volume without aberrations.

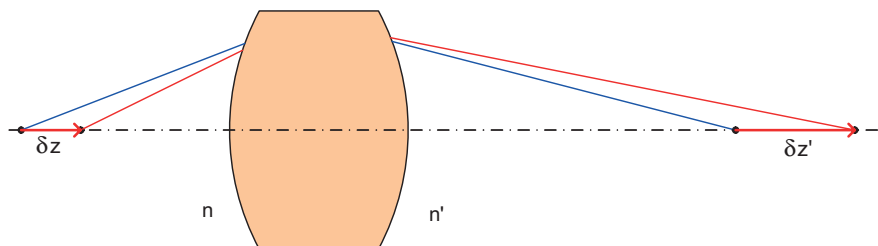


Figure 29-57: Longitudinal magnification.

29.8.3

Aplanatism and Isoplanatism

The sine condition and the OSC are the basis for understanding aplanatism and isoplanatism. A system is called aplanatic when both the spherical aberration and the linear coma are zero. The OSC is sometimes also called the isoplanatic condition, and a system is called isoplanatic when the OSC equals zero [29-12]. The significance of isoplanatism is the stationarity of the aberration with regard to a small shift in the field. This means that when the object point is shifted by a small amount, the aberration at the new image point will be the same as it was before the shift. For the axial image point the only aberration which may change with a small image shift, is the linear coma. So the system is isoplanatic (on the axis) if the OSC equals zero. In the theory of image formation, isoplanatism is an important principle and the term isoplanatism is generally not restricted to the axial image point as it means local stationarity of the aberrations over an expanded field. For the axial case, the OSC is a very simple method of checking for isoplanatism. Aplanatism, in general, means the same stationarity but with zero aberrations.

29.8.4

Aldis Theorem

The Aldis theorem is a noteworthy aberration formula which calculates the finite transverse aberrations dx' and dy' in the Gaussian image plane of an arbitrary ray with respect to the associated paraxial chief ray [29-2], [29-13]. Currently this result can be obtained by simple raytracing. The outstanding feature of Aldis theorem is that these formulae give the finite aberrations as a sum of the surface contributions. This is extremely useful for locating the source of an aberration, especially when it is used in parallel with the Seidel third-order formulae, because the Aldis theorem gives the entire aberration, including all orders.

Let the system to be analyzed have k spherical surfaces. To use the Aldis theorem only two rays must be traced through the system, one paraxial ray and one finite ray. The paraxial ray emerging from the axial object point with an arbitrary aperture (the aperture cancels out in the formulae) provides the refraction invariant (see eq. (29-21)) $A_j = n_j i_j = n_j (h_j c_j + u_j)$ at each surface of the system and also the final aperture angle u'_k . Further, the Lagrange invariant H must be calculated based on the chosen paraxial aperture and also based on the field, for which the aberrations are to be calculated. For instance, H can be calculated in the image space as $H = n'_k u'_k y'$ with the paraxial image height y' .

The finite ray emerging from the object related to y' is the ray for which the transverse aberrations dx' and dy' are to be calculated. This finite ray provides the incidence points on each surface with coordinates x_j, y_j, z_j and the direction cosines L_j, M_j, N_j in each space.

Let $\sigma(x) = x' - x$ designate the increment on refraction. With these quantities the transverse aberrations dx' and dy' are, according to the Aldis theorem,

$$dx' = \frac{-1}{n'_k u'_k N'_k} \sum_{j=1}^k \left[A_j z_j^{\prime\prime} L_j + \frac{A_j x_j}{N'_j + N_j} \prime \left(L_j^2 + M_j^2 \right) \right], \quad (29-118)$$

$$dy' = \frac{-1}{n'_k u'_k N'_k} \sum_{j=1}^k \left[A_j z_j^{\prime\prime} M_j + \frac{A_j y_j - H}{N'_j + N_j} \prime \left(L_j^2 + M_j^2 \right) \right]. \quad (29-119)$$

The proof of the Aldis theorem is not really difficult but it is rather lengthy and it does not give any additional insight into this interesting matter [29-2]. But there is a useful specialization of the Aldis theorem, which relates to longitudinal spherical aberration only, see the next paragraph.

The difference between the Seidel surface contributions and the Aldis surface contributions can be demonstrated by a rather simple example. For this purpose the spherical aberration of an achromat, see figure 29-58, will be analyzed.

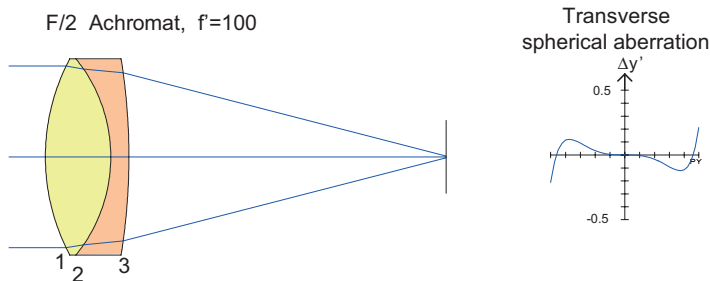


Figure 29-58: An achromat.

In table 29-10 the Seidel surface contributions according to eq. (29-23) as well as the Aldis surface contributions according to eq. (29-119) are listed. As the Aldis contributions depend on the chosen ray, in table 29-10 the Aldis contributions are given for two rays, for full aperture $\rho = 1$ and for half the aperture $\rho = 0.5$.

Table 29-10: Seidel and Aldis surface contributions for the achromat in figure 29-58.

Surface	Seidel S_j	Aldis $\prime \gamma'(\rho = 0.5)$	Aldis $\prime \gamma'(\rho = 1)$
1	0.455	-0.118	-1.047
2	-0.730	0.209	2.604
3	0.577	-0.151	-1.345
Sum	0.302	-0.060	0.212

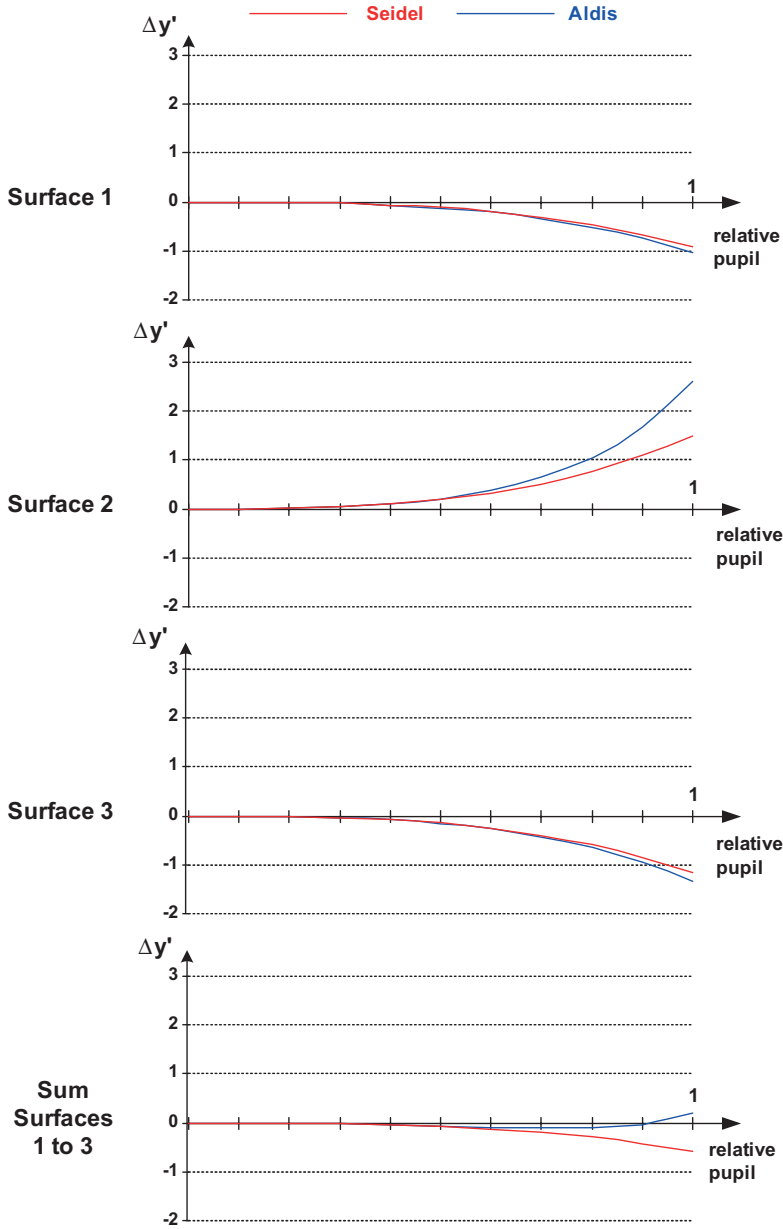


Figure 29-59: Aldis versus Seidel surface contributions for the achromat shown in figure 29-58.

Because Seidel and Aldis contributions are quite different types of quantities, a direct comparison is difficult. Much more insight is gained when the corresponding transverse aberrations are calculated from the Seidel coefficients according to eq. (29-59). The result is shown in figure 29-59. It can be seen that the main difference between Seidel and Aldis occurs at surface 2, the cemented surface. As can be seen from figure 29-58 at this surface the marginal ray has the largest incidence angles and so surface 2 contributes the largest amount of fifth-order aberration. The Aldis curve for the sum over all surfaces at the bottom of figure 29-59 exactly represents the transverse spherical aberration as is also shown in figure 29-58.

29.8.5

Spherical Aberration, a Surface Contribution Formula

A simple formula for longitudinal spherical aberration on a surface-by-surface basis can be derived from the Aldis theorem (29-118) and (29-119) or it can be derived directly [29-14]. The formula reads

$$ds' = \frac{-1}{n'_k u'_k \sin U'_k} \sum_{j=1}^k A_j'' Q_j \tag{29-120}$$

with the same nomenclature as in (29-118) and (29-119). ds' is the total longitudinal spherical aberration of the ray with the finite aperture angle U'_k in image space and Q_j designates the perpendicular distance from the surface vertex to the ray at the surface number j as shown in figure 29-60. To make use of this surface contribution formula (29-120) from the axial pencil a finite and an arbitrary paraxial ray must be traced.

From (29-120) it can be seen that the surface number j does not introduce spherical aberration if $A_j = n_j i_j$ or $'' Q_j$ is zero. For further discussion it is useful to modify eq. (29-120) so that $'' Q_j = Q'_j - Q_j$ is expressed as a product of different quantities that can vanish independently. Let \overline{AP} designate the chord from the surface vertex A to the intersection point P of the ray, as shown in figure 29-60.

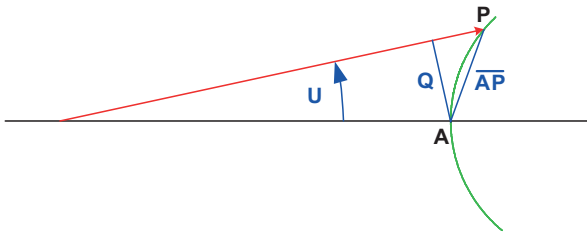


Figure 29-60: Chord \overline{AP} and perpendicular Q , to the ray.

By using the ray tracing formulae and pure geometrical arguments it is not difficult [29-14] to prove that

$$'' Q_j = Q'_j - Q_j = 2\overline{AP}_j \sin \frac{I'_j - I_j}{2} \sin \frac{I'_j + U_j}{2} \tag{29-121}$$

where I_j and I'_j represent the finite incidence angles as usual.

Now, together with eq. (29-120) it follows that there are four cases where a spherical surface does not introduce spherical aberration:

1. $\overline{AP} = 0$
2. $ni = 0$
3. $I' - I = 0$
4. $I' + U = 0$.

The discussion of these four cases follows.

1. The case $\overline{AP} = 0$

The calculated finite aperture ray hits the surface at the vertex as shown in figure 29-61. It is obvious that the surface cannot introduce spherical aberration to this ray. If the paraxial object and image location is on the surface vertex and if there is no spherical aberration in the incoming axial pencil, then for each ray in this pencil one obtains

$$U = -I \text{ and } U' = -I'. \quad (29-122)$$

To check the sine condition we have

$$\frac{n \sin U}{n' \sin U'} = \frac{n \sin I}{n' \sin I'} = 1 \quad (29-123)$$

and with the corresponding paraxial quantities the magnification in this case is

$$m = \frac{nu}{n'u'} = \frac{ni}{n'i'} = 1 \quad (29-124)$$

so that the sine condition is fulfilled and we have an aplanatic case.

2. The case $ni = n'i' = 0$

This is the only paraxial condition of the four cases. There are two possibilities. First, $i = i' = 0$, which means that the location of the intermediate object and intermediate image is the center of curvature of the surface. So the paraxial ray does not change the direction when passing the surface. This situation looks similar to case 3, but here the finite incoming axial pencil may have a lot of spherical aberration as we have the paraxial condition. That means the finite marginal rays may not be perpendicular to the surface. Nevertheless, as long as the condition $i = i' = 0$ is fulfilled, the surface contribution to the spherical aberration will be zero. This configuration is called the concentric surface. The paraxial intersection lengths are equal to the radius of curvature,

$$s' = s = r, \quad (29-125)$$

and the magnification for the concentric surface is

$$m = \frac{ns'}{n's} = \frac{n}{n'}. \quad (29-126)$$

If there is no spherical aberration in the incoming axial pencil, of course the sine condition is fulfilled and again we have an aplanatic case. There is also a second possibility; that is the trivial case, $n = n'$ where the surface is a dummy surface and has no effect on the rays.

3. The case $I' - I = 0$

Similar to case 2, but here related to a finite aperture ray. Again we have the two possibilities, the first is: $I' = I = 0$, the perpendicular incidence of the chosen aperture ray, see figure 29-61. The surface contribution to spherical aberration for this special finite ray is zero. But for all other rays in the pencil for which the condition $I' = I = 0$ is not fulfilled (due to spherical aberration in the incoming pencil) the surface contribution will not generally be zero. The second possibility is the trivial one, $n = n'$, the surface is a dummy surface and there is no effect on the rays from this surface.

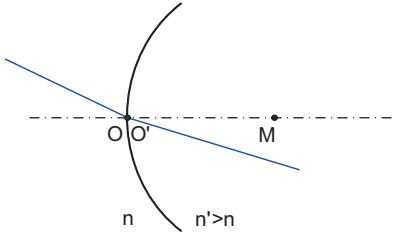
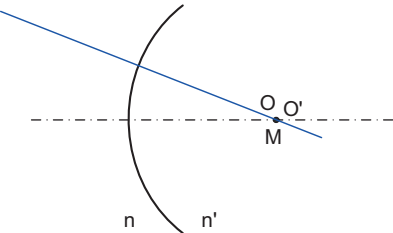
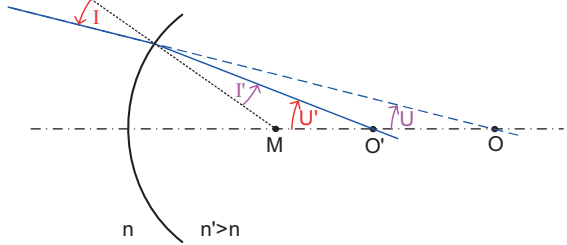
<p>1. Case: Object O and image O' at surface vertex $m = 1$</p>	
<p>2. Case: Paraxial ray 3. Case: Finite ray Object O and image O' at the center of curvature M $m = \frac{n}{n'}$</p>	
<p>4. Case: $I' + U = I + U' = 0$ "Aplanatic" condition $m = \left(\frac{n}{n'}\right)^2$</p>	

Figure 29-61: Cases with zero spherical aberration.

4. The case $l' + U = 0$

This is the most important case, which turns out to describe the actual aplanatic condition, and which plays an important role in lens design. But in the first instance this condition is again a condition for a chosen finite aperture ray, similar to that discussed in cases 1 and 3. In this generality we have the statement of a zero spherical aberration contribution only for the chosen aperture ray. But when the incoming axial pencil has no (or in practice at least low) spherical aberration, this configuration becomes highly interesting. A more detailed discussion for this case follows in the next paragraph.

From figure 29-61 it can be seen that, for a given surface with radius r and refractive indices n and n' , there are three special object positions which exhibit zero spherical aberration: Case 1: object at the vertex of the surface; Case 2: object at the center of curvature, that is the concentric position; Case 3: object at the aplanatic object point. For all other object positions the spherical aberration must be unequal to zero, as we used a rigorous mathematical case differentiation to find these three conditions. In figure 29-62 the spherical aberration is shown as a function of the object position. As an example, a surface with a positive power and positive radius is chosen as in figure 29-61. The radius is $r = 100$, the aperture angle U is held constant. The position of the aplanatic point depends on the ratio of the refractive indices. In the example shown,

$$S = \frac{n + n'}{n} r = 2.5r.$$

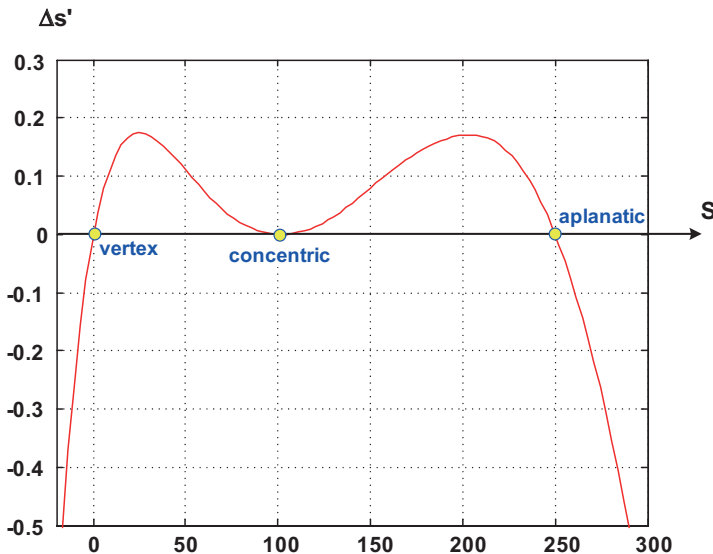


Figure 29-62: Spherical longitudinal aberration $\Delta s'$ as a function of the object position S for a constant aperture angle with $\sin U = 0.2$. The surface radius is $r = 100$ and the refractive indices are $n = 1$ and $n' = 1.5$.

However, figure 29-62 must not be misinterpreted. In most situations, when the intersection length S is not so close to the surface vertex as in the main range of this figure, a surface with positive power exhibits under-correction as can be seen in figure 29-62 for $S < 0$ and for $S > 2.5r$. It is interesting to note that, for an intersection length, which is about 80% of the aplanatic intersection length, the spherical aberration possesses a maximum over-correction. Occasionally this can be used in the design process.

29.8.6

Aplanatic Surface and Aplanatic Lens

If the condition $I' + U = 0$ is fulfilled for each ray in the axial pencil, then the aplanatic surface is described. From this condition several conclusions relating to the intersection lengths S and S' of the finite aperture rays, the magnification m , and the radius of curvature r can be derived from the raytracing formulae (see section 11.6.2):

$$I' + U = I + U' = 0, \quad (29-127)$$

$$S = \frac{n + n'}{n} r \quad \text{and} \quad S' = \frac{n + n'}{n'} r. \quad (29-128)$$

Eq. (29-128) shows that any given surface with parameters r , n , n' , can work as an aplanatic surface with the appropriate object and image position. These object and image points are called the aplanatic points of the surface. On the other hand,

$$r = \frac{nS}{n + n'} = \frac{n'S'}{n + n'} = \frac{SS'}{S + S'}. \quad (29-129)$$

Eq. (29-129) shows that for any given object intersection length S (or image intersection length S') and given refractive indices n and n' the surface radius of curvature r can be determined, in such a way that the surface will be an aplanatic one:

$$nS = n'S', \quad (29-130)$$

$$\frac{1}{S} + \frac{1}{S'} = \frac{1}{r}, \quad (29-131)$$

$$m = \frac{ns'}{n's} = \frac{nS'}{n'S} = \left(\frac{n}{n'}\right)^2, \quad (29-132)$$

$$\frac{n \sin U}{n' \sin U'} = \frac{n \sin I'}{n' \sin I} = \left(\frac{n}{n'}\right)^2 = m. \quad (29-133)$$

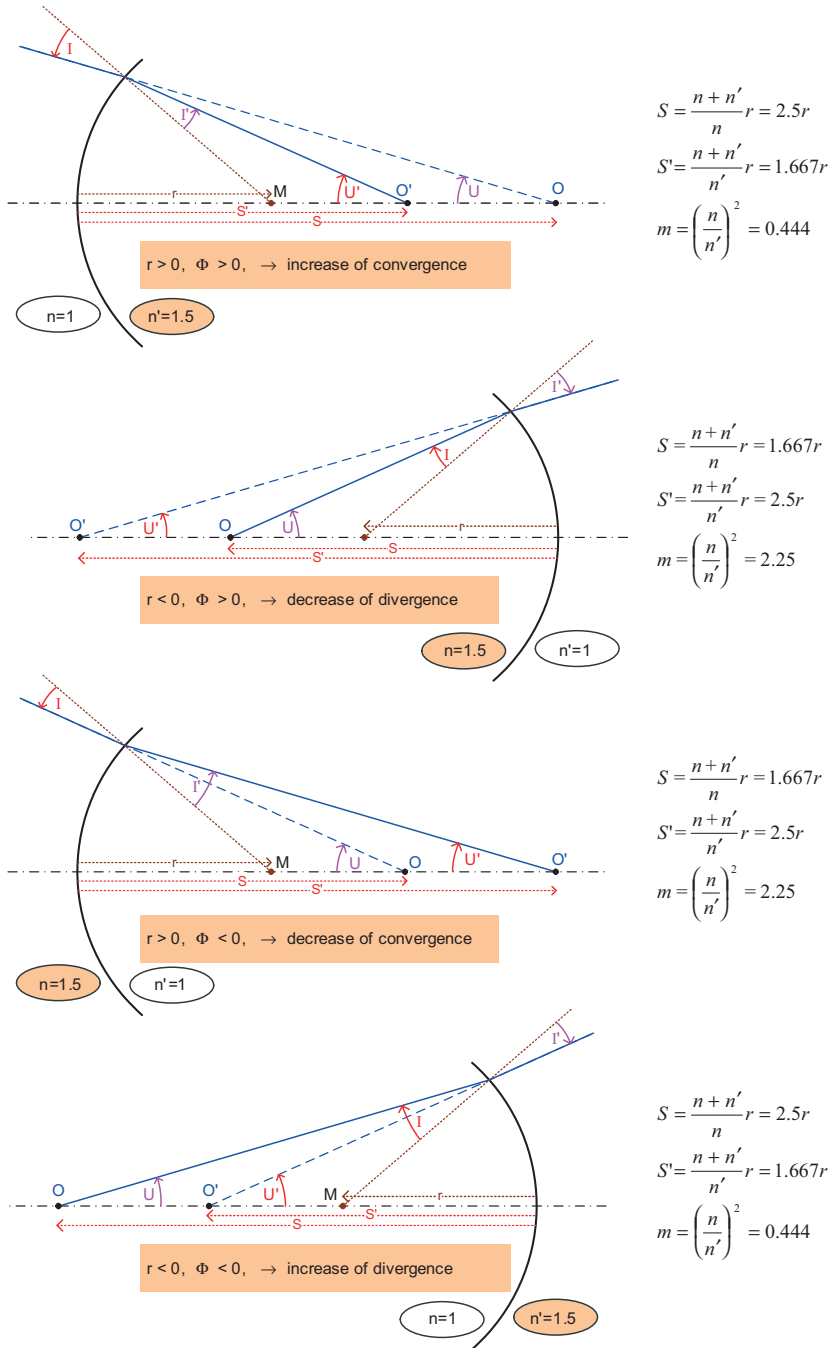


Figure 29-63: Aplanatic surfaces with positive and negative radii r , with positive and negative power δ .

Eq. (29-133) represents the fulfillment of the sine condition, so that the designation “aplanatic surface” is justified. In figure 29-63 aplanatic surfaces are shown for surfaces with positive and negative radii r and, due to the arrangement of the refractive indices, for positive and negative powers δ . As seen from eqs. (29-128) the object and image distance S and S' both have the same sign as the surface radius r . As the object and image are always on the same side of the surface the imaging of an aplanatic surface is always virtual. This can also be seen from the magnification m , which is always positive, see eq. (29-132). As shown in figure 29-63, the effect of an aplanatic surface is to change the magnification, or in other words, to increase or to decrease the convergence or the divergence, as the case may be.

It should be understood that the aplanatic condition $l' + U = 0$ makes the surface contribution to the spherical aberration zero only for the rays which exactly fulfill the condition. If there is some spherical aberration in the incoming axial ray bundle, most of the rays cannot fulfill the condition and so these rays will have some spherical aberration at the surface considered. If there is no spherical aberration in the incoming ray bundle, all rays can fulfill the condition and then there will be no contribution to spherical aberration at the surface considered.

To construct a single aplanatic lens the combination of concentric and aplanatic surfaces can be used. The possible configurations are discussed in section 11.6.3 as shown in figure 11-22. It should be kept in mind that such aplanatic surfaces as well as aplanatic lenses make aplanatic imaging suitable only for the appropriate object and image position.

29.9

Literature

- 29.1 I. P. A. Agurok, Double expansion of wave-front deformation in Zernike polynomials over the pupil and the field-of-view of optical systems: Lens design, testing and alignment, Proc. SPIE 3430, 80 (1998).
- 29.2 W.T. Welford, Aberrations of Optical Systems (Adam Hilger, Bristol, 1986).
- 29.3 C. G. Wynne, A Comprehensive First-order Theory of Chromatic Aberration Secondary Spectrum Correction without Special Glasses, J. Mod. Opt. 25, 627 (1978).
- 29.4 W. E. Wöltche, Structure and Image forming properties of asymmetrical wide angle lenses for 35-mm Photography, Appl. Opt. 7, 343 (1968).
- 29.5 Mori Ikuo, US Patent 4203653, Inverted telephoto type wide angle lens system. (1980).
- 29.6 C. G. Wynne, Primary Aberrations and Conjugate Change, Proc. Phys. Soc. Lond. 65B, 429 (1952).
- 29.7 V. N. Mahajan, Optical Imaging and Aberrations, Part I (SPIE Press, Bellingham, 1998).
- 29.8 C. Hofmann, Die optische Abbildung (Akademische Verlagsgesellschaft, Leipzig, 1980).
- 29.9 Horimoto Mitsuki, US Patent 4256373, Fish eye lens system. (1981).
- 29.10 Shimizu Yoshiyuki, US Patent 4588264, Microscope objective lens. (1986).
- 29.11 J. Braat, The Abbe sine condition and related imaging conditions in geometrical optics, Proc. SPIE 3190, 59 (1997).
- 29.12 M. Shibuya, The relation between offence against the sine condition in the presence of spherical aberration and lateral aberration, Opt. Commun. 70, 12 (1989).
- 29.13 A. Cox, A System of Optical Design (Focal Press, New York, 1964).
- 29.14 R. Kingslake, Lens Design Fundamentals (Academic Press, New York, 1978).