Aberrations of optical systems

W. T. Welford, FRS

Aberrations of optical systems

W. T. Welford, FRS
The Blackett Laboratory,
Imperial College of Science and Technology,
University of London

Adam Hilger Ltd
Bristol and Boston

Preface and Acknowledgements

Aberration theory as used in optical design has changed considerably since 1974 when my book "Aberrations of the Symmetrical Optical System" (Academic Press) appeared. Among these changes are the use of non-axially symmetric systems and diffractive optical elements in quite complex designs such as head-up displays and the increasing use of scanning systems with laser illumination. The present book is based to a considerable extent on that of 1974 and I acknowledge the use of much material from that book; however, I have changed much and added material on the subjects mentioned above and others suggested by colleagues. It is a pleasure to acknowledge help and suggestions given by these friends, including Prudence Wormell, Richard Bingham, Charles Wynne, Michael Kidger and many others. I am grateful also to Jim Revill of Adam Hilger Ltd for his advice and help in the preparation of the book.

W. T. Welford

Contents

1 Optical Systems and Ideal Optical Images

The symmetrical optical system, i.e. a system with symmetry about an axis of revolution, is the type of system most frequently met as a design problem; this includes systems folded by means of plane mirrors or prisms, since it is trivial to unfold them for optical design purposes. However, non-symmetrical systems are not uncommon, e.g. some kinds of spectacle lens, spectrographic systems, anamorphic projection systems and systems containing holographic optical elements. In this book we shall be mainly concerned with symmetrical systems but some discussion of non-symmetrical systems will be given, chiefly in connection with raytracing.

1.1 Initial assumptions

The treatment will be based mainly on the geometrical optics model but there will be occasional references to physical optics in the form of scalar wave theory; this is needed for dealing with aberration tolerances. In geometrical optics the essential concept is the ray of light; in this chapter we assume this as an intuitive notion, deferring more precise definition to Chapter 2. It is then possible to formulate definitions of ideal image formation using only the concept of rays and the assumption that to one ray entering the system there corresponds one and only one ray emerging. We do not at this stage invoke the laws of reflection and refraction, and we make no assumptions about how the transformation from object to image space is accomplished: i.e. there might be non-spherical surfaces, media of continuously varying refractive index, etc., in the system. However, it is convenient to assume that the entering and emerging rays are straight line segments, or in physical terms that there are clearly defined regions in the object and image spaces in which the respective refractive indices are constant. Ideal image formation for a general system then means that a pencil of rays from a point in object space becomes a pencil also passing through a single point in image space and that this holds for some one- or two-dimensional object surface. This does not get us very far but if we assume a symmetrical system we can obtain many other properties of ideal image formation to which the performance of a real well-corrected system should approximate.

The first notions of ideal image formation through symmetrical systems are due to A. F. Möbius (1855). A few years later James Clerk Maxwell (1856, 1858) formalized the concept of an ideal system without invoking any physical image-forming mechanism. It is essentially Maxwell's concept which we describe in this chapter.

1.2 Ideal image formation in the symmetrical optical system

Take the z-axis of a right-handed Cartesian coordinate system as the axis of revolution of a symmetrical optical system, as in Fig. 1.1, and the y-axis in the plane of the diagram: the origin O is taken as any convenient point on the axis.

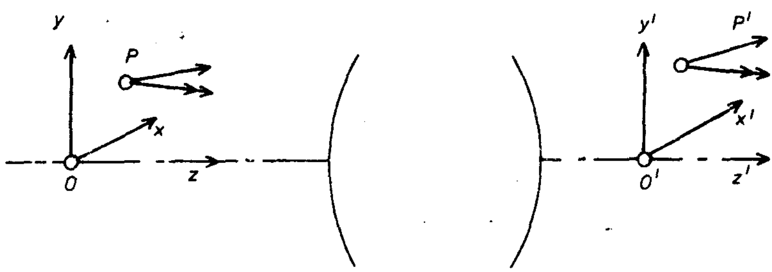

FIG. 1.1. Coordinates for the symmetrical optical system

If, as is customary, the light is supposed to travel from left to right then this coordinate system is in the object space and we take a similar system $O'x'y'z'$ in the image space, the respective axes being parallel to each other.

All points and rays in object space are referred to $Oxyz$ and those in image space to $O'x'y'z'$. The rays are shown as straight line segments but we introduce immediately the generalization that they are to be regarded as extending indefinitely in either direction; thus the object space extends right through the optical system and through image space and similarly image space is extended infinitely in both directions. This is an essential convention for dealing with the details of image formation in the intermediate spaces of a system, where the image from one optical element is the object for the next, since this image-*cum*-object is very frequently virtual.

Ideal image formation from the x–y plane to the x'–y' plane can then be defined as follows. All rays through any point P on the x–y plane must pass through one point P′ on the x'–y' plane and the coordinates (x', y') are proportional to (x, y); the constant of proportionality, which is, of course, the magnification, depends on the nature of the optical system and on the axial positions of the two planes. This can be summarized by saying that any figure on the x–y plane is perfectly imaged as a geometrically similar figure on the x'–y' plane. It will be shown that if there are two such pairs of conjugate planes then any plane in object space is imaged ideally on another plane in image space.

There is considerable interest in examining this Maxwellian ideal image formation because, as will be seen in Chapter 3, the image formation in any real symmetrical system approximates to the ideal in a narrow region sufficiently close to the optical axis. We shall therefore study this ideal image formation in more detail.

Let Oxy and O$_1x_1y_1$ be two planes in object space and let O′$x'y'$ and O$'_1x'_1y'_1$ be the corresponding planes in image space. We suppose the image formation to be perfect between these pairs of planes; thus, in Fig. 1.2, if $\boldsymbol{a}$

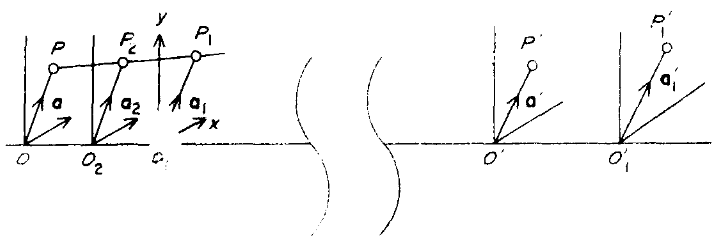

FIG. 1.2. Ideal image formation

and $\boldsymbol{a}_1$ are two-dimensional vectors in the planes Oxy and O$_1x_1y_1$ from the origins to points P and P$_1$, the corresponding vectors in the image planes will be given by

$$\left.\begin{aligned} \boldsymbol{a}' &= m\boldsymbol{a} \\ \boldsymbol{a}'_1 &= m_1\boldsymbol{a}_1 \end{aligned}\right\} \qquad (1.1)$$

where m and m_1 are appropriate magnification factors between these planes. These equations imply that *all* rays through P pass through P′ and similarly for P_1 and P_1', so that they express the assumption of perfect image formation between the two planes.

Now we consider points P_2 on a third plane $O_2x_2y_2$; we enquire whether all rays through P_2 also pass through one point P_2' in image space and, if so, whether this point always lies on one plane perpendicular to the axis and is related to P_2 by an equation similar to eqn (1.1). We have

$$\boldsymbol{a}_2 = (z_2/z_1)(\boldsymbol{a}_1 - \boldsymbol{a}) + \boldsymbol{a} \tag{1.2}$$

where z_1 and z_2 are the coordinates of O_1 and O_2 with respect to O, this being true for all points P and P_1 which are on a ray through P_2. If a point P_2' with the above properties exists, then we must also have

$$\boldsymbol{a}_2' = (z_2'/z_1')(\boldsymbol{a}_1' - \boldsymbol{a}') + \boldsymbol{a}' \tag{1.3}$$

and

$$\boldsymbol{a}_2' = m_2\boldsymbol{a}_2. \tag{1.4}$$

Equations (1.3) and (1.2) can be rearranged as

$$\boldsymbol{a}_2' = (m_1z_2'/z_1')\boldsymbol{a}_1 + m(1 - z_2'/z_1')\boldsymbol{a} \tag{1.3'}$$

$$\boldsymbol{a}_2 = (z_2/z_1)\boldsymbol{a}_1 + (1 - z_2/z_1)\boldsymbol{a} \tag{1.2'}$$

and these will be consistent with eqn (1.4) if we can find z_2' and m_2 to satisfy

$$m_1z_2'/z_1' = m_2z_2/z_1 \quad \text{and} \quad m(1 - z_2'/z_1') = m_2(1 - z_2/z_1); \tag{1.5}$$

clearly this can be done since we have two equations and two unknowns, and the values will hold good for all points P_2. Thus we have shown that ideal image formation for two pairs of conjugate planes implies ideal imagery for all other pairs.

1.3 Properties of an ideal system

We can now develop many properties of ideal systems which are to be used in the paraxial approximation. For this purpose, we indicate the optical system schematically as in Fig. 1.3, but it must be understood that it may extend for a considerable distance along the axis and the constructions to be explained may take place inside the physical system, i.e. with virtual parts of rays.

In Fig. 1.3, let $\boldsymbol{r}_1$ be a ray from the point at infinity on the axis in object space, i.e. a ray parallel to the axis. It meets the axis in image space at some point F′; this must be the image of the axial point at infinity in object space since two rays pass through both these points, namely the ray $\boldsymbol{r}_1$ and the ray

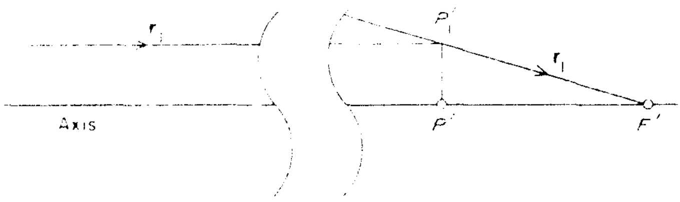

FIG. 1.3. Principal focus and principal point

along the axis, and we are assuming ideal image formation.† The point F′ is called the second principal focus or image-side principal focus.

Let the segment of ray r_1 in object space be produced until it meets the segment in image space at P_1'; a plane normal to the axis through P_1' meets the axis at P′, the second or image-side principal point, the plane itself being the second or image-side principal plane.

Similar constructions and definitions lead to the object-side principal focus and principal point. These constructions can be made on the same diagram with the rays r_1 and r_2 parallel to the axis chosen to be at equal distances from

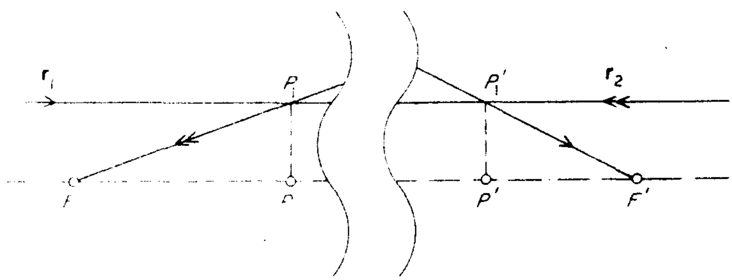

FIG. 1.4. Unity magnification between the principal points

the axis, as in Fig. 1.4, which shows all four points F, F′, P, P′. The two segments of r_1 meet at P_1' and those of r_2 at P_1. Both rays r_1 and r_2 pass through P_1 and P_1', so these points must be object and image; furthermore, using the properties of ideal image formation, the planes normal to the axis through P_1 and P_1' must be conjugates and since by construction $PP_1 = P'P_1'$ the magnification between these planes must be unity. For this reason, P and P′ are sometimes called unit points.

By their definitions, F and P are always in object space and F′ and P′ are always in image space; however, it may very well happen, for example, that F′ and P′ may physically lie to the left of the system, although they are still

† We defer until Section 3.4 the special case in which the ray r_1 emerges parallel to the axis.

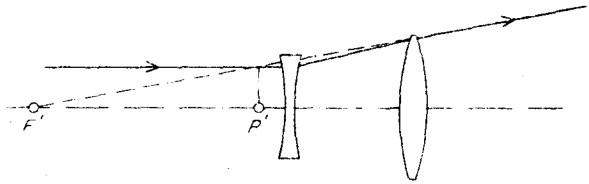

FIG. 1.5. An optical system with P' and F' physically on the left of the system

in image space. Fig. 1.5 shows a system consisting of two components in which this occurs.

The distance P'F' is the second or image-side focal length, denoted by f', and PF $= f$ is the object-side focal length. These magnitudes take signs according to the order of the letters in the definition in relation to the positive z direction, so that in Fig. 1.4 f would be negative and f' positive.

The four points F, F', P, P', along the axis fix the properties of the ideal optical system completely. We can use them in a construction to find the position and size of the image of any object, as in Fig. 1.6 where the optical system is represented merely by these four points and the principal planes. Let the object be OO_1 and let the ray r_1 be drawn through O_1 parallel to the axis to meet the first principal plane in P_1; it must emerge through P_1' at the same distance from the axis, and pass through F'. In the same way, the ray r_2 from O_1 through F is drawn through P_2 and P_2'. The point O_1' in which the image side segments of r_1 and r_2 meet must be the image of O_1, and O' must therefore be on the perpendicular from O_1' to the axis. This construction will be recognized as a simple generalization of the elementary construction for image formation by a thin lens.

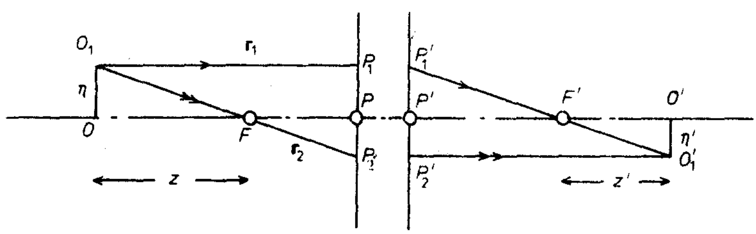

FIG. 1.6. Geometrical construction for conjugates

Figure 1.6 also yields simple formulae relating the positions and sizes of the object and image. Let η and η' be the object and image heights; these take signs according to the y-axis in the coordinate system (Fig. 1.1), so that in Fig. 1.6 η is positive and η' negative. Let FO $= z$, F'O' $= z'$, so that these quantities specify the axial conjugate positions; they are taken as directed segments with signs according to the z-axis, so that in Fig. 1.6 z is negative and z' is positive. From the similar triangles FOO_1 and FPP_2 we have

$$\eta'/\eta = -f/z \tag{1.6}$$

and likewise from $F'O'O'_1$ and $F'P'P'_1$,

$$\eta'/\eta = -z'/f'. \tag{1.7}$$

Combining eqns (1.6) and (1.7), we have

$$zz' = ff'; \tag{1.8}$$

this is known as a conjugate distance equation, since it relates z and z'; it is generally called Newton's conjugate distance equation. Isaac Newton gave it for a single surface ("Opticks", Book 1, Part 1, axiom 6, Dover 1952, based on the 4th edition 1730).

At the same time, we have obtained important expressions for the magnification $m = \eta'/\eta$ in eqns (1.6) and (1.7); these are generally written

$$z = -f/m, \qquad z' = -mf'. \tag{1.9}$$

It is also useful to have a conjugate distance equation in terms of the distances of object and image from the respective principal planes. Let PO $= l$, P′O′ $= l'$, again with signs implied by the fact that PO is a directed segment; thus l is negative and l' positive in Fig. 1.6. We have

$$l = z + f, \qquad l' = z' + f'; \tag{1.10}$$

if the values of z and z' are substituted from eqn (1.9) and m is eliminated, we obtain

$$\frac{f'}{l'} + \frac{f}{l} = 1, \tag{1.11}$$

the required equation. We also have, analogous to eqn (1.9),

$$l = f\left(1 - \frac{1}{m}\right), \qquad l' = f'(1 - m) \tag{1.12}$$

and so

$$m = -\frac{l'f}{lf'}. \tag{1.13}$$

The considerable difference in form between the two conjugate distance equations, eqns (1.8) and (1.11), is because in eqn (1.11) the conjugates are referred to origins in object and image space which are themselves conjugates, namely the two principal points, whereas the principal foci are *not* conjugates. This brings to light a slight inconsistency in notation; primed and unprimed letters normally refer to conjugates or to some other quantities, e.g. angles of incidence and refraction, which have the relationship "before and after going through the optical system or one surface of it". The principal foci do not fit into this scheme and they are therefore occasionally labelled F_1 and F_2, but F and F′ is the more usual notation.

A third useful pair of points on the axis can be defined, the nodal points N and N′: these are such that a ray entering through N emerges from N′ parallel to its initial direction. We can find the positions of the nodal points by starting with the usual skeleton system specified by F, F′, P, P′ as in Fig. 1.7: we construct any ray r_1 through F, meeting the principal planes in P_1

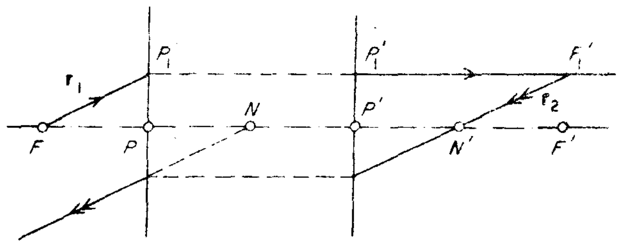

FIG. 1.7. Construction for the nodal points

and P_1' and meeting the image-side focal plane at F_1'; next we draw the ray r_2 through F_1' in image space and make it parallel to the segment of r_1 in object space. Since r_1 and r_2 meet at an image point F_1' which is on the image-side focal plane, they must come from an object point at infinity, i.e. they are parallel in object space: thus, both segments of r_2 are parallel to the segment of r_1 in object space and so r_2 must intersect the axis in the nodal points. It is easily seen by similar triangles that

$$FN = f', \qquad F'N' = f. \tag{1.14}$$

The six points F, F′, P, P′, N, N′, are sometimes called the cardinal points. If either of the quadruples F, F′, P, P′, or F, F′, N, N′, is known the properties of the system are determined completely, since the conjugate distance equation and the magnification formulae are known. The points can occur in any order and relative positions on the optical axis, subject only to the restrictions implied by eqn (1.14).

The relation between axial object and image points given by eqns (1.8) and (1.11) is in effect a one-to-one correspondence between pairs of points on a line, the optical axis; it is an *involution*, in the terminology of projective geometry. Similarly, the transformation which expresses the image segment of a ray in terms of the object segment is a one-to-one correspondence between lines in the same three-dimensional space (a *collineation*) with axial symmetry.

The more detailed theory of involutions and collineations is not important in geometrical optics, but they are mentioned here to establish the point that the most general one-to-one correspondences take this form; on the other hand, it will be seen that in *real* optical image formation the relationship between object and image entities is more complex. For example, more than one axial "image point" may correspond to a single object point, on account of spherical aberration, and a point which is common to three rays in object space may not be common to the corresponding three rays in image space. Thus real optical image formation is essentially more complicated than the ideal case we have been discussing.

2 Geometrical Optics

2.1 Rays and geometrical wavefronts

We obtained in Chapter 1 a simple model of image formation with axial symmetry, and we pointed out there that this was based on assumptions about optical systems which are only valid under certain restrictions. In this chapter and the next we explain these restrictions and develop further the theory of optical systems within them. In order to do this, we have to introduce a further concept, the *geometrical wavefront*, in addition to the *ray*, already used.

The concept of a geometrical wavefront appears in the work of Fermat (1667), Malus (1808), Hamilton (1820–30) and others, as a surface of constant optical path from the source or a surface orthogonal to the rays from a source point. More recently the shape of the geometrical wavefronts has been used to characterize the aberrations of an optical system directly, rather than regarding the ray patterns as fundamental; one of the earliest authors to do this was G. Yvon ("Contrôle des surfaces optiques", Paris 1926). This usage of the geometrical wavefronts provides a link with the physical concepts which originated with C. Huygens (1690) and A. Fresnel (1866) and developed into the Kirchhoff diffraction theory (1891). A very full treatment of the early history of these topics is given by E. T. Whittaker (1951), "History of the Theories of Aether and Electricity", Vol. I, revised edition, Nelson, London.

To an adequate approximation, we can regard rays as the paths along which the radiation energy travels; this breaks down near foci and near the edges of shadows, owing to diffraction effects, but it is essential to geometrical optics that these are ignored. Now, let a point source of light be placed in