MODERNIZING THE TEACHING OF ADVANCED GEOMETRIC OPTICS

Allen Nussbaum

Electrical Engineering Department, University of Minnesota
Minneapolis MN 55455 USA

Abstract

Advanced geometric optics as traditionally presented uses the approach of Conrady [1]. The subject can be made easier to teach and learn through the use of two powerful tools: matrix-based paraxial optics and computer-based non-paraxial analysis and design. Programs which handle many different kinds of problems will be discussed and misunderstandings concerning the definitions of the point aberrations will be corrected.

Keywords: education, paraxial matrices, skew ray tracing, aberration calculation

1. Paraxial Matrix Optics

Paraxial ray tracing in an optical system can be accomplished with the use of 2 x 2 matrices. These, as obtained [2,3] from the small-angle form of Snell's law, $n\theta = n'\theta'$, and the definition of the refracting power, $k = (n' - n)/r$, are the refraction matrix R and the translation matrix T, respectively. They are defined as

$$R = \begin{pmatrix} 1 & -k \\ 0 & 1 \end{pmatrix} \qquad T = \begin{pmatrix} 1 & 0 \\ t'/n' & 1 \end{pmatrix} \qquad (1)$$

where t' is the lens thickness, and n and n' are the indices on either side of an interface. Multiplying in the correct order a matrix R for each interface and a matrix T for each translation in an optical system of arbitrary complexity, we obtain the Gaussian constant matrix

$$\begin{pmatrix} b & -a \\ -d & c \end{pmatrix} \qquad (2)$$

These constants determine the location of the cardinal points and planes: the focal points F and F', the unit points H and H'--corresponding to unit linear magnification--and the nodal points N and N'--corresponding to unit angular magnification. This set of six cardinal points or planes completely models the system and eliminates the necessity of determining intermediate images. In object space, it is shown [2,3,4] that the unit plane H is located at a distance $n(1 - b)/a$ from the first vertex, the focal point F is at a distance $-nb/a$ and the nodal point N is at a distance $(n' - nb/a$, with similar expressions for the image space cardinal points H', F', and N'. It is also found that the object position t and the image position t' are related by the generalized Gauss law $t'/n' = [(cl/n) + d]/[(at/n) + b]$ and that the magnification is $m = c - (at'/n')$ or $1/m = b + (at/n)]$. Since the multiplication of more than a few matrices is tedious, a simple computer program in BASIC can be used locate the cardinal points and to generate a ray tracing diagram [3]. The position and nature of the image due to a specified object point can be found with a

scale drawing, which traces one ray parallel to the axis as far as the unit plane H′ and after refraction through F′ and traces a second ray from the same point through F to H where it becomes parallel to the axis. As a check, students are told to verify that the ray from an object point through N is precisely parallel to the ray from N′ to the image point, and also to compare their calculated image distance and magnification with the drawing. A further check involves verifying the fact that bc - ad = 1, a consequence of the fact that all individual matrices have a determinant of unity, and so their product must have the same determinant.

The paraxial theory presented above supports a systematic procedure which gives students confidence in their knowledge of the subject because it contains a number of built-in checks. As an example, consider an object that is 15 units in front of a double convex thin lens (defined as having zero thickness), and the lens is followed by a concave spherical mirror that is 20 units away. Let the lens surfaces have radii of 10 units each and let it have an index of 1.5, while the mirror surface has a radius of 16 units. The student uses these specifications to calculate the refraction powers at each of the three curved interfaces (remembering that the index after a reflection is -1) and writes the matrix equation for the complete process as

$$\begin{pmatrix} b & -a \\ -d & c \end{pmatrix} = \begin{pmatrix} 1 & -(-1 - 1)/-16 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 20 & 1 \end{pmatrix}\begin{pmatrix} 1 & -(1 - 3/2)/-10 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & -(3/2 - 1)/10 \\ 0 & 1 \end{pmatrix}$$

(3)

or

$$a = -1/40, \quad b = -3/2, \quad c = -1, \quad d = -20$$

and

$$(-3/2 \text{ x } -1) - (-20 \text{ x } -1/40) = 1$$

Using these Gaussian constants in the expressions for the positions of the object space cardinal points, their locations with respect to the first vertex are

$$F: \ -60, \quad H: \ -100, \quad N: \ -20$$

and the image space cardinal point locations with respect to the mirror vertex are

$$F': \ -40, \quad H': \ -80, \quad N': \ 0$$

The separation between F and H is the object space focal length (-40) and the separation between F′ and H′ is the image space focal length (40); these are equal but opposite, as they should be. Although this arrangement of cardinal points is most unusual, students can construct the scale ray tracing diagram of Figure 1, following the procedures mentioned above. An inverted, slightly reduced real image, located near the mirror, is obtained. This result is checked with the Gaussian lens equation
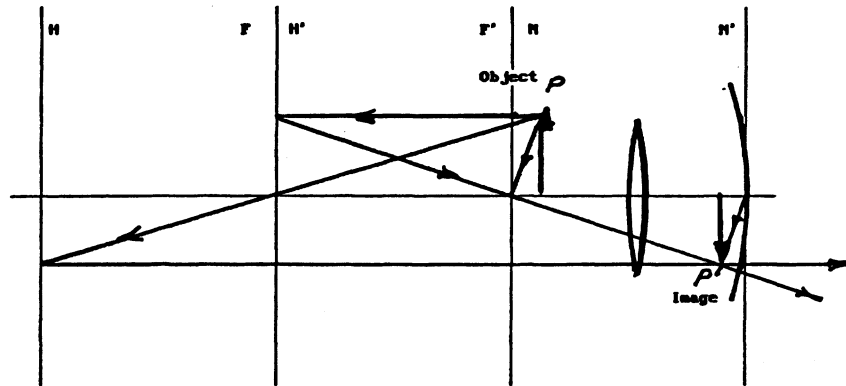
$$t'/n' = (ct + d)/(at + b) = 40/9$$

or

$$t' = -40/9$$

and the magnification is confirmed in two ways

$$1/m = at + b = -9/8, \quad m = c - at'/n' = =8/9$$
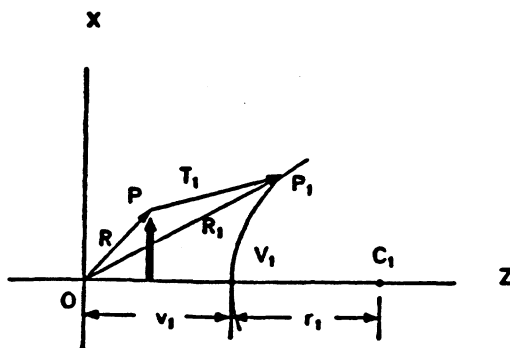
Figure 1

Finally, it is seen that the rays from object point to N and from N' to image point are parallel. All of these built-in checks provide confidence in the validity of the analysis of the system.

2. Non-paraxial Meridional

We consider next the procedure for tracing non-paraxial rays which lie in the meridional plane (i.e., a plane containing the symmetry axis). Figure 2 shows a meridional ray leaving an object point P and striking the first surface of a lens at the point P_1. This ray is regarded as the vector T_1. Figure 3 shows the same ray, now specified as a vector with length equal to the index in air, and refracted as it enters the lens. By the use of vector algebra and the equation of a circle, we obtain a quadratic expression for the magnitude of T_1. Determination of T_1 requires a knowledge of the starting coordinates and slope of the ray as well as the position and radius of the first surface of the lens.

Figure 2

Figure 3
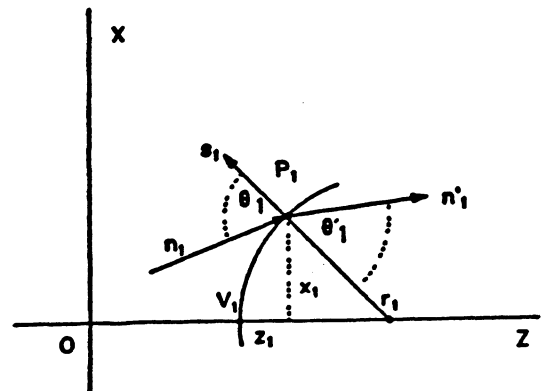
The coordinates of P_1 obtained in this calculation are then used to find the Snell's law angle after refraction, and this in turn permits the determination of the slope of the refracted ray. A derivation of the expressions needed in this exact meridional ray tracing procedure will be found in a previous publication [5] and the associated short BASIC program will appear soon [3] (or may be obtained from the author). The matrices used in non-paraxial ray tracing have the form

$$T_1 = \begin{pmatrix} 1 & 0 \\ T_1/n_1 & 1 \end{pmatrix}, \quad R_1 = \begin{pmatrix} 1 & -K_1 \\ 0 & 1 \end{pmatrix} \tag{4}$$

These 2 x 2 matrices are generalizations of the equivalent paraxial matrices; the constant translation t as measured along the z-axis is replaced by a variable distance T_1 measured along the ray and the constant refracting power k is replaced by a variable refracting power K_1 which involves the Snell's law angles.

3. Spherical Aberration

The procedure just described, when applied to a set of meridional rays coming from infinity and passing through a double convex lens, gives the ray trace of Figure 4. This non-paraxial ray tracing provides a simple definition of spherical aberration: it is the failure of meridional rays to obey the paraxial approximation. This statement is unlike that of most optics books; for example, Smith [6] states that it is the variation of focus with aperture. This contradicts the fact that the concepts of cardinal points or planes are purely paraxial; for large angles, the image formation is an ambiguous process, since no two rays from a given point on an object will necessarily meet at the image. The explicit definition given here removes this ambiguity.

Figure 4
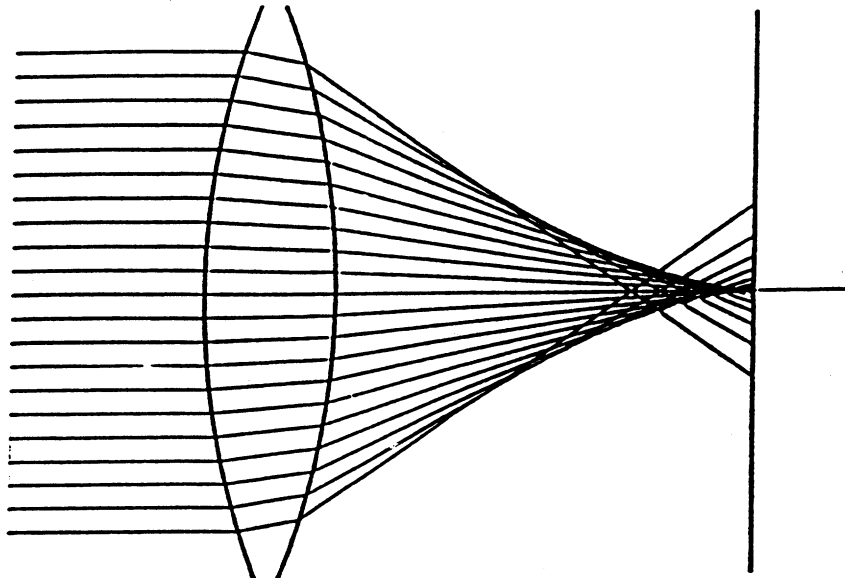
Although the shape of the caustic curve and the location of the circle of least confusion are shown, these concepts are of not too much pedagogical value. Instead, we quantify this aberration by defining the longitudinal spherical aberration as the distance from the point where a marginal ray crosses the axis to the paraxial focus. Using this concept, the BASIC programming language permits a thorough study of the lens bending process in a very direct manner; it is possible to change variables from the keyboard and immediately see the change in lens shape and the way in which the longitudinal spherical aberration increases or decreases. Figure 5 shows a symmetrical lens displayed by the the monitor. . The arrow keys permit changing the front radius by convenient amounts, and the program recalculates the aberration magnitude while maintaining the focal length at its original value. This computer procedure has many advantages over the optical bench; it is quicker, easier, and far more accurate.

Figure 5
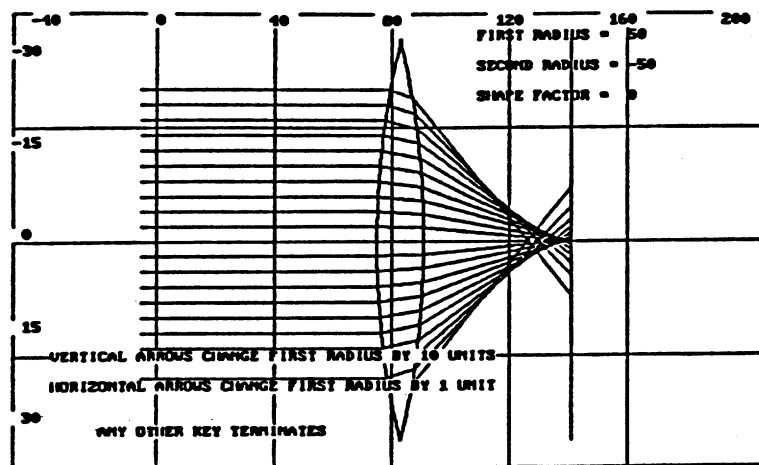
4. Skew Ray Optics and Coma

The procedure for tracing skew rays (those out of the meridional plane) is a simple extension of that described in Section 2; any term involving an x coordinate or direction-cosine will appear in identical form as a term in y or a y-axis direction-cosine. To understand the aberration due to skew rays, let a cylinder of light strike the front surface of a lens (Figure 6). All these meridional rays meet at a common image point, namely the apex of a symmetrical cone. Now have the cylinder axis make an angle with the lens axis in such a way that the apex of the cone moves upward in the meridional plane. When students are asked to describe the resulting image, they invariably guess that the cone will be slightly distorted and the apex will become less sharp. It is subsequently realized that tilting the cylinder converts all the rays except the two at the top and bottom from meridional to skew. These two meridional rays must meet at a point somewhat above the lens axis and this point determines an image plane. The remaining skew rays in the front half of the cylinder will then create a pattern on the image plane which starts at the meridional intersection and terminates there as well, so that the original sharp apex is now a closed curve. And the back half of the cylinder shows the same pattern. This is confirmed by Figure 7, which shows the ray trace for a double convex lens, and illustrates the aberration

Figure 6
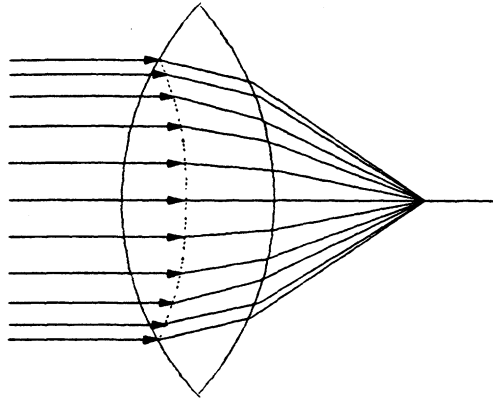
known as coma. Because coma is an unusual double-valued phenomenom, the standard texts state that the resulting image is a pair of superimposed circles and that for a series of incoming cylinders whose size decrease towards the axis, the

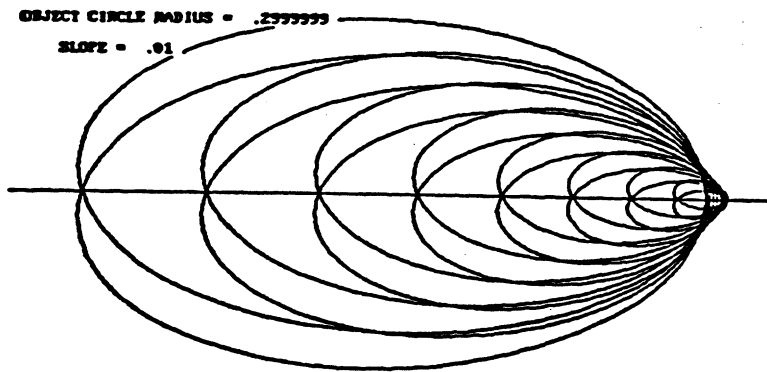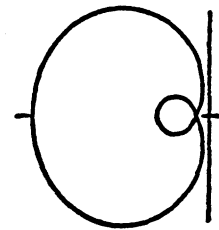Figure 7   Figure 8

image is a series of decreasing circles with shifted centers . It is not made clear that this is correct only when coma is the only aberration present. Otherwise, the two circular images do not register perfectly, and the cardioid pattern of Figure 8 is obtained from the skew ray trace. This pattern was also given by Franke [7] and in a photograph by Sherman [8]. We recognize that coma can be defined as the failure of skew rays to match the behavior of meridional rays; the definition given by Smith [6] is that it is the variation of magnification with aperture. Again

we have a misapplication of a paraxial concept to a non-paraxial situation. Another misconception that exists is illustrated in Figure 9, taken from Jenkins and White [16]. This figure shows the use of lens bending to reduce both spherical aberration and coma, and it is stated that coma can actually be made to vanish. This is probably a misinterpretation of a result given by Kingslake [9], who makes extensive use of the well-known Abbe sine condition. He shows that it is the offense against the sine condition (OSC) which is the quantity that vanishes when the coma is a minimum. The OSC is found by tracing meridional rays to estimate the magnitude of coma. A more convincing way of observing the effect of lens bending is to use the program that produced Figure 7. Figure 10 shows the coma pattern at the minimum for the lens of Figure 7; its magnitude is 4.5 times smaller than the symmetrical lens.

Figure 11 shows how the pattern changes when the bending goes beyond the minimum shape. It is interesting to note that the description of coma presented here is not universal. In his introduction to a study of coma, Zimmer [15] states: "We have formulae which enable us to trace any desired skew ray, but there is no exact theory which can explain and correlate the many effects and problems that arise."
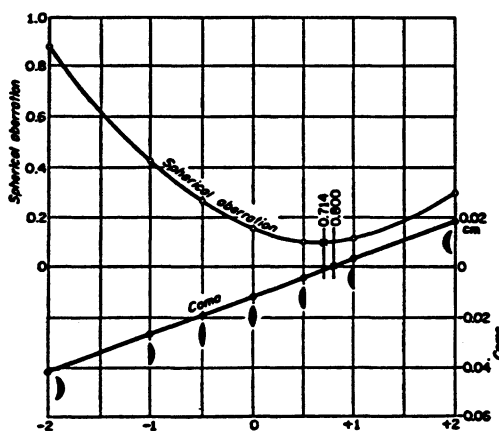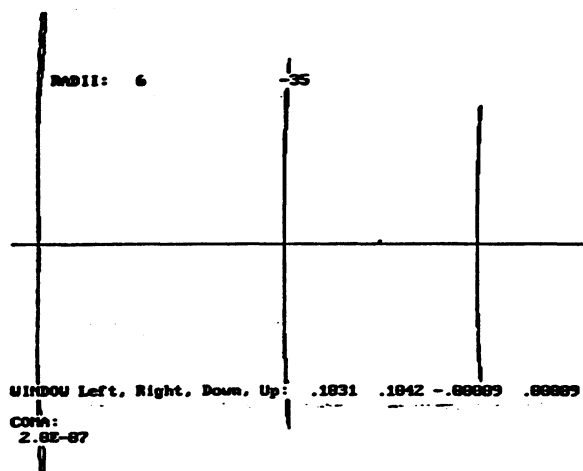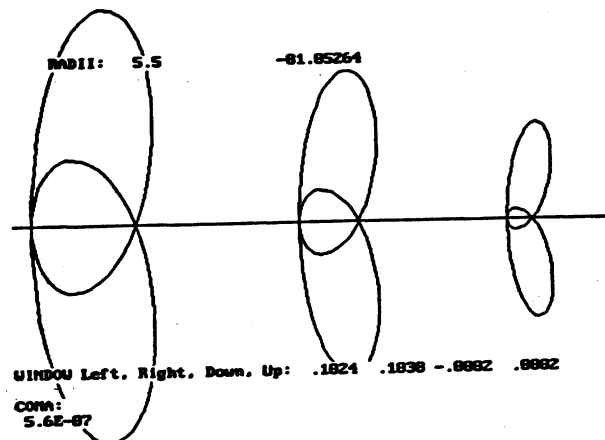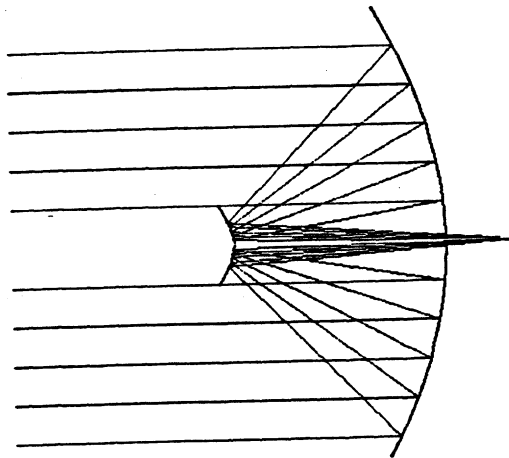
Figure 9

Figure 10

Figure 11

40

5. Aspheric Components

One of the truly remarkable aspects of the approach to optics teaching given here is the ease of extension to lenses and mirrors with conic-section profiles. Smith [6] gives a prescription using iteration for ray tracing with aspheric surfaces of any order. However, by replacing the equation for a sphere with that for the general second-order surface, the programs already described are readily extended to give a direct solution of the quadratic equation for T_1. The conic surfaces are specified by the simple expression

$$\frac{C(x^2 - Sz^2)}{2} - z = 0 \tag{5}$$

where C is the vertex curvature and S is the shape factor. It has the value S = 0 for a parabola, while a positive value specifies a member of a family of hyperbolas and a negative value specifies a family of ellipses, with S = -1 for a circle. Ray tracing for the Hubble space telescope as designed (Figure 12) shows the perfect meridional focusing obtained with two hyperbolic mirrors. A simpler system is the use of a Ross correcting eyepiece to reduce the
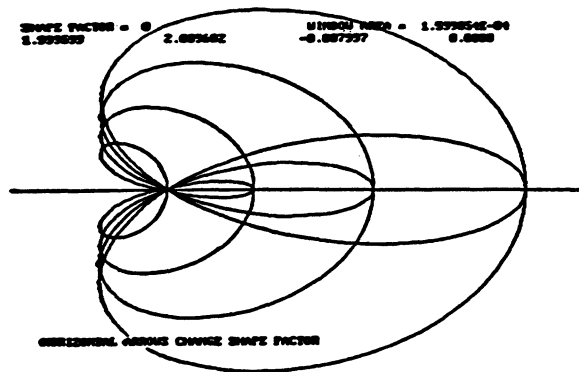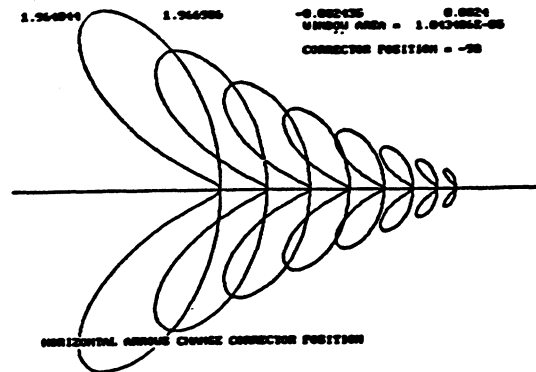
Figure 12

coma of a parabolic mirror. An air-spaced doublet is placed near the image plane of the telescope. Figure 13 shows the coma pattern of the mirror alone and Figure 14 shows the effect of a Ross corrector designed by Kingslake [9]. This design involved the use of the Abbe sine principle; it was found that when the amount of coma was specified as being proportional to the area of the pattern which just filled the monitor screen, a further reduction could be obtained by a slight adjustment of the corrector position. Another application of the conic ray tracing procedure involves the coma in a lens with one spherical and one elliptical surface. A very clever demonstration by Hecht and Zajac [17] indicates that if such a lens has an index equal to the reciprocal of the eccentricity, it will be free of spherical aberration. Plotting the coma for such a lens (Figure 15) shows the virtually perfect circles which almost coincide and which have asymptotes at an angle of 60°, as the classical textbook pictures indicate.
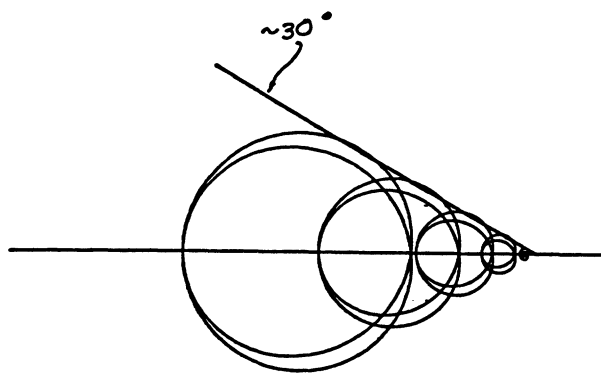
Figure 13

Figure 14

Figure 15

6. Astigmatism

The third point aberration to be studied in an advanced optics course is astigmatism, as illustrated in Figure 16. This diagram, found in most optics texts, is known to opthalmologists as Sturm's conoid [11]. It shows a meridional or tangential fan leaving a point P and coming to a perfect focus at the point P_T' and a sagittal fan from the same point coming to a perfect focus at P_S'. The separation of these two points (Sturm's interval) is taken as a measure of astigmatism. The perfect image for the tangential fan indicates that spherical aberration is absent and the perfect sagittal image indicates that coma has been fully corrected. It is then logical to define this astigmatism as the failure of the correction for meridional rays to agree with the correction for skew rays. However, an optical system having no spherical aberration or coma but with noticable astigmatism would be rather unusual. It is more practical to deal with systems in common use. Figure 17 shows the output of a ray tracing program for a triplet (US Patent No 2 453 260). The intersection of marginal meridional and skew rays was determined aand their distance from the paraxial image plane calculated and plotted. The spacing between these two curves is a direct measure of the astigmatism. Note that the designer arranged for it

to vanish at lens edge, thus limiting the maximum value.

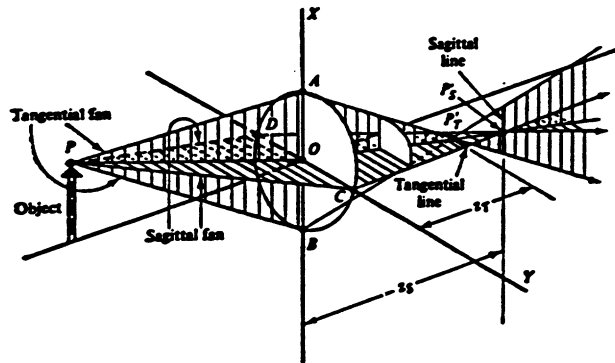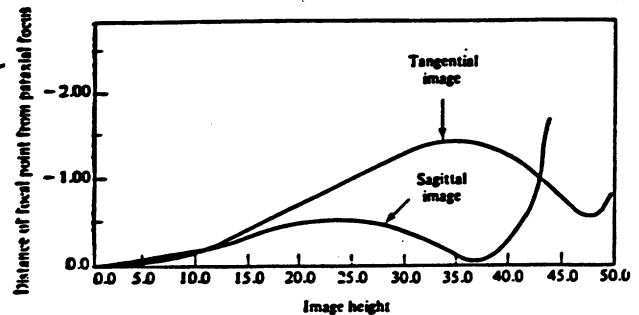Figure 16   Figure 17

It is apparent that astigmatism is difficult to reduce and is inherent in most systems. There appears to be a misunderstanding in the literature regarding this fact. For example, Smith and Atchison [12] state that astigmatism of the eye is due to its imperfect shape; a perfectly spherical eye should not have this defect. The figure above challenges this conclusion.

7 Summary

Advanced instruction in geometric optics can be made easier to present and and to understand by using matrix methods for paraxial components (as originally presented by Brouwer [10]) and using computers for exact ray tracing. Programming reveals the nature of the point aberrations and their correction procedures in both a quantitative and graphical way, in contrast to an involved analytic presentation such as that of Mahajan [13], which does not meet the promise implied in its title. A recent text [12] cites the very complicated approach of Hopkins [14], involving the use of meridional ray tracing to estimate the aberrations due to skew rays and doing these calculations one surface at a time. The above material has been used in one-term and short courses with indications that students find it helpful.

8. References

[1] A E Conrady, Applied Optics and Optical Design, Oxford (1929)
[2] A Nussbaum and R A Phillips, Contemporary Optics for Scientists and Engineers, Prentice-Hall (1976)
[3] A Nussbaum, Optical System Design, Prentice-Hall (1997) (in press)
[4] E L O'Neill, Introduction to Statistical Optics, Addison-Wesley (1963)
[5] A Nussbaum, Am J Phys 47, 351 (1979)

[6] W Smith, <u>Modern Optical Engineering</u>, McGraw-Hill (1990)
[7] G Franke, <u>Physical Optics in Photography</u>, Focal Press (1966)
[8] B Sherman, Modern Photography <u>32</u>, 118 (1968)
[9] R Kingslake, <u>Lens Design Fundamentals</u>, Academic Press (1978)
[10] W Brouwer, <u>Matrix Methods in Optical Instrument Design</u>, W A Benjamin (1964)
[11] H Davson, <u>The Eye</u>, Vol 4, Academic Press (1962)
[12] G Smith and D Atchison, <u>The Eye and Visual Instruments</u>, Cambridge (1997)
[13] V N Mahajan, <u>Aberration Theory Made Simple</u>, SPIE (1991)
[14] H H Hopkins, <u>The Wave Theory of Aberrations</u>, Oxford (1950)
[15] H-G Zimmer, <u>Geometrical Optics</u>, Springer (1970)
[16] F A Jenkins and H E White, <u>Fundamentals of Optics, 3rd ed</u>, McGraw-Hill (1957)
[17] E Hecht and A Zajac, <u>Optics</u>, Addison-Wesley (1974)