GA's in Optical Design - Addendum:

D.C. van Leijenhorst

Computing Science Institute/

CSI-R9806 January 1998
GA’s in Optical Design - Addendum:

Including a very short Introduction to Geometrical Optics

D.C. van Leijenhorst

March 23, 1998

Affiliation: Department of Computer Science
University of Nijmegen, NL
e-mail bolke@cs.kun.nl

ENGLISH VERSION (followed by the dutch version).

The main purpose of this report is to serve as an explanatory addendum to the article “Optical Design with the Aid of a Genetic Algorithm” by Carlos Lucasius, Jos Thijssen and myself, which appeared in “Biosystems” in 1996.

The first part forms a very succinct introduction to the field of geometrical optics, though it is assumed that the reader remembers some "schoolbook" optics. It summarizes exactly the theory necessary to perform paraxial optics and exact ray-tracing, and as such may be very useful to programmers.

In the second part, a number of ideas are given, some of which were implemented in the programs associated with for the above article; some not.
1 Lens systems.

Light rays travel from the left to the right (= the direction of the x-axis). The origin on the x-axis may be chosen in an arbitrary way.

Lens system. This is a series of surfaces $O_1, \ldots, O_k$, each rotationally symmetric around the x-axis. $O_i$ intersects the x-axis in $x = x_i$.

The object space lies to the left of $O_1$; the image space to the right of $O_k$.

The object space of $O_i$ lies to the left of $O_i$; the image space of $O_i$ to the right of $O_i$.

The equation of $O_i$: $F_i(x, y, z) = 0$ or $G_i(x, r) = 0$ with $r = \sqrt{y^2 + z^2}$. We assume that the surfaces $O_i$ do not intersect within a radius of $r < r_{\text{max}}$. Usually $F_i$ is quadratic; usually a quadric; usually part of a sphere.

The spherical surface.

In the latter case we denote: center $(x_i + R_i, 0, 0)$; radius $R_i$; equation $F_i(x, y, z) = (x - x_i - R_i)^2 + y^2 + z^2 = R_i^2$ with the further restriction $x < x_i + R_i$ ($O_i$ is convex: “bulges to the left”) or $x > x_i$ ($O_i$ concave).

Convention: if $O_i$ is convex, we choose $R_i > 0$; otherwise, $< 0$.

Media

With $O_i$ one associates the medium $M_i$ directly on the right of the surface ($1 \leq i \leq k$). $M_0$ is air; as well as $M_k$. 
$M_i$ will be described by: refractive indices $n_i$, $n'_i$ en $n''_i$. This will be done for three standard wavelengths, hence $M_i$ is described by 3 parameters. More wavelengths are allowed, however; in the table of optical glasses in the book “Telescope Optics, Evaluation and Design” by Rutten and van Venrooij seven are given. Dispersion is a derived quantity that depends on the refractive indices; we shall not need it in our calculations.

Fig. 1: A typical lens system.

2 Ray-Tracing-Exact

Light rays.

A light ray is given as a base point $p$ and a unit vector $\xi$.

The points on the ray then have the form $p + \lambda \xi; \lambda \geq 0$. Initially, $p$ lies in the object space of $O_i$; $p_i = p_{i-1}$, say; with unitvector $\xi_i = \xi_{i-1}; 1 \leq i \leq k$.

Refraction at a surface

Let us now intersect a ray with the surface $O_i$. The new base point is obtained as the intersection
point of the ray with \( O_i : F_i(p_{i-1} + \lambda \xi_{i-1}) = 0 \). This is a quadratic equation in \( \lambda \) (in the case of quadrics); the right choice \( \lambda_{i-1} \) van \( \lambda \) follows from the above mentioned restriction (\( O_i \) convex or concaaf). The base point becomes:

\[
\begin{align*}
\mathbf{p}_i &= p_{i-1} + \lambda_{i-1} \xi_{i-1}.
\end{align*}
\]

The new unit vector \( \xi_i \) follows from Snell’s Law: let \( \nabla F_i = (\partial F_i/\partial x, \partial F_i/\partial y, \partial F_i/\partial z)\vert_{x=p} \); then \( g = \nabla F_i/\||\nabla F_i||\); let \( g \) be the “inward pointing” normal vector of length 1 on \( O_i \) in \( x = p \).

Let \( \gamma = (g, \xi_{i-1}) \) (ordinary scalar product) and \( \nu = \frac{n_i-1}{n_i} \). Then \( \xi_i \) can be written as \( \alpha \xi_{i-1} + \beta g \).

where

\[
\begin{align*}
\alpha &= \sqrt{\frac{1 - \nu^2 \gamma^2}{1 - \gamma^2}} \quad \text{and} \quad \beta = (\nu - \alpha) \gamma.
\end{align*}
\]

The spot diagram.

In this way one may proceed with the next lens surface, etc. Finally one calculates the intersection point of the ray with the “ideal image plane” (from the paraxial approximation, see below), thus obtaining a “spot” of the spot diagram. Thereby, the ray trace (of this single ray) is completed.

3 Paraxial imaging.

3.1 The paraxial image

The paraxial image is a first order (linear and shape-preserving) approximation for rays travelling close to the axis in one plane with this axis. We consider this the “ideal” image formation which we try to extend into a region farther from the axis, to be specified at forehand.

Ideal lens surfaces.

The lens surface will now be represented in an “idealized” way as planes perpendicular to the x-axis
Radius of curvature and focal length.

Let $O_i$ have radius of curvature $R_i$ on the $x$-axis, i.e. in $(x_i, 0, 0)$. For spherical surfaces this radius equals our former $R_i$. One has: $1/R_i = -((\partial^2 x/\partial y)^2)|_{y=0}$ in absolute value and sign, where one obtains $x$ as a function of $y$ from the equation $F_i(x, y, 0) = 0$ and in a sufficiently small neighbourhood of $x = x_i$.

The radius of curvature is connected with the paraxial focal length $f_i$, given by: $f_i = R_i/(n_i - n_{i-1})$.

3.2 Paraxial ray-tracing

This is done by a $2 \times 2$ matrix transformation. The relevant notions now follow.

Paraxial rays.

Let $p_{i-1}$ be a point in the object space of $O_i$ with $y$-coordinate $y_{i-1}$ and consider in the $xy$-plane a light ray from $p_{i-1}$ with direction vector $\xi_{i-1}$. We shall describe this ray by $y_{i-1}$ and $\phi_{i-1}$, where $\sin(\phi_{i-1}) = (\xi_{i-1}, (1, 0, 0))$ and $\cos(\phi_{i-1}) = -(\xi_{i-1}, (0, 1, 0))$.

This means: $\phi_{i-1}$ is the angle of the ray with the positive $x$-direction. We assume that $\phi_{i-1}$ is small, and we reckon it positive for rays pointing downward. Anywhere, the sin and tg of the small angles $\phi$ will be replaced by $\phi$ itself.

Translation- and refraction matrices

Let $p_i$ be a point on the ray having $y$-coordinate $y_i$; let $p_{i-1}$ have distance $d_{i-1}$ to $p_i$ along the
$x$-axis. Then the new ray from $p_i$ is described by the matrix transformation

$$
\begin{pmatrix}
  y_i \\
  \phi_i
\end{pmatrix} =
\begin{pmatrix}
  1 & -d_{i-1} \\
  0 & 1
\end{pmatrix}
\begin{pmatrix}
  y_{i-1} \\
  \phi_{i-1}
\end{pmatrix}
$$

(translation).

If furthermore also $p_i$ lies on $L_i$ (so $d$ is the distance from $p_{i-1}$ to $L_i$) then refraction will occur as well; this can be described by the transformation

$$
\begin{pmatrix}
  y_i \\
  \phi_i
\end{pmatrix} =
\begin{pmatrix}
  1 & 0 \\
  1/n_i f_i & n_i/n_{i-1}
\end{pmatrix}
\begin{pmatrix}
  y_{i-1} \\
  \phi_{i-1}
\end{pmatrix}
$$

(refraction)

The ray trace

After this, another translation occurs towards $L_{i+1}$ plus a refraction at that location; in this way the total paraxial ray trace is described as a product of $2 \times 2$ matrices. Beyond the last surface $L_k$ a final translation takes place to the image plane.

3.3 The paraxial parameters of the system.

Let $M = \begin{pmatrix} p & q \\ r & s \end{pmatrix}$ be the product matrix without the two translations to $L_1$ and from $L_k$.

Several things can be read off from this matrix:

**The image.**

The paraxial mapping yields for every point $p_0$ in object space an image point in image space (object and/or image may lie at infinity; in this case, they have to be considered directions.)

Let $v$ be the distance of $p_0$ to $L_1$ and let $b$ be the (possibly negative) distance of $L_k$ to some point on the broken ray through the system.
The complete transformation to the image is then described by the matrix

\[
\begin{pmatrix}
1 & -b \\
0 & 1
\end{pmatrix}
\begin{pmatrix}
p & q \\
r & s
\end{pmatrix}
\begin{pmatrix}
1 & -v \\
0 & 1
\end{pmatrix}
= \begin{pmatrix}
p - br & -vp + q - b(-vr + s) \\
r & -vr + s
\end{pmatrix}.
\]

We note that an image is formed if the final \( y \)-value does not depend on the incidence angle \( \phi_0 \) of the ray, i.e. the entry in the upper right hand side of the matrix is 0. We find: \( b = \frac{vp - q}{vr - s} \).

Caveat

Note that \( v \) and \( b \) are always measured from the first and the last lens surface, respectively; \( v \) being positive if the point to be imaged lies to the left of \( L_1 \) and negative otherwise; \( b \) however is positive if the image point lies to the right of \( L_2 \) and negative otherwise. In order to apply the constructions learnt at school, one measures the object distance (\( v' \), say) and the image distance (\( b' \), say) from the so-called “first and second principal planes”. Familiar formulas like “\( 1/b' + 1/v' = 1/f' \)” then hold (the definition of \( f \) and the principal planes follow below.)

So in our notation, \( v \) and \( b \) have a different meaning!

Linear magnification. The linear magnification is equal to the quotient of the last and the first \( y \)-value, i.e. \( p - br \).

Principal planes.

Hence, the magnification is 1 if \( b = (p - 1)/r \). This is the location of the second principal plane. Its original is the first principal plane; from \( b = \frac{vp - q}{vr - s} \) one obtains \( v = q - bs \).

This means: the intersection point of a ray travelling through the system with the first principal
plane and its intersection with the second principal plane lie at the same height above the axis.

**Angular magnification.** Now consider the angular magnification. If \( \phi_1 \) and \( \phi_2 \) are two incidence angles of rays from \( p_0 \), the relation with the respective angles \( \phi'_1 \) and \( \phi'_2 \) of the emerging rays are given, using the above matrix, as \( \phi'_1 = ry + (-vr + s)\phi_1 \) and \( \phi'_2 = ry + (-vr + s)\phi_2 \). Therefore, \( \Delta \phi' = (-vr + s)\Delta \phi \): the **angular magnification**.

**Nodal planes.**

The angular magnification is 1 if \( -vr + s = 1 \) so \( v = (s - 1)/r \). Then \( b = (-p(s - 1) + qr)/r \). These \( v \) and \( b \) determine the planes with angular magnification 1: the so-called **first and second nodal plane**, respectively.

**Coincidence of principal and nodal planes.**

Note, that \( \det(M) = (n_1/n_0).(n_2/n_1)\ldots(n_k/n_{k-1}) = n_k/n_0 \) implies: \( ps - qr = n_k/n_0 \). For the second nodal plane it follows that \( b = (p - n_k/n_0)/r \); for the first principal plane that \( v = (s - n_k/n_0)/r \).

If we compare this with the second principal plane: \( b = (p - 1)/r \) and with the first nodal plane: \( v = (s - 1)/r \), respectively, then it is clear that the **principal and nodal planes coincide if** \( n_k = n_0 \); for instance if the first and the last medium are air.

*This means:* rays that (possibly after extension) intersect the first principal plane in the same point \( p \) (e.g., on the \( x \)-axis) emerge from the second principal plane not only at the same height, but even at the same angle. The region between both principal/nodal planes can be disregarded! Thus, the lens system is equivalent with one lens only with some specific focal distance \( f \) (see below). The elementary schoolbook constructions are applicable now.
Relation between linear and angular magnification.

A further simple consequence of the value of $det(M)$: (linear magnification) $\times$ (angular magnification) $= n_k/n_0$.

Shape preservation.

The matrix $M$ determines a mapping from object planes to image planes that from a geometric point of view preserves shapes near the axis only (since we identify angles with their sin and tg.) For larger angles, however, we can assume that the calculated linear magnification also occurs.

Focal distance.

If $v$ approaches infinity, we see that in the formula for the image distance $b$ approaches $p/r$. The focal distance $f$ of the system is measured from the second principal plane, and thus equals $p/r - (p - 1)/r = 1/r$.

Determining the aperture $D$.

We take $D = 2r_{\text{max}} = $ the diameter of the front lens. In fact one has to calculate of each diaphragm in the system the inverse image (leftward) with respect to the left hand subsystem and then take the minimum diameter of these images. This is called the entrance pupil. In practice, most of the time it turns out to be the front lens diameter.

Three-dimensionally the paraxial image of a point $p_0$ is calculated in the above way within the plane through this point and the $x$-axis. What matters is only the distance to the $x$-axis.

Also, the principal planes and the nodal planes can be calculated directly from $M$; for the time being however, this is of no importance to us.
4 Calculation of the Airy disk.

For this purpose there exists the formula $2.44\lambda/D$ radiants, $\lambda$ being given in nm (e.g., 555 nm for green light). $D = 2r_{\text{max}}$ is the aperture. For astronomical objects this amounts to $2.44\lambda f/D$ mm.

Note that $f/D$ is the aperture used in photography.

More in general, one has to take $2.44\lambda B/D$ mm in the image plane, with $B$ the image distance from the second principal plane. However, let us restrict ourselves to telescopes for the time being.

5 A program (first ideas.)

5.1 Graphical software

These display:
- lens system
- paraxial lens system
- object plane and image plane
- ray paths
- spotdiagram
- Airy disk

Secondary software.

Procedure traceray (should be fast). Traces a ray through a lens system.

Procedure spotdiag: calculates a spotdiagram from a given point.

Procedure parax: calculates matrix $M$ of the paraxial approximation and the Airy disk.

Visualisation of ray trace and paraxial ray traces (at first only as a check; later maybe in an expanded PC package?).
5.2 The Genetic algorithm.

Elementary objects in the population.

Lens system: given as a parameter array plus matrix of the paraxial approximation. Parameter array: initially only for spherical surfaces. $O_i$ will then be described by the radius $R_i$ and centre $x_i + R_i$ (in accordance with our sign conventions.)

Proposal for the “chromosome”:

At first one might try $x_1, R_1, n_1, n'_1, R_1, n_2, x_2, n_2, n'_2, \ldots, x_k, R_k$; notation in binary (each $R_i$ has its own sign bit).

(N.b.: $n_0, n'_0, n''_0, n_k, n'_k$ and $n''_k$ are usually those of plain air; hence they are omitted since they do not play a role in the evolution.)

This is also in accordance with the heuristic based on Holland’s Schemata Theorem, to put related pieces of information close together on the chromosome.

However: usually the number of glasses to choose from will be rather small, and the parameters $n_i, n'_i$ and $n''_i$ will not vary independently. In this case, also from the point of view of complexity theory, it seems wise to replace the triples $n_i, n'_i$, and $n''_i$ by a small binary number $a_i$: the address of the $i$th glass in a small table or database (note that this is, in fact, a form of data compression!). The chromosome then takes the form $x_1, R_1, a_1, x_2, R_2, a_2, \ldots, x_k, R_k$.

The precision to which the numbers $x_i, R_i, n_i, n'_i$, and $n''_i$ are given may be made to depend on the state of the genetic algorithm (whether one is close to an optimum or not). Another small problem: the sign bit of $R_i$ is not allowed to mutate too much; preferably only if $O_i$ is “almost flat”, i.e. $|R_i|$ is very large.
To each bit sequence (chromosome) $K$ (from now on to be called “telescope”) belongs a paraxial ray-trace matrix $M$ which has to be recalculated after mutation and crossover, at least if $K$ changes. The same holds for the aperture $D$; in practice (if we do not use small diaphragms (aperture stops)) we may fix it at $D = 2r_{\max}$.

5.3 The fitness function.

In order to calculate the fitness function, take an appropriate set of test points in object space, lying as a pattern in a certain plane $x = x_1 - v$.

From each test point $P$ there emerges an appropriate bundle of rays. These have to be chosen carefully: the intersection points with $O_1$ should be evenly distributed over this surface. To implement this, one might take a fixed set of points on $O_1$ and send a ray from $P$ to each of these; or one might send a bundle from $P$ towards $O_1$, evenly distributed qua solid angle. In this way recalculation of the bundle is avoided.

Remark.

During exact as well as paraxial ray tracing, it may happen that $v$ is infinite. Each point of the pattern then determines a direction; the ray bundle from this point consists of parallel lines. At finite distance one might take a basis point $P_0$ on each ray and from there send a ray into the system having the direction of the parallel bundle. Similar precautions should be taken if the image distance might become infinite.

The pattern itself has a paraxial image at distance $b$ from the rightmost lens surface. This is the ideal image. In practice, one might first define the ideal image and then calculate its inverse image, going from right to left, in order to obtain the object. This is useful since one has to define
at forehand a maximal circle in the image plane where sharpness should occur (e.g., “up to 20 mm from the axis”.)

The test points have paraxial test point images lying on the ideal image. For each wavelength there is an Airy disk the size of which depends on the wavelength and the aperture (see above).

For each test point $P$ calculate the spot diagram of its ray bundle and count the number of spots inside the Airy disk around the test point image; let this number be $u_P$. As a first tentative measure for the fitness one might think of $\min_p(i_P/(u_P + i_P))$ or something similar. Eventually, 90 to 95% of all the rays from each $P$ have to end inside the Airy disk.

**Choice of the pattern.**

The pattern has to satisfy:

1. At various, evenly distributed distances from the axis test points should occur.
2. Pinhole and cushion distortion should be detectable.

One can use the symmetry of the system around the $x$-axis to constrain the number of test points. As we shall see, due to this symmetry 2. follows from 1. in our test method.

A tryout: the left part of the following figure is the pattern, lying in the object plane perpendicular to the axis. The plane meets the axis in the centre of the circle.
First, one chooses as test points the fat points on the pattern. Since these are on a grid, distortion in image space is easily detectable and can be corrected by our algorithm. However, one may restrict oneself to the six encircled points at the right: if these are imaged well, then the rest too (by symmetry)! Furthermore, these six points may be moved along their circles until they lie on one radius (at their original distances from the axis.)

Remark.

Since the “blurring” aberrations increase with the distance to the axis, one might increase program speed by just taking one or a few points along the rim. Then however, pincushion and barrel distortions are out of our control.

Bottom line: it suffices to choose a radius together with an appropriate set of points on it - preferably, the circular projection of a grid as depicted above.

Number of points and number of rays from a point.

Just like the precision of the parameters, it is also possible to vary the number of rays from a point
during execution of the genetic algorithm: at first, take just a few and later, during the last calculations near the optimum, more. Unfortunately, this does not hold for the number of points in the pattern, since distortion cannot be corrected if the test points number too few.

**Constraining the number of ray-traces.**

In practice the diameter of the spot diagram appears to grow going outward from the axis. Therefore, initially it will be sufficient to calculate a complete spotdiagram for the outermost point for all colours. for the inner points it is sufficient to calculate one ray-trace for each point, i.e. a spot diagram consisting of only one spot per colour! Calculation of the fitness function is easier now: for the inner points the distribution of the spots is not important, but the geometrical distance of the single spot (per colour) to the ideal image point. This distance should be smaller than the radius of the Airy disk.

**REMARK:** THE SOFTWARE USED IN THE FINAL PAPER DIFFERS IN PLACES FROM THE IMPLEMENTATIONS SUGGESTED ABOVE. USE IS MADE OF A STANDARD GENETIC ALGORITHM PACKAGE, AND THE LENS SYSTEM IS BEING RESCALED CONTINUOUSLY DURING THE ALGORITHM.

*****

DCvL, 1997
Dit rapport is in de eerste plaats bedoeld als toelichting op het artikel “Optical Design with the Aid of a Genetic Algorithm” van Carlos Lucarius, Jos Thijssen en mijzelf, verschenen in “Biosystems” in 1996.

Het eerste stuk is een zeer beknopte inleiding in de geometrische optica. De lezer dient zich wel enige middelbare school-optica te herinneren. Het is een samenvatting van precies die theorie, nodig om paraxiale optica te bedrijven alsmede exacte ray-traces, en kan daardoor van veel nut zijn voor programmeurs.

In het tweede gedeelte staan nog een aantal ideeën, waarvan sommige wel, andere niet werden geïmplementeerd in de software bij dat artikel.

6 Lenzensystemen

Lichtstralen lopen van links naar rechts (= richting x-as). De oorsprong van de x-as is willekeurig te kiezen.

Lenzensysteem Dit is een rij oppervlakken $O_1, \ldots, O_k$, elk omwentelingssymmetrisch om x-as. $O_i$ snijdt de x-as in $x = x_i$.

De voorwerpsruimte ligt links van $O_1$; de beeldruimte rechts van $O_k$.

De voorwerpsruimte van $O_i$ ligt links van $O_i$; de beeldruimte van $O_i$ rechts van $O_i$.

De vergelijking van $O_i$: $F_i(x, y, z) = 0$ of $G_i(x, r) = 0$ met $r = \sqrt{y^2 + z^2}$. We nemen aan dat de oppervlakken $O_i$ geen intersectie hebben binnen een straal $r < r_{\text{max}}$. Meestal is $F_i$ kwadratisch;
meestal een omwentelingskegelsnede; meestal deel van een bol.

**Boloppervlak**

In laatste geval: middelpunt \((x_i + R_i, 0, 0)\); straal \(R_i\); vergelijking \(F_i(x, y, z) = (x - x_i - R_i)^2 + y^2 + z^2 = R_i^2\) met verdere restrictie \(x < x_i + R_i\) (\(O_i\) convex ("staat bol naar links toe") of \(x > x_i\) (\(O_i\) concaaf).

Conventie: als \(O_i\) convex is, kiezen we \(R_i > 0\); anders \(R_i < 0\).

**Media**

Met \(O_i\) is geassocieerd het medium \(M_i\) direct rechts van het oppervlak \((1 \leq i \leq k)\). \(M_0\) is lucht; \(M_k\) ook.

\(M_i\) wordt beschreven door: brekingsindices \(n_i, n'_i\) en \(n''_i\). Dit bij drie standaardgolflengten zodat we \(M_i\) beschrijven door 3 parameters. Evt. mogen dat meer golflengten worden; bij de tabel van optische glassoorten in het boek "Telescope Optics, Evaluation and Design" van Rutten en v. Venrooij worden er zeven gegeven. De **dispersie** is een afgeleide grootheid die meevarieert met de brekingsindices.

![Example for k=6](image.png)

Image plane
7 Ray-Tracing-Exact

Lichtstralen

Een lichtstraal wordt gegeven door een basispunt $p$ en een eenheidsvector $e$.

De punten op de straal hebben dan de vorm $p + \lambda e; \lambda \geq 0$. Initieel ligt $p$ in de voorwerpsruimte van $O_i$: zeg $p = p_{i-1}$; met eenheidsvector $e = e_{i-1}; 1 \leq i \leq k$.

Breking aan een oppervlak

We snijden nu de straal met $O_i$. Bepaal daarbij het nieuwe basispunt als snijpunt van de straal met $O_i : F_i(p_{i-1} + \lambda e_{i-1}) = 0$. Dit is een kwadratische vergelijking in $\lambda$ (bij kegelsneden); de goede keuze $\lambda_{i-1}$ van $\lambda$ volgt uit de eerder vermelde restrictie ($O_i$ convex of concaaf). Het basispunt wordt:

$$p_i = p_{i-1} + \lambda_{i-1} e_{i-1}.$$

De nieuwe eenheidsvector $e_i$ volgt uit de wet van Snellius: laat $\text{grad}(F_i) = (\partial F_i/\partial x, \partial F_i/\partial y, \partial F_i/\partial z)|_{x = p_i}$ en $g = \text{grad}F_i/||\text{grad}F_i||$; dan is $g$ de “naar binnen wijzende” normaalvector van lengte 1 op $O_i$ in $x = p_i$. Laat $\gamma = (g,e_{i-1})$ (gewoon scalar product) en $\nu = (n_{i-1})/n_i$. Dan is $e_i$ te schrijven als $\alpha e_{i-1} + \beta g$ waarin

$$\alpha = \sqrt{\frac{1 - \nu^2 \gamma^2}{1 - \gamma^2}} \text{ en } \beta = (\nu - \alpha) \gamma.$$

Het spotdiagram

Zo kan men voortgaan met het volgende lensoppervlak, etc. Tenslotte snijdt men de straal met het berekende ideale beeldvlak (uit de paraxiale benadering; zie onder) en is een “spot” van het spotdiagram bekend! (en deze ray-trace voltooid).
8 Paraxiale beeldvorming

8.1 De paraxiale afbeelding

De paraxiale afbeelding is een eerste orde (lineaire en vormgetrouwe) benadering voor stralen die dicht langs de as lopen en in één vlak met deze as. Dit beschouwen we als de “ideale” afbeelding die we zoeken uit te breiden tot verder van de as in een vooraf te specificeren gebied.

Ideale lensoppervlakken

De lensoppervlakken worden nu geïdealiseerd voorgesteld als vlakken loodrecht op de x-as (grafisch in het xy-vlak dus: lijnen); $O_i$ is dan te tekenen als de loodlijn $L_i$ in $x_i$. Het brekend vermogen van $L_i$ wordt beschreven door de brandpuntsafstand $f_i$; zie hieronder.

Krommingsstraal en brandpuntsafstand

Stel $O_i$ heeft op de x-as, dus in $(x_i, 0, 0)$, krommingsstraal $R_i$. Voor boloppervlakken is deze gelijk aan de vroegere $R_i$. Er geldt: $1/R_i = -\left(\frac{\partial^2 x}{\partial y^2}\right)|_{y=0}$ naar absolute waarde en teken, waarin we $x$ als functie van $y$ halen uit $F_i(x, y, 0) = 0$ en werken in de buurt van $x = x_i$.

Met de krommingsstraal hangt samen de paraxiale brandpuntsafstand $f_i$, gegeven door: $f_i = R_i/(n_i - n_{i-1})$.

8.2 Paraxiale ray-tracing

Dit gaat via een 2 bij 2 matrixtransformatie. De relevante begrippen volgen nu.

Paraxiale stralen

Zij $p_{i-1}$ een punt in de voorwerpsruimte van $O_i$ met $y$-coördinaat $y_{i-1}$ en beschouw in het $xy$-vlak
een lichtstraal uit \( P_{i-1} \) met richtingsvector \( \mathbf{e}_{i-1} \). We beschrijven die straal door \( y_{i-1} \) en \( \phi_{i-1} \), waar

\[
\sin(\phi_{i-1}) = (\mathbf{e}_{i-1}, (1, 0, 0)) \quad \text{en} \quad \cos(\phi_{i-1}) = -(\mathbf{e}_{i-1}, (0, 1, 0)).
\]

*Betekenis:* \( f_{i-1} \) is de (kleine, voor omlaag gerichte stralen positief gerekende) hoek die de straal maakt met de positieve \( x \)-richting. We vervangen overal \( \sin \) en \( \tan \) van de kleine hoeken \( \phi \) door \( \phi \) zelf.

**Translatie- en brekingsmatrices**

Laat \( P_i \) een punt op de straal zijn met \( y \)-coördinaat \( y_i \); laat \( P_{i-1} \) langs de \( x \)-as afstand \( d_{i-1} \) tot \( P_i \) hebben. Dan wordt de nieuwe straal uit \( P_i \) beschreven door de matrixtransformatie

\[
\begin{pmatrix}
  y_i \\
  \phi_i
\end{pmatrix}
= 
\begin{pmatrix}
  1 & -d_{i-1} \\
  0 & 1
\end{pmatrix}
\begin{pmatrix}
  y_{i-1} \\
  \phi_{i-1}
\end{pmatrix}
\]

*(translatie).*

Als \( P_i \) bovendien op \( L_i \) ligt (dus \( d \) is de afstand van \( P_{i-1} \) tot \( L_i \)) dan vindt bovendien breking plaats; deze is te beschrijven door de transformatie

\[
\begin{pmatrix}
  y_i \\
  \phi_i
\end{pmatrix}
= 
\begin{pmatrix}
  1 & 0 \\
  1/n_i f_i & n_i/n_{i-1}
\end{pmatrix}
\begin{pmatrix}
  y_{i-1} \\
  \phi_{i-1}
\end{pmatrix}
\]

*(breking)*

**De ray-trace**

Hierna vindt dan weer een translatie plaats naar \( L_{i+1} \) plus een breking aldaar; aldus wordt de hele paraxiale ray-trace beschreven met een product van 2 bij 2 matrices. Na het laatste oppervlak \( L_k \) vindt nog een laatste translatie plaats tot het beeldvlak.
8.3 De paraxiale parameters van het stelsel

Zij \( M = \begin{pmatrix} p & q \\ r & s \end{pmatrix} \) de productmatrix zonder de twee translaties tot \( L_1 \) en vanaf \( L_k \).

We kunnen uit deze matrix verscheidene zaken aflezen:

Het beeld

De paraxiale afbeelding geeft van elk punt \( p_0 \) uit de voorwerpsruimte een beeldpunt in de beeldruimte (voorwerp en/of beeld kunnen oneindig ver weg liggen en moeten dan worden opgevat als richtingen).

Zij \( v \) de afstand van \( p_0 \) tot \( L_1 \) en zij \( b \) de (mogelijk negatieve) afstand van \( L_k \) tot het een punt op de gebroken straal uit het systeem.

De totale transformatie tot het beeld wordt dan beschreven door de matrix

\[
\begin{pmatrix}
1 & -b \\
0 & 1
\end{pmatrix}
\begin{pmatrix}
p & q \\
r & s
\end{pmatrix}
\begin{pmatrix}
1 & -v \\
0 & 1
\end{pmatrix}
= \begin{pmatrix}
p - br & - vp + q - b(-vr + s) \\
r & -vr + s
\end{pmatrix}.
\]

We zien dat er beeldvorming optreedt als de slot- \( v \)-waarde niet van de invalshoek \( \phi_0 \) van de straal afhangt, dus wanneer rechtsboven een 0 staat; dus als \( b = \frac{vr - \phi}{vr - s} \).

Caveat

Let wel \( v \) en \( b \) worden hier steeds gerekend vanaf resp. eerste en laatste lens, en wel is \( v \) positief als het af te beelden punt links ligt van \( L_1 \) en anders negatief; \( b \) is echter positief als het beeldpunt rechts ligt van \( L_k \) en anders negatief. Wil men de constructies van de middelbare school toepassen, dan moet men voorwerpsafstand (zeg \( v' \)) en beeldafstand (zeg \( b' \)) rekenen vanaf resp. “eerste en tweede hoofdvlak”; er gelden dan bekende formuletjes als “\( 1/b' + 1/v' = 1/f' \)” (voor \( f \) en definitie
hoofdvlakken: zie onder).

Bij ons hebben $v$ en $b$ dus een andere betekenis!

**Lineaire vergroting** Verder is de *lineaire vergroting* gelijk aan de verhouding van eerste en laatste $y$-waarde, i.e. $p - br$.

**Hoofdvlakken**

De vergroting is dus indien $b = (p - 1)/r$. Dit is de plaats van het zgn. *tweede hoofdvlak*. Het origineel daarvan is het *eerste hoofdvlak*; uit $b = (vp - q)/(vr - s) \rightarrow v = q - bs$.

*Betekenis*: het snijpunt van een in het systeem vallende straal met het eerste hoofdvlak en het snijpunt van de uiteindelijke straal uit het systeem met het tweede hoofdvlak liggen op gelijke hoogte.

**Hoekvergroting** Bezien we nu de hoekvergroting. Als $\phi_1$ en $\phi_2$ twee invalshoeken zijn van stralen uit $p_0$, dan wordt het verband met de resp. uitvalshoeken $\phi'_1$ en $\phi'_2$ van de door het lenzen-systeem gebroken stralen, volgens bovenstaande matrix gegeven door $\phi'_1 = ry + (-vr + s)\phi_1$ en $\phi'_2 = ry + (-vr + s)\phi_2$ dus $\Delta \phi' = (-vr + s)\Delta \phi$; de *hoekvergroting*.

**Knopenvlakken**

De hoekvergroting is 1 als $-vr + s = 1$ dus $v = (s - 1)/r$. Dan is $b = (-p(s - 1) + qr)/r$. Deze $v$ en $b$ bepalen de vlakken met hoekvergroting 1; het zgn. eerste en tweede knopenvlak(resp.).

**Samenvallen van hoofd-en knopenvlakken**

Merk op dat uit het $M = (n_1/n_0). (n_2/n_1) \ldots (n_k/n_{k-1}) = n_k/n_0 \Rightarrow ps - qr = n_k/n_0$. Voor het tweede knopenvlak volgt dan $b = (p - n_k/n_0)/r$ en voor het eerste hoofdvlak $v = (s - n_k/n_0)/r$. 

22
Vergelijken we dit met resp. het tweede hoofdvlak: \( b = (p - 1)/r \) en met het eerste knopenvlak: 
\( v = (s - 1)/r \); dan zien we dat hoofd-en knopenvlakken samenvallen indien \( n_k = n_0 \); b.v. als eerste en laatste medium lucht zijn.

*Betekenis:* stralen die na verlenging hoofdvlak 1 snijden in hetzelfde punt \( p \) (b.v. op de \( x \)-as) komen niet alleen op dezelfde hoogte, maar zelfs onder dezelfde onderlinge hoek weer uit hoofdvlak 2. Het gebied tussen beide hoofd/knopenvlakken valt gewoon weg te denken! Het lenzensysteem is aequivalent met één lens met brandpuntstand \( f \) (z.o.) op plaats van de hoofdvlakken. Daarop zijn de elementaire middelbare school-constructies toepasbaar. Wij bevinden ons met onze programmatuur (voorlopig) in dit geval.

**Verband hoek-en lineaire vergroting**

Ook volgt nog eenvoudig uit de waarde van het \( M \): (lin. vergroting).(hoekvergroting) = \( n_k/n_0 \).

**Vormgetrouwe afbeelding**

De matrix \( M \) bepaalt een afbeelding van voorwerpsvlakken naar beeldvlakken die geometrisch gesproken alleen vormgetrouw is nabij de as (wegens identificatie van hoeken met hun sin en tg). Voor grotere hoeken kunnen we echter aannemen dat er dezelfde lineaire vergroting optreedt als uitgerekend.

**Brandpuntsafstand**

Voor \( v \) naar oneindig nadert (in de formule voor de beeldafstand) \( b \) tot \( p/r \). De *brandpuntsafstand* \( f \) van het stelsel wordt gerekend vanaf het tweede hoofdvlak en is dus \( p/r - (p - 1)/r = 1/r \).

**Het bepalen van de opening \( D \)**
Hiertoe nemen we $D = 2r_{\text{max}}$ = de diameter van de frontlens. In feite moet men van alle diafragma’s het inverse beeld (naar links) bepalen in het linkere sub-lenzenstelsel en van de gevonden beelden de minimum diameter: dit heet de intreepupil. Maar gewoonlijk blijkt dat de frontlensdiameter te zijn.

*Driedimensionaal* wordt het paraxiale beeld van een punt $p_0$ bepaald door in het vlak van dat punt en de $x$-as op bovenstaande wijze te werken; van belang is dus slechts de afstand tot de $x$-as. Ook de zgn. hoofdvakken en knopenvlakken van het systeem zijn met behulp van $M$ te bepalen; dit is echter voor ons voorlopig van geen belang.

9 Het berekenen van de airy disk

Hiervoor bestaat de formule $\lambda.2,44/D$ radialen, met $\lambda$ uitgedrukt in nm (b.v. 555 nm voor groen licht). Daarin is $D = 2r_{\text{max}}$ de opening. Bij afbeelding van een astronomisch object komt dit overeen met $\lambda.2,44.f/D$ mm. Merk op: $f/D$ is de in de fotografie gebruikelijke openingsverhouding. (In het algemeen zal men in het beeldvlak $\lambda.2,44.B/D$ mm moeten nemen, met $B$ de beeldafstand vanaf het tweede zgn. hoofdvak. Laten wij dit voorlopig niet doen en ons beperken tot kijkers).

10 Een mogelijk programma (eerste ideeën)

10.1 Grafische programmatuur

Deze maakt zichtbaar:
- lenzensysteem
- paraxiaal lenzensysteem
- voorwerpvlak en beeldvlak
- loop van stralen
- spotdiagram
- Airy disk

**Hulpprogrammatuur.**

Procedure traceray (letten op snelheid). Trace’t een ray door een lenzensysteem.

Procedure spotdiag: berekent spotdiagram vanuit gegeven punt.

Procedure parax: berekent matrix M bij paraxiale benadering en de Airy disk.

Visualisatie van ray-trace en paraxiale ray trace (voorlopig alleen ter controle; later wellicht in uitgebreide versie in PC-pakket?).

**10.2 Genetisch algorithme.**

**Elementaire objecten in de populatie.**

Lenzensysteem: als rij parameters plus matrix van bijbehorende paraxiale benadering. Rij parameters: voorlopig beperken we ons tot boloppervlakken. $O_i$ wordt dan beschreven door de straal $R_i$ en middelpunt $x_i + R_i$ (klopt bij onze tekenconventie).

**Voorstel voor het “chromosoom”:**

In eerste instantie zou men kunnen denken aan $x_1, R_1, n_1, n'_1, n''_1, x_2, R_2, n_2, n'_2, n''_2, \ldots, x_k, R_k$; binair opgeschreven (elke $R_i$ heeft een apart tekenbit).

(N.b.: $n_0, n'_0, n''_0, n_k, n'_k$ en $n''_k$ zijn gewoonlijk die van lucht en worden weggelaten omdat ze niet meedoen in de evolutie).

Hiermee is tevens recht gedaan aan de heuristiek, gebaseerd op het schemata theorema van Holland, om zaken die samenhangen dicht bij elkaar in het chromosoom te zetten.

Echter: gewoonlijk zal het aantal te kiezen glassoorten vrij beperkt zijn, en variëren de parameters $n_i, n'_i$, en $n''_i$ niet onafhankelijk. In dat geval lijkt het, ook complexiteitstheoretisch, verstandig
om de drietalen \( n_i, n'_i, \) en \( n''_i \) te vervangen door één klein binair getal \( a_i; \) het adres van de \( i^\text{e} \) glassoort in een kleine tabel of database (in feite is dit een vorm van datacompressie!). Het chromosoom krijgt dan dus de vorm \( x_1, R_1, a_1, x_2, R_2, a_2, \ldots, x_k, R_k. \)

De precisie waarin de getallen \( x_i, R_i, n_i, n'_i, \) en \( n''_i \) worden gegeven kan afhangen van het stadium waarin je genetisch algorithme verkeert (of je al dicht bij een optimum zit).

Voorts is er nog een probleempje: het tekenbit van \( R_i \) mag niet (teveel) muteren; liefst alleen als \( O_i \) “bijna vlak” is, dus \(|R_i|\) erg groot.

Bij elke bitrij (chromosoom) \( K \) (vanaf nu “kijker” te noemen) hoort een paraxiale ray-trace matrix \( M \) die na mutatie en crossover opnieuw even moet worden berekend, tenminste als \( K \) verandert.

Evenzo moet er nog bijgehouden worden de opening \( D \); deze ligt echter meestal in de praktijk (als we geen kleine diafragma’s monteren) aan het begin reeds vast \((D = 2r_{\text{max}})\).

### 10.3 De fitnessfunctie.

Om deze te berekenen beschouwde men in eerste instantie een geschikt stel testpunten in de voorwerpsruimte, liggend op een vlak patroon in zeker vlak \( x = x_1 - v. \)

Vanuit elk testpunt \( P \) komt een geschikte bundel stralen. Deze moet zorgvuldig gekozen worden: de snijpunten met \( O_1 \) moeten gelijkmatig verdeeld zijn over dit oppervlak. Hoe dit te realiseren is nog de vraag. Men zou op \( O_1 \) een vast stel punten kunnen nemen en stralen van \( P \) daarnaar toe; men zou ook vanuit \( P \) een qua ruimtehoek gelijkmatig verdeelde bundel kunnen sturen en afwachten waar \( O_1 \) getroffen wordt. Dit heeft als voordeel dat de stralenbundel niet steeds opnieuw behoeft te worden uitgerekt.
Opmerking

Het komt zowel bij exact als bij paraxiaal ray-tracen vaak voor dat de voorwerpsafstand $v$ oneindig is. Elk punt van het patroon komt dan overeen met een zekere richting; de bundel stralen eruit bestaat uit *evenwijdige lijnen*. Op *eindige afstand* zou men op elke straal een basispunt $p_0$ kunnen nemen en van daaruit een straal met de bedoelde richting het lenzensysteem in. Soortgelijke voorzorgen nemen men voor als de beeldafstand oneindig is.

Het patroon zelf wordt paraxiaal afgebeeld op afstand $b$ van de laatste lens. Dit beeld is de *ideale figuur*. In de praktijk kan men dit ideale figuur eerst definiëren en van rechts naar links terugrekenen door het lenzensysteem om het bijbehorende voorwerp te bepalen. Dit is vooral nuttig omdat men vooraf een maximale cirkel in het beeldvlak definieert waarbinnen scherpte moet heersen (b.v. “tot 20 mm van de as”).

De testpunten worden paraxiaal afgebeeld tot *testbeeldpunten* liggend op de ideale figuur. Bij elke golflengte hoort rond elk testbeeldpunt een Airy disk met, zie boven, een straal die afhangt van de golflengte, en de afmetingen (openingsverhouding) van het lenzensysteem.

Bij elk testpunt $P$ bepale men het spotdiagram van de bijbehorende bundel en telt hoeveel spots er binnen de bijbehorende verstrooingscirkel rond het testbeeldpunt liggen, zeg $i_P$ stuks, en hoeveel erbuiten; zeg $u_P$ stuks. Als eerste tentatieve maat voor de fitness zou men kunnen denken aan $\min_P(i_P/(u_P + i_P))$ o.i.d. Uiteindelijk is het zo dat 90 à 95 % van de stralen binnen de Airy disk moeten komen te vallen.

**De keuze van het patroon**

Dit moet voldoen aan:
1. Op diverse, gelijkmatige afstanden van de as moeten er testpunten liggen.

2. Ton-en kussenvormige vervorming moeten detecteerbaar zijn.

Men kan hier gebruik maken van omwentelingssymmetrie om het aantal testpunten te beperken. We zullen zien: wegens die symmetrie volgt 2. uit 1. bij onze testmethode.

Een poging: de volgende figuur links stelt het patroon voor, in het voorwerpsvlak loodrecht op de as. Het midden van de cirkel is het snijpunt met de as.

![Patroon](image)

Fig. 2: Een mogelijk patroon.

In eerste instantie kiest men als testpunten de vette punten op het patroon. Omdat deze op een rooster liggen is eventuele vervorming in de beeldruimte direct zichtbaar en door ons algoritme corrigeerbaar. Men mag zich echter beperken tot de zes omcirkelde punten rechts: als die goed afgebeeld worden dan, wegens omwentelingssymmetrie, de rest ook! Zelfs zou men deze zes punten alle nog kunnen verplaatsen tot ze op een radius liggen (maar wel met hun oorspronkelijke afstanden tot het middelpunt).
Opmerking

Aangezien de “versmerende” beeldfouten toenemen vanaf de as, zou men voor de snelheid zelfs kunnen volstaan met één of enkele punten langs de buitenrand - ware het niet dat dan de ton- of kussenvormige vertekening niet onder controle is!

Resumerend: men hoeft slechts een radius te kiezen en een geschikt stel punten daarop - liefst volgendo uit de projectie van een rooster, zoals boven getekend.

Aantal punten en aantal stralen vanuit een punt

Net als de precisie van de parameters (brekingsindices etc.) kan men ook het aantal stralen vanuit een punt tijdens het genetisch algorithme laten variëren: men kan er eerst een paar nemen en later, bij de laatste verfijningen dicht bij het optimum, meer. Dit geldt helaas niet voor het aantal punten uit het patroon waarvoor men een spotdiagram maakt, want bij te weinig punten houd je de vervorming niet in de hand.

Beperking van het aantal ray-traces

In de praktijk blijkt de diameter van het spotdiagram vanaf de as naar buiten toe groter te worden. Daarom zal het in eerste instantie voldoende zijn om tijdens de optimalisatie van het buitenste punt een volledig spotdiagram bij te houden voor alle kleuren. Voor de meer naar binnen gelegen punten kan men volstaan met voor elk van de kleuren één ray te tracen, dus een spotdiagram van één spot slechts (per kleur)! Ook de berekening van de fitnessfunctie is dan uiteraard eenvoudiger: voor de binnenste beeldpunten gaat het nu niet om de spreiding van de spots, maar om de geometrische afstand van de éne spot (per kleur) tot het ideale beeldpunt, welke kleiner moet zijn dan de Airy disc radius.
OPMERKING: IN HET UITEINDELIJKE ARTIKEL IS DE PROGRAMMATUUR ENIGSZINS ANDERS. GEBRUIK WORDT GEMAAKT VAN EEN STANDAARD GENETISCH PAKKET TERWIJL HET LENZENSYSTEEM STEEDS GEHERSCHAALD WORDT.

***** DCvL, 1997