LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

UCRL-TR-200436

Huygens Integral Transformation for A 4x4 Ray Matrix

D. Phillion

October 21, 2003

This work was performed under the auspices of the U.S. Department of Energy by University of California, Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.

HUYGENS INTEGRAL TRANSFORMATION FOR A 4x4 RAY MATRIX

D. W. Phillion

A general asymmetrical optical system can be represented to first order by the 4x4 ray matrix

$$\begin{bmatrix} x_2 \\ y_2 \\ n_2\, x_2' \\ n_2\, y_2' \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ n_1\, x_1' \\ n_1\, y_1' \end{bmatrix} \tag{1}$$

where **A**, **B**, **C**, and **D** are 2x2 sub-matrices. For an optical system with axial symmetry, **A**, **B**, **C**, **D** are the scalars A, B, C, D making up the 2x2 ray matrix multiplied by the 2x2 identity matrix **I**. The primes represent derivatives with respect to the propagation direction z. Thus x_1', y_1', x_2', and y_2' are dimensionless slopes. Here n_1 and n_2 are the refractive indices on sides 1 and 2, respectively. We have chosen to work with the generalized slopes n_1x_1', n_1y_1', n_2x_2', and n_2y_2' in order to greatly simplify the mathematics. Only ten of the sixteen real elements in this 4x4 ray matrix are independent. The first order eikonal is a homogeneous quadratic in x_1, y_1, x_2, and y_2. There are ten combinations: There are ${}_4C_2$ or six combinations in which the two variables in a quadratic term differ and four combinations in which the two variables are the same. We will derive explicit constraint equations which show that only ten elements are independent in the 4x4 ray matrix.

Let $\theta(x_1,y_1,x_2,y_2)$ be the eikonal in millimeters. It is the optical path of the ray connecting x_1,y_1 to x_2,y_2. In first order optics, this ray always exists and is unique. The following four equations must be integrated:

$$\begin{aligned} \frac{\partial\theta}{\partial x_1} &= -n_1\, x_1' \\ \frac{\partial\theta}{\partial y_1} &= -n_1\, y_1' \\ \frac{\partial\theta}{\partial x_2} &= n_2\, x_2' \\ \frac{\partial\theta}{\partial y_2} &= n_2 y_2' \end{aligned} \tag{2}$$

The last two equations are readily obtained by considering a ray bundle from a fixed point source on side 1. The wavefronts on side 2 are normal to the ray directions on side 2. Similarly, the first two equations are obtained by considering a ray bundle which converges to a point on side 2. The wavefronts on side 1 are normal to the ray directions on side 1.

We assume equality of mixed partial derivatives. Equality of mixed partial derivatives gives the equations of constraint for the 4x4 ray matrix. However, we will derive these equations of constraints in a very straightforward way by looking at the types of ray matrices that we can multiply together to form the total 4x4 ray matrix.

Given the solution for $\theta(x_1,y_1,x_2,y_2)$, the Huygens integral transformation can be written down as:

$$E_2\left(x_2,y_2\right)=\iint E_1\left(x_1,y_1\right)K\left(x_2,y_2,x_1,y_1\right)dx_1\,dy_1 \tag{3}$$

where the kernel is given by

$$K\left(x_2,y_2,x_1,y_1\right)=N\exp\left[i\,k\,\theta\left(x_2,y_2,x_1,y_1\right)\right] \tag{4}$$

and N is a normalizing constant chosen so that

$$\iint\left|E_2\left(x_2,y_2\right)\right|^2 dx_2\,dy_2=\iint\left|E_1\left(x_1,y_1\right)\right|^2 dx_1\,dy_1 \tag{5}$$

It is not immediately obvious that N depends only upon the 4x4 ray matrix. We will show that this is so and derive an expression for N.

There are two key intermediate results which express either the side 1 slopes or the side 2 slopes in terms of x_1, y_1, x_2, and y_2. We have from Eq (1):

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix}=\mathbf{A}\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}+\mathbf{B}\begin{bmatrix} n_1\,x_1' \\ n_1\,y_1' \end{bmatrix}$$
$$\begin{bmatrix} n_2\,x_2' \\ n_2\,y_2' \end{bmatrix}=\mathbf{C}\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}+\mathbf{D}\begin{bmatrix} n_1\,x_1' \\ n_1\,y_1' \end{bmatrix} \tag{6}$$

We can eliminate the side 1 slopes to obtain an equation for the side 2 slopes:

$$\begin{bmatrix} n_2\,x_2' \\ n_2\,y_2' \end{bmatrix}=\left(\mathbf{C}-\mathbf{D}\,\mathbf{B}^{-1}\,\mathbf{A}\right)\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}+\mathbf{D}\,\mathbf{B}^{-1}\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} \tag{7}$$

Similarly, the side 1 slopes are given by:

$$\begin{bmatrix} n_1 x_1' \\ n_1 y_1' \end{bmatrix} = -\mathbf{B}^{-1}\,\mathbf{A}\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \mathbf{B}^{-1}\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} \tag{8}$$

Use Eq (2) to obtain

$$\begin{bmatrix} \dfrac{\partial\theta}{\partial x_2} \\ \dfrac{\partial\theta}{\partial y_2} \end{bmatrix} = \left(\mathbf{C} - \mathbf{D}\,\mathbf{B}^{-1}\,\mathbf{A}\right)\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \mathbf{D}\,\mathbf{B}^{-1}\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} \tag{9}$$

and

$$\begin{bmatrix} \dfrac{\partial\theta}{\partial x_1} \\ \dfrac{\partial\theta}{\partial y_1} \end{bmatrix} = \mathbf{B}^{-1}\,\mathbf{A}\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} - \mathbf{B}^{-1}\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} \tag{10}$$

Integrating Eq (9) and Eq (10) we get:

$$\theta = \frac{1}{2}\left\{\begin{bmatrix} x_2 & y_2 \end{bmatrix}\mathbf{D}\,\mathbf{B}^{-1}\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} + 2\begin{bmatrix} x_2 & y_2 \end{bmatrix}\left(\mathbf{C} - \mathbf{D}\,\mathbf{B}^{-1}\,\mathbf{A}\right)\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \begin{bmatrix} x_1 & y_1 \end{bmatrix}\mathbf{B}^{-1}\,\mathbf{A}\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}\right\} \tag{11}$$

A number of other forms are possible depending upon the order in which the integrations are done. We shall show that all these forms are equivalent because of the constraint equations which we shall derive next. We shall also be able to simplify Eq (11) using these constraint equations. Note that this becomes the following form for the case of axial symmetry:

$$\theta = \frac{1}{2\,B}\left\{D\left(x_2^2 + y_2^2\right) - 2\left(x_1\,x_2 + y_1 y_2\right) + A\left(x_1^2 + y_1^2\right)\right\} \tag{12}$$

Here we have used the fact that the determinant of the generalized 2x2 ray matrix equals one. This is Eq (12) on page 781 in section 20.1 of Siegman's *Lasers*.

We now derive the constraint equations. There are two possible kinds of 4x4 matrices: propagation by a length and refraction or reflection at a curved surface. One can also allow rotation of coordinates. From appendices A and B, we see that refraction or reflection at a curved surface gives a 4x4 ray matrix which can be expressed as the product of three 4x4 matrices of the form:

$$\begin{bmatrix} \mathbf{d}_2^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{d}_2 \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{C} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{d}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{d}_1^{-1} \end{bmatrix} \tag{13}$$

where $\mathbf{C}$ is a symmetric 2x2 matrix, $\mathbf{0}$ is the 2x2 zero matrix, $\mathbf{I}$ is the 2x2 identity matrix, and $\mathbf{d}_1$ and $\mathbf{d}_2$ are 2x2 diagonal matrices. This is in a special coordinate system where the ray is obliquely incident in the xz plane upon a general quadratic refracting or reflecting surface, but we will consider rotations of coordinates separately. Here the value of working with the generalized slopes shows. This matrix would not have had this simple form if we hadn't. There are actually two kinds of matrices of matrices appearing in the product in Eq (13). The center matrix has an inverse given by:

$$\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{C} & \mathbf{I} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{C} & \mathbf{I} \end{bmatrix} \tag{14a}$$

The left and right matrices are of the same form and have an inverse obtainable from

$$\begin{bmatrix} \mathbf{d} & \mathbf{0} \\ \mathbf{0} & \mathbf{d}^{-1} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{d}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{d} \end{bmatrix} \tag{14b}$$

Propagation by a length L gives the 4x4 ray matrix:

$$\begin{bmatrix} \mathbf{I} & L\mathbf{I} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \tag{15}$$

which has the inverse:

$$\begin{bmatrix} \mathbf{I} & L\mathbf{I} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{I} & -L\mathbf{I} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \tag{16}$$

For completeness, the 4x4 matrix for a rotation of coordinates has the form:

$$\begin{bmatrix} \cos\theta\mathbf{I} & +\sin\theta\mathbf{I} \\ -\sin\theta\mathbf{I} & \cos\theta\mathbf{I} \end{bmatrix} \tag{17}$$

which has the inverse:

$$\begin{bmatrix} \cos\theta\mathbf{I} & +\sin\theta\mathbf{I} \\ -\sin\theta\mathbf{I} & \cos\theta\mathbf{I} \end{bmatrix}^{-1} = \begin{bmatrix} \cos\theta\mathbf{I} & -\sin\theta\mathbf{I} \\ +\sin\theta\mathbf{I} & \cos\theta\mathbf{I} \end{bmatrix} \tag{18}$$

A 4x4 ray matrix **Q** is a product of the form $\mathbf{Q}_1\ \mathbf{Q}_2\ \mathbf{Q}_3 \ldots \mathbf{Q}_n$ where each of the $\mathbf{Q}_i$ is a 4x4 matrix of one of the four forms given in Eq's (14a), (14b), (15), and (17).

For each of the four different kinds of Q matrices given in Eq's (14a), (14b), (15), and (17), the following relation is valid:

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{T} \begin{bmatrix} \mathbf{0} & -\mathbf{I} \\ \mathbf{I} & \mathbf{0} \end{bmatrix} \tag{19}$$

Here the superscript T denotes matrix transposition. Eq (19) is of course not true for an arbitrary 4x4 matrix, but it is true for a 4x4 ray matrix of the kind in Eq (14a), (14b), (15), or (17). For the kind of matrix in Eq (14a), it is imperative that $\mathbf{C}=\mathbf{C}^T$, as is indeed the case.

Define the matrix **S** by

$$\mathbf{S} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{I} & \mathbf{0} \end{bmatrix} \tag{20}$$

We have

$$\mathbf{S}\left(\mathbf{Q}_1\,\mathbf{Q}_2\,\mathbf{Q}_3 \ldots \mathbf{Q}_n\right)^{-1}\mathbf{S}^{-1} = \left(\mathbf{S}\mathbf{Q}_n^{-1}\mathbf{S}^{-1}\right)\left(\mathbf{S}\mathbf{Q}_{n-1}^{-1}\mathbf{S}^{-1}\right)\ldots\left(\mathbf{S}\mathbf{Q}_2^{-1}\mathbf{S}^{-1}\right)\left(\mathbf{S}\mathbf{Q}_1^{-1}\mathbf{S}^{-1}\right) \tag{21}$$

Using Eq (19) we get

$$\mathbf{S}\left(\mathbf{Q}_1\,\mathbf{Q}_2\,\mathbf{Q}_3 \ldots \mathbf{Q}_n\right)^{-1}\mathbf{S}^{-1} = \mathbf{Q}_n^T\ \mathbf{Q}_{n-1}^T \ldots \mathbf{Q}_2^T\ \mathbf{Q}_1^T = \left(\mathbf{Q}_1\ \mathbf{Q}_2 \ldots \mathbf{Q}_n\right)^T \tag{22}$$

Therefore

$$\left(\mathbf{Q}_1\,\mathbf{Q}_2\,\mathbf{Q}_3\ldots\mathbf{Q}_n\right)^{-1} = \mathbf{S}^{-1}\left(\mathbf{Q}_1\,\mathbf{Q}_2\ldots\mathbf{Q}_n\right)^T\mathbf{S} \tag{23}$$

Now

$$\mathbf{Q}_1\,\mathbf{Q}_2\,\mathbf{Q}_3\ldots\mathbf{Q}_n = \begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{bmatrix} \tag{24}$$

Therefore

$$\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{bmatrix}^{-1} = \begin{bmatrix}\mathbf{D}^T & -\mathbf{B}^T\\ -\mathbf{C}^T & \mathbf{A}^T\end{bmatrix} \tag{25}$$

We are very near having the constraint equations. Since a matrix times its inverse is the identity matrix and since it doesn't matter whether we multiply a matrix either on the left or on the right by its inverse, we have:

$$\begin{bmatrix}\mathbf{D}^T & -\mathbf{B}^T\\ -\mathbf{C}^T & \mathbf{A}^T\end{bmatrix}\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{bmatrix} = \begin{bmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\end{bmatrix} \tag{26}$$

and

$$\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{bmatrix}\begin{bmatrix}\mathbf{D}^T & -\mathbf{B}^T\\ -\mathbf{C}^T & \mathbf{A}^T\end{bmatrix} = \begin{bmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\end{bmatrix} \tag{27}$$

This gives us the constraint equations:

$$\mathbf{A}\,\mathbf{D}^T - \mathbf{B}\,\mathbf{C}^T = \mathbf{I} \tag{28}$$

$$\mathbf{D}\,\mathbf{A}^T - \mathbf{C}\,\mathbf{B}^T = \mathbf{I} \tag{29}$$

$$\mathbf{D}^T\,\mathbf{A} - \mathbf{B}^T\,\mathbf{C} = \mathbf{I} \tag{30}$$

$$\mathbf{A}^T\,\mathbf{D} - \mathbf{C}^T\,\mathbf{B} = \mathbf{I} \tag{31}$$

$$\mathbf{B}\,\mathbf{A}^T = \mathbf{A}\,\mathbf{B}^T \tag{32}$$

$$\mathbf{C}\,\mathbf{D}^T = \mathbf{D}\,\mathbf{C}^T \tag{33}$$

$$\mathbf{D}^T\ \mathbf{B} = \mathbf{B}^T\ \mathbf{D} \tag{34}$$

$$\mathbf{A}^T\ \mathbf{C} = \mathbf{C}^T\ \mathbf{A} \tag{35}$$

Eq (29) and Eq (31) are just the transposes of Eq (28) and Eq (30), respectively. We can simplify the $\mathbf{C} - \mathbf{D}\ \mathbf{B}^{-1}\ \mathbf{A}$ matrix that appears in Eq (11). Using first Eq (31) and then Eq (34) we obtain

$$\mathbf{C}^T = \mathbf{A}^T\ \mathbf{D}\,\mathbf{B}^{-1} - \mathbf{B}^{-1} = \mathbf{A}^T \left(\mathbf{D}\,\mathbf{B}^{-1}\right)^T - \mathbf{B}^{-1} \tag{37}$$

Thus

$$\mathbf{B}^{-1} = \mathbf{A}^T \left(\mathbf{D}\,\mathbf{B}^{-1}\right)^T - \mathbf{C}^T = -\left(\mathbf{C} - \mathbf{D}\,\mathbf{B}^{-1}\ \mathbf{A}\right)^T \tag{38}$$

Thus we may rewrite Eq (11) as

$$\theta = \frac{1}{2}\left\{\begin{bmatrix} x_2 & y_2 \end{bmatrix}\mathbf{D}\,\mathbf{B}^{-1}\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} - 2\begin{bmatrix} x_1 & y_1 \end{bmatrix}\mathbf{B}^{-1}\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} + \begin{bmatrix} x_1 & y_1 \end{bmatrix}\mathbf{B}^{-1}\ \mathbf{A}\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}\right\} \tag{39}$$

We have used the fact that a scalar equals its transpose. This agrees with Eq (84) on page 618 in section 15.6 of Siegman's *Lasers*. Siegman quotes this result without derivation. The original reference is Nazarathy's D. Sc. dissertation written in Hebrew for the Israel Institute of Technology.

The normalizing factor N mentioned earlier may be readily shown to be given by

$$N = \frac{1}{\sqrt{|\mathbf{B}|}\ \lambda} \tag{40}$$

except for a phase factor. The absolute phase information is lost when obtaining the eikonal from the 4x4 ray matrix. Siegman chooses the phase factor to agree with the phase factor for a propagation length in free space, but even if given the 4x4 ray matrix for a pure propagation length, one of course doesn't know that the real optical system is in fact just a pure propagation length. The key mathematical formula in this straightforward calculation is the formula

$$\iint \exp\left(\mathbf{r}_1^T\ \mathbf{Q}\,\mathbf{r}_2\ \right) d^2\mathbf{r}_2 = 2\pi\ \delta^{(2)}\left(\mathbf{r}_1^T\ \mathbf{Q}\right) = 2\pi\left|\mathbf{Q}\right|^{-1}\ \delta^{(2)}\left(\mathbf{r}_1\right) \tag{41}$$

where $\delta^{(2)}(\mathbf{x})$ is the two-dimensional delta function.

The final result for the kernel of the Huygens integral transformation is:

$$K\left(x_2,y_2,x_1,y_1\right)=\frac{1}{\sqrt{|\mathbf{B}|}\,\lambda}\exp\left[i\frac{\pi}{\lambda}\left(\mathbf{r}_2^T\,\mathbf{D}\,\mathbf{B}^{-1}\,\mathbf{r}_2-2\,\mathbf{r}_1^T\,\mathbf{B}^{-1}\,\mathbf{r}_2+\mathbf{r}_1^T\,\mathbf{B}^{-1}\,\mathbf{A}\,\mathbf{r}_1\right)\right] \tag{42}$$

The difference in the exponent signs between this result and the result quoted in Siegman is due to differing phase conventions.

We still have the problem of implementing this numerically. The $r_1^T\,B^{-1}\,A\,r_1$ and $r_2^T\,D\,B^{-1}\,r_2$ terms represent quadratic phase factors to be applied beforehand and afterwards, respectively. It is the $-2\,r_1^T\,B^{-1}\,r_2$ term that represents the Fourier transform. Let $r_1=(x_1,y_1)$ and $r_2=(x_2,y_2)$. We have:

$$\mathbf{r}_1^T\,\mathbf{B}^{-1}\,\mathbf{r}_2=\left(\mathbf{B}^{-1}\right)_{11}x_1\,x_2+\left(\mathbf{B}^{-1}\right)_{12}x_1\,y_2+\left(\mathbf{B}^{-1}\right)_{21}y_1\,x_2+\left(\mathbf{B}^{-1}\right)_{22}y_1\,y_2 \tag{43}$$

Make the side 2 change of variables:

$$\begin{aligned}\overline{x}_2&=\left(\mathbf{B}^{-1}\right)_{11}x_2+\left(\mathbf{B}^{-1}\right)_{12}y_2\\ \overline{y}_2&=\left(\mathbf{B}^{-1}\right)_{21}x_2+\left(\mathbf{B}^{-1}\right)_{22}y_2\end{aligned} \tag{44}$$

We then have

$$\mathbf{r}_1^T\,\mathbf{B}^{-1}\,\mathbf{r}_2=\mathbf{r}_1^T\;\overline{\mathbf{r}}_2 \tag{45}$$

Let dx_1 and dy_1 be the side 1 grid for which the complex wavefront is known. We wish to propagate to side 2. Make the further change of variables:

$$\begin{aligned}x_1&=k\;dx_1 \quad & y_1&=m\;dy_1\\ \overline{x}_2&=l\;d\overline{x}_2 \quad & \overline{y}_2&=n\;d\overline{y}_2\end{aligned} \tag{46}$$

We then obtain:

$$\exp\left[-i\frac{2\pi}{\lambda}\,\mathbf{r}_1^T\,\mathbf{B}^{-1}\,\mathbf{r}_2\right]=\exp\left[-i\,2\pi\,k\,l\frac{dx_1\,d\overline{x}_2}{\lambda}\right]\exp\left[-i\,2\pi\,m\,n\frac{dy_1\,d\overline{y}_2}{\lambda}\right] \tag{47}$$

We now require that

$$\frac{dx_1\, d\overline{x}_2}{\lambda} = \frac{1}{N_{columns}} \tag{48}$$

and

$$\frac{dy_1\, d\overline{y}_2}{\lambda} = \frac{1}{N_{rows}} \tag{49}$$

Here N_{rows} and $N_{columns}$ are the numbers of rows and columns in the side 1 image, respectively. They must be of the form $2^m 3^n$ for suitable non-negative integers m and n since I have implemented a general mixed radix FFT for the radix values 2 and 3. If the image dimensions aren't of this form, the image must be padded to make it of this form. The spacings dx_1 and dy_1 are known, so that the grid on side 2 is determined. We finally obtain the result:

$$\exp\left[-i\frac{2\pi}{\lambda}\mathbf{r}_1^T\,\mathbf{B}^{-1}\,\mathbf{r}_2\right] = \exp\left[-\frac{2\pi\, i\, k\, l}{N_{columns}}\right]\exp\left[-\frac{2\pi\, i\, m\, n}{N_{rows}}\right] \tag{50}$$

This is the correct form for the FFT.

It is always necessary to do the Huygens integral transformation in two steps. The first step is from surface 1 to focus while the second step is from focus to surface 2. A propagation to focus may be fictitiously created by representing the ray matrix in the form:

$$\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{bmatrix} = \left\{\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{bmatrix}\left(\begin{bmatrix}\mathbf{I} & f\,\mathbf{I}\\ \mathbf{0} & \mathbf{I}\end{bmatrix}\begin{bmatrix}\mathbf{I} & 0\\ -\mathbf{I}/f & \mathbf{I}\end{bmatrix}\right)^{-1}\right\}\left(\begin{bmatrix}\mathbf{I} & f\,\mathbf{I}\\ \mathbf{0} & \mathbf{I}\end{bmatrix}\begin{bmatrix}\mathbf{I} & 0\\ -\mathbf{I}/f & \mathbf{I}\end{bmatrix}\right) \tag{51}$$

The ray matrix from the first surface to focus is what's in the first parentheses while the ray matrix from focus to the second surface is what's in the curly brackets in Eq (51). The propagation to focus may be done by using the Huygens integral transformation for a 2x2 ray matrix, of course. The focus f is somewhat arbitrary. It should be chosen so that the f# is reasonably high, but it shouldn't be chosen hugely large. One wants the focal spot to be very small compared to the sizes of the images on surfaces 1 and 2. I choose the focus f to either be the distance from the top pinhole to the CCD for lensless imaging or on the order of two or three times what I think is a representative "focal length" for the system.

APPENDIX A Derivation of the Huygens 4x4 ray matrix for oblique incidence on a quadratic refractive surface

I shall first derive the 2x2 ray matrices for the special cases of tangential and sagittal incidence on a spherical surface. I will then generalize the equations used to derive these 2x2 ray matrices so that the 4x4 ray matrix for a general quadratic surface may be obtained.

Tangential Incidence

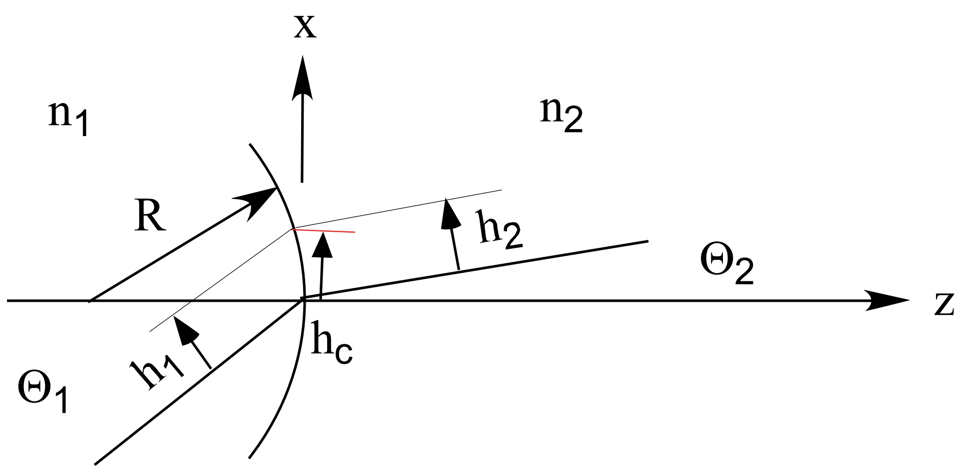

FIGURE A1

Let the ray on side 1 be incident at angle Θ_1 to the normal, and, after refraction, be at angle Θ_2 to the normal on side 2. Snell's law states that:

$$n_1 \sin\Theta_1 = n_2 \sin\Theta_2 \tag{A.1}$$

Differentiation gives

$$n_1 \cos\Theta_1 \;\delta\Theta_1 = n_2 \cos\Theta_2 \;\delta\Theta_2 \tag{A.2}$$

so that changing the angle of a ray to the normal gives:

$$\delta\, n_2\Theta_2 = \frac{\cos\Theta_1}{\cos\Theta_2}\,\delta\, n_1\Theta_1 \tag{A.3}$$

We also have

$$h_c = h_1 \sec\Theta_1 = h_2 \sec\Theta_2 \tag{A.4}$$

For a ray incident at the origin, the angle with respect to the normal is the same as the angle with respect to the z axis. However, in general the angles are different. Let θ_1 and θ_2 be the angles with respect to the z axis and Θ_1 and Θ_2 be the angles with respect to the surface normal.

Let the ray be incident at $x = h_c$. Then we have the following relationships between the angles in the paraxial approximation where $h_c << R$:

$$\begin{aligned} \delta\Theta_1 &= \delta\theta_1 - \frac{h_c}{R} \\ \delta\Theta_2 &= \delta\theta_2 - \frac{h_c}{R} \end{aligned} \tag{A.5}$$

Therefore

$$n_2\left(\delta\,\theta_2 - \frac{h_c}{R}\right) = \frac{\cos\Theta_1}{\cos\Theta_2} n_1\left(\delta\,\theta_1 - \frac{h_c}{R}\right) \tag{A.6}$$

From Eq's (4) and (6) we obtain the 2x2 ray matrix

$$\begin{bmatrix} h_2 \\ \delta n_2\theta_2 \end{bmatrix} = \begin{bmatrix} \dfrac{\cos\theta_2}{\cos\theta_1} & 0 \\ \dfrac{n_2\cos\theta_2 - n_1\cos\theta_1}{R\cos\theta_1\cos\theta_2} & \dfrac{\cos\theta_1}{\cos\theta_2} \end{bmatrix} \begin{bmatrix} h_1 \\ \delta n_1\theta_1 \end{bmatrix} \tag{A.7}$$

This agrees with the 2x2 ray matrix in Table 15.1 (f) on page 586 of Siegman's *Lasers*. Note that in the spirit of a first order approximation, we are free to interchange θ_1 and Θ_1 and to interchange θ_2 and Θ_2 everywhere except inside a variation δ.

Sagittal Incidence

Let us first obtain the relation between the sagittal angles $\delta\Theta_{y1}$ and $\delta\Theta_{y2}$ for a ray hitting the origin. The simplest way is to simply perform a rotation by a small angle ε about the z axis. One then obtains:

$$\begin{aligned} \delta\Theta_{y1} &= \varepsilon\sin\Theta_1 \\ \delta\Theta_{y2} &= \varepsilon\sin\Theta_2 \end{aligned} \tag{A.8}$$

Therefore

$$\delta n_2 \Theta_{y2} = \frac{n_2 \sin\Theta_2}{n_1 \sin\Theta_1} \delta n_1 \Theta_{y1} \tag{A.9}$$

Using Snell's law Eq. (1), this simplifies to:

$$\delta n_2 \Theta_{y2} = \delta n_1 \Theta_{y1} \tag{A.10}$$

We also have

$$h_{y2} = h_{y1} \tag{A.11}$$

We also have the relations:

$$\begin{aligned} \delta\Theta_{y1} &= \delta\theta_{y1} - \frac{h_y}{R}\cos\Theta_1 \\ \delta\Theta_{y2} &= \delta\theta_{y2} - \frac{h_y}{R}\cos\Theta_2 \end{aligned} \tag{A.13}$$

where since $h_{1y} = h_{2y}$, I have called their common value h_y. Combining Eq's (10) and (13) gives:

$$n_2\left(\delta\theta_{y2} - \frac{h_y}{R}\cos\Theta_2\right) = n_1\left(\delta\theta_{y1} - \frac{h_y}{R}\cos\Theta_1\right) \tag{A.14}$$

From Eq's (11) and (14), we obtain the 2x2 ray matrix for sagittal incidence:

$$\begin{bmatrix} h_{2y} \\ \delta n_2 \theta_{2y} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \dfrac{n_2 \cos\theta_2 - n_1 \cos\theta_1}{R} & 1 \end{bmatrix} \begin{bmatrix} h_{1y} \\ \delta n_1 \theta_{1y} \end{bmatrix} \tag{A.15}$$

Again, in the spirit of a first order approximation, I have freely interchanged θ_1 and Θ_1 and have also freely interchanged θ_2 and Θ_2 everywhere except inside a variation δ. This agrees with Table 15.1 (g) on page 586 in Siegman's *Lasers*.

Generalization

Suppose the quadratic surface has the equation:

$$z = a x^2 + 2 b x y + c y^2 \tag{A.16}$$

Alternatively we can write this surface equation as

$$f(x, y, z) = z - \left(a x^2 + 2 b x y + c y^2\right) = 0 \tag{A.17}$$

The normal is

$$\left[-2(a x + b y) \quad -2(b x + c y) \quad 1\right] \tag{A.18}$$

For the quadratic approximation to a concave spherical surface, note that

$$a = -\frac{1}{2R} \quad b = 0 \quad c = -\frac{1}{2R} \tag{A.19}$$

Eq (5) generalizes to

$$\begin{aligned} \delta\Theta_{1x} &= \delta\theta_{1x} + 2(a x + b y) \\ \delta\Theta_{2x} &= \delta\theta_{2x} + 2(a x + b y) \end{aligned} \tag{A.20}$$

Substituting Eq (20) into Eq (3) gives the generalization of Eq (6):

$$n_2\left(\delta\,\theta_{2x} + 2(a x + b y)\right) = \frac{\cos\Theta_1}{\cos\Theta_2} n_1\left(\delta\,\theta_{1x} + 2(a x + b y)\right) \tag{A.21}$$

Similarly, Eq (13) generalizes to:

$$\begin{aligned} \delta\Theta_{y1} &= \delta\theta_{y1} + 2(b x + c y) \\ \delta\Theta_{y2} &= \delta\theta_{y2} + 2(b x + c y) \end{aligned} \tag{A.22}$$

Substitutin Eq (22) into Eq (10) gives the generalization of Eq (14):

$$n_2\left(\delta\theta_{y2} + 2(b x + c y)\cos\Theta_2\right) = n_1\left(\delta\theta_{y1} + 2(b x + c y)\cos\Theta_1\right) \tag{A.23}$$

Since

$$\begin{aligned} x &= h_{1x}\sec\Theta_1 \\ y &= h_{1y} \end{aligned} \tag{A.24}$$

Combining Eq's (4), (11), (21), (23), and (24), we obtain the 4x4 ray matrix:

$$\begin{bmatrix} h_{2x} \\ h_{2y} \\ \delta n_2\theta_{2x} \\ \delta n_2\theta_{2y} \end{bmatrix} = \begin{bmatrix} \dfrac{\cos\theta_2}{\cos\theta_1} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ C_{11} & C_{12} & \dfrac{\cos\theta_1}{\cos\theta_2} & 0 \\ C_{21} & C_{22} & 0 & 1 \end{bmatrix} \begin{bmatrix} h_{1x} \\ h_{1y} \\ \delta n_1\theta_{1x} \\ \delta n_1\theta_{1y} \end{bmatrix} \tag{A.25}$$

where the elements of the 2x2 C sub-matrix are given by:

$$\begin{aligned} C_{11} &= -2a\left(\frac{n_2\cos\theta_2 - n_1\cos\theta_1}{\cos\theta_2\cos\theta_1}\right) \\ C_{12} &= -2b\left(\frac{n_2\cos\theta_2 - n_1\cos\theta_1}{\cos\theta_2}\right) \\ C_{21} &= -2b\left(\frac{n_2\cos\theta_2 - n_1\cos\theta_1}{\cos\theta_1}\right) \\ C_{11} &= -2c\left(n_2\cos\theta_2 - n_1\cos\theta_1\right) \end{aligned} \tag{A.26}$$

As usual, in the spirit of a first order approximation, I have freely interchanged θ_1 and Θ_1 and have also freely interchanged θ_2 and Θ_2 everywhere except inside a variation δ. The result in Eq (26) can be readily generalized to a reflecting surface by letting n_2 go to $-n_1$ and by letting the side 2 slope $\delta\theta_{2x}$ go to - $\delta\theta_{2x}$ and the side 2 slope $\delta\theta_{2y}$ go to - $\delta\theta_{2y}$. Note that the side 2 generalized slopes $\delta\ n_2\theta_{2x}$ and $\delta\ n_2\theta_{2y}$ go to themselves.

APPENDIX B ALTERNATIVE FORM FOR THE 4x4 RAY MATRIX FOR A OBLIQUES INCIDENCE ON A GENERAL REFRACTIVE QUADRATIC SURFACE

The geometry in figure A1 in appendix A is used. The quadratic surface has an equation of the form:

$$z(x,y) = a\,x^2 + 2b\,x\,y + c\,y^2 \tag{B.1}$$

The coordinate system has been chosen so that the chief ray on side 1 is in the xz plane at so that it hits the refractive surface at the origin. This chief ray is incident at angle θ_1 to the surface normal so that it has direction cosines ($\sin\theta_1$, 0, $\cos\theta_1$). The refractive index on side 1 is n_1 and the refractive index on side 2 is n_2. Appendix A shows that the 4x4 ray matrix for the heights and generalized slopes is given by:

$$\begin{bmatrix} h_{2x} \\ h_{2y} \\ \delta n_2\theta_{2x} \\ \delta n_2\theta_{2y} \end{bmatrix} = \begin{bmatrix} \dfrac{\cos\theta_2}{\cos\theta_1} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ C_{11} & C_{12} & \dfrac{\cos\theta_1}{\cos\theta_2} & 0 \\ C_{21} & C_{22} & 0 & 1 \end{bmatrix} \begin{bmatrix} h_{1x} \\ h_{1y} \\ \delta n_1\theta_{1x} \\ \delta n_1\theta_{1y} \end{bmatrix} \tag{B.2}$$

where the elements of the 2x2 C sub-matrix are given by:

$$\begin{aligned} C_{11} &= -2a\left(\frac{n_2\cos\theta_2 - n_1\cos\theta_1}{\cos\theta_2\cos\theta_1}\right) \\ C_{12} &= -2b\left(\frac{n_2\cos\theta_2 - n_1\cos\theta_1}{\cos\theta_2}\right) \\ C_{21} &= -2b\left(\frac{n_2\cos\theta_2 - n_1\cos\theta_1}{\cos\theta_1}\right) \\ C_{11} &= -2c\left(n_2\cos\theta_2 - n_1\cos\theta_1\right) \end{aligned} \tag{B.3}$$

Notice that C is no longer a symmetric matrix and the two 2x2 sub-matrices on the diagonal are no longer the 2x2 identity matrix. However a simple transformation of variables puts the matrix into this form:

$$\begin{bmatrix} h_{2x}\sec\theta_2 \\ h_{2y} \\ n_2\theta_{2x}\cos\theta_2 \\ n_2\theta_{2y} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \bar{C}_{11} & \bar{C}_{12} & 1 & 0 \\ \bar{C}_{21} & \bar{C}_{22} & 0 & 1 \end{bmatrix} \begin{bmatrix} h_{1x}\sec\theta_1 \\ h_{1y} \\ n_1\theta_{1x}\cos\theta_1 \\ n_1\theta_{1y} \end{bmatrix} \tag{B.4}$$

where the new symmetric 2x2 sub-matrix C bar has the form:

$$\begin{aligned} \bar{C}_{11} &= -2a\left(n_2\cos\theta_2 - n_1\cos\theta_1\right) \\ \bar{C}_{12} &= -2b\left(n_2\cos\theta_2 - n_1\cos\theta_1\right) \\ \bar{C}_{21} &= -2b\left(n_2\cos\theta_2 - n_1\cos\theta_1\right) \\ \bar{C}_{11} &= -2c\left(n_2\cos\theta_2 - n_1\cos\theta_1\right) \end{aligned} \tag{B.5}$$

The 4x4 matrix which effects this transformation of variables has the form:

$$\bar{Q}(\theta) = \begin{bmatrix} \cos\theta & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \sec\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{B.6}$$

Notice that

$$\bar{Q}(\theta)^{-1} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{I} & \mathbf{0} \end{bmatrix} \bar{Q}(\theta)^{T} \begin{bmatrix} \mathbf{0} & -\mathbf{I} \\ \mathbf{I} & \mathbf{0} \end{bmatrix} \tag{B.7}$$

Its inverse also has this property:

$$\left[\bar{Q}^{-1}(\theta)\right]^{-1} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{I} & \mathbf{0} \end{bmatrix} \left[\bar{Q}(\theta)^{-1}\right]^{T} \begin{bmatrix} \mathbf{0} & -\mathbf{I} \\ \mathbf{I} & \mathbf{0} \end{bmatrix} \tag{B.8}$$

More generally, any matrix of the form

$$Q = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^{-1} \end{bmatrix} \tag{B.9}$$

where **A** is any symmetric non-singular 2x2 sub-matrix and **0** is the zero 2x2 sub-matrix, satisfies the equation

$$\bar{Q}^{-1} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{I} & \mathbf{0} \end{bmatrix} \bar{Q}^{T} \begin{bmatrix} \mathbf{0} & -\mathbf{I} \\ \mathbf{I} & \mathbf{0} \end{bmatrix} \tag{B.10}$$