Journal for Geometry and Graphics
Volume 4 (2000), No. 1, 1–18.

Reflections on Refractions

Georg Glaeser, Hans-Peter Schröcker

Institute for Architecture, University of Applied Arts Vienna
Oskar Kokoschka-Platz 2, A-1010 Wien, Austria
email: georg.glaeser@uni-ak.ac.at

Abstract. In computer graphics, it is often an advantage to calculate refractions directly, especially when the application is time-critical or when line graphics have to be displayed. We specify efficient formulas and parametric equations for the refraction on straight lines and planes. Furthermore, we develop a general theory of refractions, with reflections as a special case. In the plane case, all refracted rays are normal to a characteristic conic section. We investigate the relation of this conic section and the diacaustic curve. Using this, we can deduce properties of reciprocal refraction and a virtual object transformation that makes it possible to produce 2D-refraction images with additional depth information.
In the three-dimensional case, we investigate the counter image of a straight line. It is a very special ruled surface of order four. This yields results on the order of the refrax of algebraic curves and on the shading of refracted polygons. Finally, we provide a formula for the diacaustic of a circle.

Key Words: Refraction, reflection, curved perspectives, fish-eye perspectives, diacaustic, catacaustic, normal congruence, real-time rendering, underwater photography.
MSC 1994: 51N05, 51N35, 51N99, 68U05.

1. Introduction and state of the art

Refractions are to be seen everywhere in daily life. Diving in the sea or watching fish in an aquarium produces remarkable optical effects. Refractions play an important role in technical applications as well: eye glasses, optical lenses, underwater photography etc. Mathematicians have been interested in refraction phenomena for quite a while. We will now briefly describe the most important results of 300 years of research.

The first to investigate refraction and reflection were TSCHIRNHAUS and HUYGENS (around 1680). They and — a little later — Johann BERNOULLI were especially interested in *caustics*,

ISSN 1433-8157/\$ 2.50

the hull curves of a one parameter set of rays that are reflected or refracted on a plane curve.[1] Caustics produce nice optical effects because the light intensity is maximal along them. They also permit a deeper insight into many reflection and refraction phenomena. E.g., the catacaustic of a pencil $E(e)$ of rays with respect to a circle c is an algebraic curve of class four. Thus, a circle (or a sphere) has theoretically four specular points (see Fig. 1).[2]

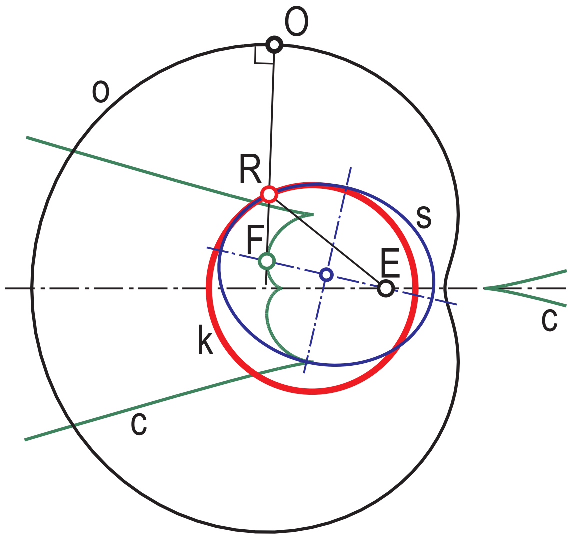

Figure 1: The catacaustic c of a circle k with respect to the light source E: The involute o of c is at the same time the orthonomic of k (a limaçon of Pascal). The conic section s with focal points E and F through R osculates k.

It also makes sense to investigate *caustic surfaces*, i.e., the focal surfaces of a two parameter set of rays refracted or reflected on a surface [8, 10]. Luckily, the spatial problem can sometimes be reduced to a planar problem.

The case where the set of rays being refracted is a pencil $E(e)$ is of special interest. It serves as a 2D-model for human perception as well as for illumination of a scene with refracting objects. In [7] and [14], a general method of constructing caustics in this case was introduced: The refracted rays are all perpendicular to a hull curve h of certain circles. Thus, the caustic is the evolute of h (compare Fig. 17).[3]

Since the days of Tschirnhaus and Huygens, the caustics for many special cases were described by a number of authors [3, 11, 15, 20].

Caustics of higher order were also studied. The light rays are not only refracted once but twice or even more often on a certain curve or surface. The problem of the n-th caustic of reflection on a circle, e.g., was solved in [9]. In this context the theorem of Malus is important: A two parameter set of straight lines is called a line *normal congruence* when it is the set of normals of a surface. The theorem of Malus now states *that a normal congruence remains normal after an arbitrary number of reflections or refractions.*[4]

In [5], the caustics of a pencil of lines $E(e)$ with respect to a plane curve $k \ \ldots \ \vec{x} = \vec{x}(t)$ were calculated in a general form. The same authors solved the problem of finding the

[1] The caustics of reflections are called *catacaustics*, those of refraction *diacaustics*. The term *caustic* refers to both cata- and diacaustics.

[2] It is quite remarkable that all four specular points can be of practical relevance (see [8]).

[3] In case of a reflection h is the orthonomic o of the reflecting curve with respect to E (compare Fig. 1).

[4] For a proof of this theorem, see [18]. Of course, an analogous theorem holds for the plane case.

"anticaustics", i.e., the one parameter set of curves producing a certain given caustic in [6]. The given formulas, however, and the related differential equations are quite complicate.

In [12] a completely different way of constructing the catacaustic of a pencil of lines $E(e)$ was introduced:

We regard a conic section s osculating the reflecting curve k and having E as one focal point. Then the corresponding point on the catacaustic is the focal point $F \neq E$ of this conic (see Fig. 1).

A lot of recent books and publications on computer graphics deal with the topic of reflection and refraction. Apart from rather basic considerations, however, they usually rely on ray-tracing methods or approximating calculations and hardly ever make use of the profound (but rather old and not well-known) theoretical background. Therefore we think it is time to present a new theory of reflection and refraction adapted to the needs of modern computer graphics.

2. The physical approach: Snell's Law, Fermat's principle

We will now present the physical basics of geometrical optics in Euclidean 2-space $\mathbb{E}^2$. We choose a straight *refracting line* s, an *eye point* $E \notin s$ and a positive real *fraction ratio* $\mathbf{r}$. For easier writing, $\mathcal{R}[s;\mathbf{r};E]$ shall denote the refraction on s with ratio $\mathbf{r}$ with respect to E. (This notation is generalization of the notation introduced in [8] for reflections.)

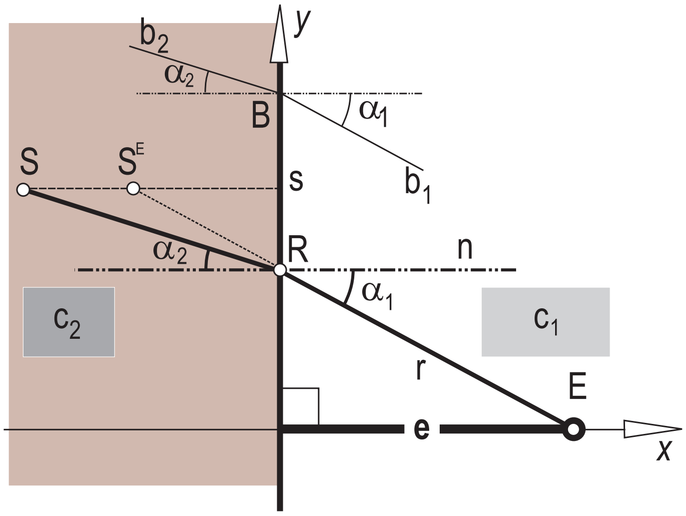

Figure 2: Refraction on a straight line

Physically speaking, $\mathbf{r}$ has the following meaning: When light propagates with speed c_1 on the side of E, it propagates with speed $c_2 = c_1/\mathbf{r}$ on the other side of s. For $\mathbf{r} > 1$, the side on E is "optically less dense". In Fig. 2, $\mathbf{r} \approx 1.33$ was chosen for the ratio of the light speed in the atmosphere (to the right) and water (to the left).

With $\mathcal{R}[s;\mathbf{r};E]$, we connect a Cartesian coordinate system as follows (Fig. 2): E is a point on the positive x-axis (position vector $\vec{e} = (\mathbf{e}, 0)^T$) and s is the y-axis.

Due to the physical law of refraction (SNELL*'s law*), a straight line b_1 (incidence angle α_1 to the normal of s) is refracted into a straight line $b_2 = \mathcal{R}[s;\mathbf{r};E](b_1)$ through $B = b_1 \cap s$ with incidence angle α_2 according to the equation

$$\sin\alpha_1 = \mathbf{r}\sin\alpha_2. \tag{1}$$

Though in principle we have $\alpha_1,\ \alpha_2\ \in [-\pi/2, \pi/2]$, there is a restriction on either α_1 or α_2: For $\mathbf{r} > 1$, the refracted ray will have a maximum angle of $|\alpha_2| \le \arcsin(1/\mathbf{r})$, for $\mathbf{r} < 1$ rays are only refracted when $|\alpha_1| \le \arcsin \mathbf{r}$. E.g., for $\mathbf{r} \approx 0.75$ (water $\longrightarrow$ air) we have $|\alpha_1| \le \alpha_1^{max} = 48.5°$. Refraction is always accompanied by reflection: The smaller the angle α_1 is, the less reflection occurs. For $|\alpha_1| \ge \alpha_1^{max}$, we have total reflection on s.

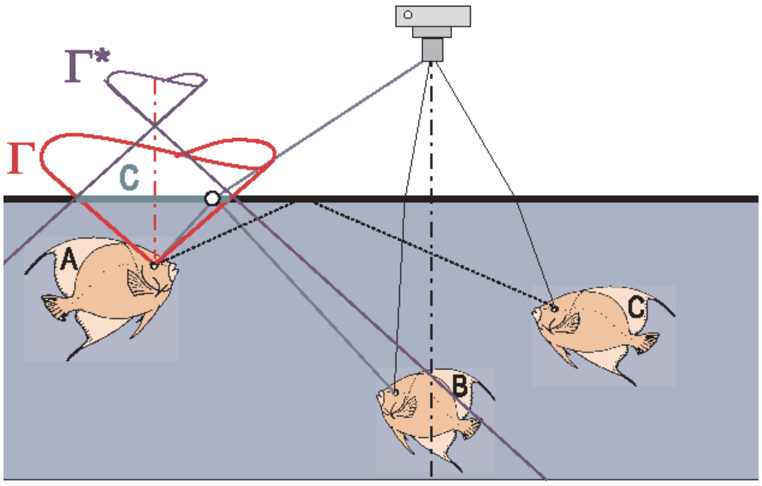

Figure 3: Who can see whom?

As a consequence, fish A in a calm pool (Fig. 3) will see

- "everything" outside the pool, though partly very distorted. The refracted image fills a circle c on the surface that stems from a cone of revolution Γ with apex angle $2 \times 48.5°$;
- the total reflections of those parts of the pool that are outside the reflected cone Γ^* (e.g., fish C);
- very dim reflections of the rest of the pool (e.g., fish B) inside c as a result of partial reflection;
- "everything" inside the pool, e.g. fish B and C.

When a person outside the pool takes a picture of the pool (e.g., from the spring board), the image will show all the fish. In the following section we will develop an efficient method of computing the seeming positions on this photo.

3. Refracting projecting rays through space points

Snell's law does not explicitely require the position of the eye point. Nevertheless, we will now take into account such a point, since we usually observe with our eye (or even two eyes, of course). Therefore, we will distinguish between projection rays through E and general rays.

Let us take a simple example: From the border of a pool, we are watching a fish swimming around. We all know that the fish is not at the position we see it. Our goal is now to solve the two problems:

1. Given the position S of a point on the fish's surface, we are looking for the projection ray r through our eye that runs through S after being refracted on the pools plane surface σ. The intersection point $R = r \cap \sigma$ will be the key for the determination of r.
2. Given the point R, can we say anything about the spatial position of S? Well, of course we cannot with only one eye, but what if we look two-eyed?

Speaking of fish: We call the photographic images created by refracting optical ultra-wide angle lenses "fish-eye perspectives". The creation of such curved perspectives is another motivation for the investigation of refractions.

To cut longer sentences short, we will henceforth use a new word:

Definition 1 *The "image" $R \in \Phi$ of a point S under the influence of a refraction on a surface Φ ($R = \mathcal{R}[\Phi; \mathbf{r}; E](S)$) is called refrax of S on Φ.*

This comes close to the word "reflex" for the image of a point in a (plane) mirror.

We now want to solve the first problem: Given a point S left to σ, we are looking for its refrax on σ, i.e., the point which we practically look at when we try to see S.

Obviously, the problem is two-dimensional: We consider the situation in an auxiliary plane ν through E and S perpendicular to σ. Due to the laws of optics, R will automatically lie in ν. In ν, the points have the coordinates $E(\mathbf{e}, 0)$, $S(s_x, s_y)$, $R(0, r_y)$. The refracting line is $s = \nu \cap \sigma$. Let again c_1 be the light speed on the eye point's side (e.g., outside the pool), and c_2 be the light speed on the other side (e.g., in the water $\Rightarrow$ $\mathbf{r} = c_1/c_2 \approx 1.33$). Snell's physical approach was now to minimize the time the light ray needs to propagate from S to E. Actually, the calculation used Fermat's principle:

When light travels from E to S, it travels along a path or ray for which the time taken (the "optical length") has a stationary value with respect to infinitesimal variations of the path (see, e.g., [18]):

$$\overline{ER}/c_1 + \overline{RS}/c_2 \;\rightarrow\; \text{min.} \;\Rightarrow\; \sqrt{\mathbf{e}^2 + r_y^2} + \mathbf{r}\sqrt{s_x^2 + (r_y - s_y)^2} \;\rightarrow\; \text{min.} \tag{2}$$

We introduce the variable $y = r_y$. Then the "total-time function" or "optical length function" (Fig. 4)

$$t(y) = \sqrt{\mathbf{e}^2 + y^2} + \mathbf{r}\sqrt{s_x^2 + (y - s_y)^2} \tag{3}$$

has to have a minimum:

$$t'(y) = \frac{y}{\sqrt{\mathbf{e}^2 + y^2}} + \frac{\mathbf{r}(y - s_y)}{\sqrt{s_x^2 + (y - s_y)^2}} = 0 \tag{4}$$

This leads to an algebraic equation $f(y)$ of fourth order in y:

$$f(y) = Ny^4 - 2Ns_y y^3 + (Ns_y^2 + \frac{s_x^2}{\mathbf{r}^2} - \mathbf{e}^2)y^2 + 2\mathbf{e}^2 s_y y - e^2 s_y^2 = 0 \quad \text{with } N = \frac{1}{\mathbf{r}^2} - 1. \tag{5}$$

We now prove

Theorem 1 *The calculation of the "refrax" $R = \mathcal{R}[s; \mathbf{r}; E](S)$ leads to the determination of the roots of an algebraic polynomial* (5) *of degree four. The only root y_0 that lies in the interval $[0, s_y]$ is the practical solution.*

Proof: Let $u(y) = \sqrt{\mathbf{e}^2 + y^2}$ and $v(y) = \mathbf{r}\sqrt{s_x^2 + (y - s_y)^2}$. All the solutions of (5) then fulfill $t'(y) = u'(y) + v'(y) = 0$, or $\widetilde{t}'(y) = u'(y) - v'(y) = 0$, respectively. Only those solutions that fulfill $t'(y) = 0$ (4) are practical solutions. We will now show that exactly only one of the roots y_i fulfills (4), whereas the residual ones fulfill $\widetilde{t}' = 0$ (Fig. 4).

We have to verify: $t(y)$ has only one position y_0 of extremal value. y_0 is the position of a minimum and is in the interval $[0, s_y]$. The positions of extremal values of $\widetilde{t}(y)$ are outside this interval. This is exactly the contents of the following Lemma 1 with $x_0 = 0$, $x_1 = s_y$. □

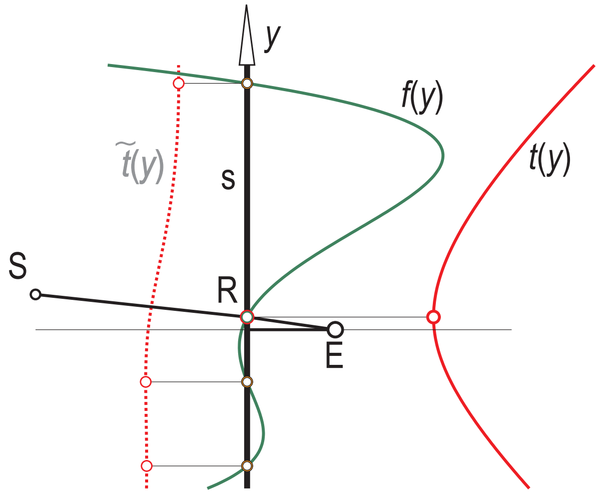

Figure 4: The four roots of $f(y)$

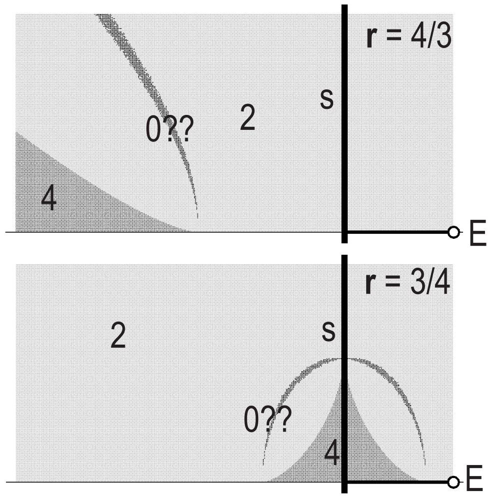

Figure 5: Number of roots

Lemma 1 *Let $I \subset \mathbb{R}$ be a real interval and u, v two strictly convex functions in $\mathcal{C}^\infty[I, \mathbb{R}]$. x_0 and $x_1 > x_0$ be positions of minimum of u and v, respectively. Then we have:*

1. *The function $t := u + v$ has a unique position of minimum x_m and $x_m \in [x_0, x_1]$.*
2. *The function $\widetilde{t} := u - v$ has no position of extremum in $[x_0, x_1]$.*

Proof: t is strictly convex as well, and has therefore at most one position of minimum and no position of maximum. Furthermore, strictly convex functions have strictly monotonous derivatives which gives $t'(x_0) = v'(x_0) < 0$ and $t'(x_1) = u'(x_1) > 0$. Now by the theorem of intermediate values there exists a real $x_m \in [x_0, x_1]$ satisfying $t'(x_m) = 0$. x_m is the uniquely determined position of minimum of t.

Let us now suppose that $x_m \in I$ is a position of minimum of $\widetilde{t}$. Then we necessarily get $u'(x_m) = v'(x_m)$. As $x_m \neq x_0, x_1$ and $u' > 0$ and $v' < 0$ in (x_0, x_1), x_m cannot be in $[x_0, x_1]$. □

Fig. 5 illustrates where we can expect four real roots y_i, and where only two can be found. Small areas around certain conics are numerically instable, i.e., we will not be able to verify (4) when we declare an ε that is too small for $|t'(y)| < \varepsilon$ (in Fig. 5, $\varepsilon = 10^{-11}$ was chosen; with $\varepsilon = 10^{-6}$, the verification was always OK). We will explain this behavior in Section 4; it is closely connected with the three residual roots of equation (5).

Anyway, the fast criterion $0 \le y \le s_y$ (or $s_y \le y \le 0$, respectively) works fine for *all* points $S \in \mathbb{E}^2$, even on the side of E, since the sign of s_x does not have an impact on (5).

The up to four real solutions of the polynomial (5) can be calculated by means of well known formulas [16].[5]

[5]When less accuracy is necessary, we can find the only practical usable root of the polynomial even a bit faster by means of NEWTON's iteration, since we explicitly have the equation of $f'(x)$:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{c_4 x_n^4 + c_3 x_n^3 + c_2 x_n^2 + c_1 x_n + c_0}{4c_4 x_n^3 + 3c_3 x_n^2 + 2c_2 x_n + c_1}. \tag{6}$$

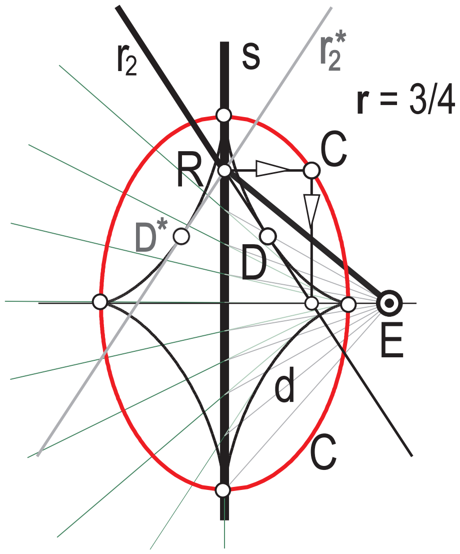

Figure 6: The characteristic ellipse

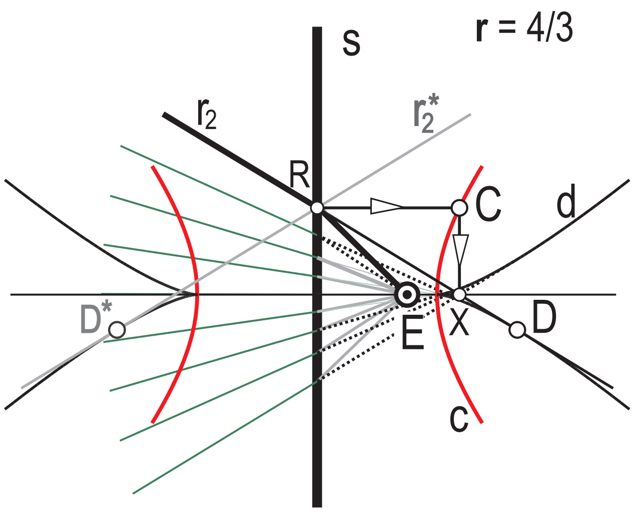

Figure 7: The characteristic hyperbola

4. Diacaustic and characteristic conic of $\mathcal{R}[s; \mathbf{r}; E]$

In this section, we will take a closer look at the diacaustic of $\mathcal{R}[s; \mathbf{r}; E]$. Actually, for given $\mathbf{r}$ and α_1 (1) has two solutions α_2^0 and $\pi - \alpha_2^0$. When we have a ray r_1, we will therefore assume two refracted rays r_2 and r_2^* that are symmetric with respect to the refracting line s. This is not appropriate in a physical model of refraction but here it makes sense as we will regard refractions in terms of algebraic geometry as well.

We have to mention a special case: The refraction $\mathcal{R}[s; 1; E]$ ($\mathbf{r} = 1$) is a *reflection*, where each line l is reflected into a pair of lines r_1 and $r_2 = l$ symmetrical with respect to s. We will exclude this case sometimes without explicitly saying so in order to perform certain calculations. In general, however, reflection is a special case of refraction.

For $\mathbf{r} < 1$, a straight line r through E must intersect the refracting line s in a point $R(0, r_y)$ with $|r_y| < \mathbf{er}(1 - \mathbf{r}^2)^{-1/2}$ in order to produce real refracted rays, else there is no restriction.

Definition 2 *The* diacaustic d *of a pencil of rays* $E(1)$ *with respect to a refracting line* s *is the hull curve of all rays* r_2, r_2^*.

Let now $r_1 = ER$ be a straight line ($R(0, y = r_y) \in s$). We refract r_1 and get a pair of straight lines r_2, r_2^*. Let $X(x, 0)$ be the intersection of r_2 with the x-axis (Fig. 6). For α_1 and α_2 we then have

$$\sin \alpha_1 = \frac{y}{\sqrt{\mathbf{e}^2 + y^2}} \qquad \text{and} \qquad \sin \alpha_2 = \frac{y}{\sqrt{x^2 + y^2}}.$$

Together with (1), we get the following quadratic relation which describes a conic c:

$$c \ \ldots \ x^2 + y^2(1 - \mathbf{r}^2) = \mathbf{e}^2\mathbf{r}^2. \tag{7}$$

We call c the *characteristic conic* of $\mathcal{R}[s; \mathbf{r}; E]$, since we can find a refracted ray by orthogonally projecting a conic point on the coordinate axes and connecting these two points (Figures 6, 7). c is an ellipse if $\mathbf{r} < 1$ (Fig. 6), a pair of parallel lines $y = \pm\mathbf{e}$ if $\mathbf{r} = 1$, and a hyperbola

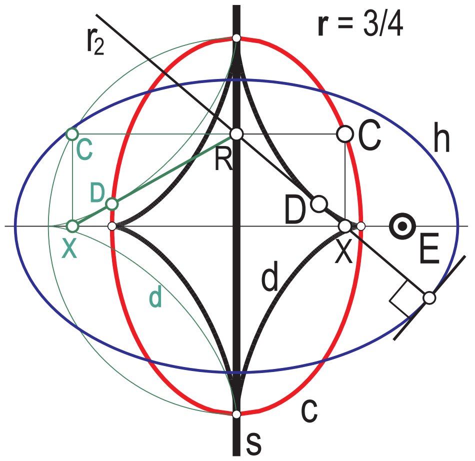

Figure 8: Diacaustic and involute ellipse

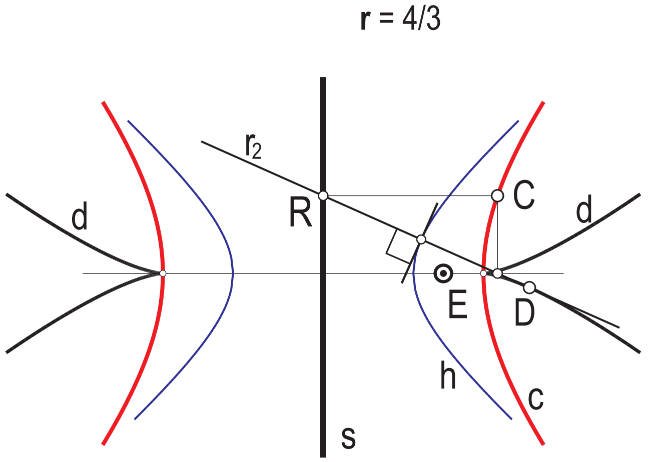

Figure 9: Diacaustic and involute hyperbola

if $\mathbf{r} > 1$ (Fig. 7). If $\mathbf{r} \neq 1$, the vertices of the conic lie on the coordinate axes and have coordinates $(\pm a, 0)^T$, $(0, \pm b)^T$ (b is imaginary for $\mathbf{r} > 1$):

$$a = \mathbf{er}, \quad b = \frac{\mathbf{er}}{\sqrt{1-\mathbf{r}^2}} \quad \Rightarrow \quad \frac{a}{b} = \sqrt{1-\mathbf{r}^2} \tag{8}$$

In case of $\mathbf{r} = 1$, all refracted (actually *reflected*) rays belong to one of the pencils with vertices E or $E^*(0, -\mathbf{e})$. Thus the diacaustic (actually *catacaustic*) degenerates into the two points E and E^*.

If c is an ellipse or a hyperbola, a simple consideration shows that the diacaustic d is the evolute of a conic h of the same type as c (Figures 8, 9). For the elliptic case we will give an elementary proof:

We apply an affine transformation to the characteristic conic c such that it appears as circle with radius b (Fig. 8). Then the line XR has constant length b and the affine hull curve is the result of an elliptic motion, i.e., an astroid. Thus, d is affine to an astroid and evolute of a conic [20]. The equation of this involute conic is

$$a^2x^2 + b^2y^2 = \frac{a^4b^4}{(a^2-b^2)^2} \qquad \text{or} \qquad \mathbf{r}^2(1-\mathbf{r}^2)x^2 + \mathbf{r}^2y^2 = \mathbf{e}^2(1-\mathbf{r}^2). \tag{9}$$

Obviously, E is focus of h (see also [15]).

The refraction $\mathcal{R}[s; \mathbf{r}; E]$ is fully described by the numbers a and b, since we then can calculate $\mathbf{r}$ and $\mathbf{e}$ from a and b:

$$\mathbf{r} = \sqrt{1 - \frac{a^2}{b^2}}, \quad \mathbf{e} = \frac{a}{\mathbf{r}}. \tag{10}$$

We can say:

Theorem 2 *Each refraction $\mathcal{R}[s; \mathbf{r}; E]$ is characterized by the conic (7). The diacaustic of the pencil $E(r_1)$ with respect to $\mathcal{R}[s; \mathbf{r}; E]$ is the evolute of a conic of the same type and has its four real cusps in the vertices of the characteristic conic.*

It is now time to reveal the secret of Fig. 5. The region where we can expect four real solutions is the *interior $\mathcal{I}(d)$ of the diacaustic* ($\mathcal{I}(d)$ can be defined as the set of all points through which four real tangents of d pass).[6] It is not difficult to show that the four roots of equation (5) are the y-coordinates of the intersection points of these tangents with the y-axis. Their geometrical meaning is quite remarkable, as three of them stem from the residual solutions of a merely physical problem. The regions of numerical instability in Fig. 5 are just the characteristic conics of the refraction.

Evolutes of conic sections are rational curves. Homogeneous rational parameter representations are for example

$$e_1 \;\ldots\; \vec{e}_1(t) = \begin{pmatrix} (1+t^2)^3 \\ a_1(1-t^2)^3 \\ 8b_1t^3 \end{pmatrix}, \quad e_2 \;\ldots\; \vec{e}_2(t) = \begin{pmatrix} (1-t^2)^3 \\ a_2(1+t^2)^3 \\ 8b_2t^3 \end{pmatrix}, \quad t \in \mathbb{R} \cup \{\infty\}. \tag{11}$$

In this formula e_1 is the evolute of an ellipse, e_2 the evolute of a hyperbola. a_1, b_1, a_2 and b_2 denote the (real) half-length of the axes of e_1 and e_2, respectively. The collineation

$$\kappa : \mathbb{R}^2 \to \mathbb{R}^2, \qquad X \mathrel{\widehat{=}} \vec{x} \mapsto \kappa(X) \mathrel{\widehat{=}} \mathbf{A}\vec{x},$$

where $\mathbf{A}$ is the matrix

$$\begin{pmatrix} 0 & b_1 & 0 \\ a_1b_1a_2 & 0 & 0 \\ 0 & 0 & a_1b_2 \end{pmatrix},$$

maps the point $\vec{e}_1(t)$ to the point $\vec{e}_2(t)$.[7] The evolute of an ellipse and a hyperbola are hence projectively equivalent. The same holds for the evolutes of two ellipses or two hyperbolas and can easily be verified. With respect to the refraction this means:

Theorem 3 *The diacaustics of all refractions on a straight line (with arbitrary ratio) are projectively equivalent.*

Theorem 3 is not difficult to prove, but not trivial, as the evolute of a curve is an object of *Euclidean* geometry. It has an important consequence for the real time calculation of refraction images. If we implement just one standard refraction (e.g., $\mathcal{R}[s; 4/5; E(0,1)] \Rightarrow a = 4/5$, $b = 4/3$) by creating tables, we can calculate *all* other refractions *in real time* by transforming the scenery using a simple collineation.

A well known parameter representation of d (see [1]) is

$$d \;\ldots\; \vec{d}(u) = \begin{pmatrix} aC(u) \\ bS(u) \end{pmatrix}, \qquad u \in I \tag{12}$$

where

$$\begin{array}{llll} C(u) = \cosh(u), & S(u) = \sinh(u), & I = \mathbb{R} & \text{if } c \text{ is an ellipse.} \\ C(u) = \cos(u), & S(u) = \sin(u), & I = [-\pi, \pi) & \text{if } c \text{ is a hyperbola.} \end{array}$$

Theorem 3 gives us now at once a second possible parameter representation of d:

$$d \;\ldots\; \vec{d}(u) = \frac{1}{C^3(u)} \begin{pmatrix} a \\ bS^3(u) \end{pmatrix}. \tag{13}$$

We will refer to this parameterization in the next section.

[6] The evolutes of ellipses and hyperbolas are algebraic curves of order 6 and class 4, i.e., we have at most four real tangents.

[7] Two of the cusps of e_1 are mapped to the points at infinity of e_2.

5. Reciprocal refractions

We are going to study a special pair of refractions now:

Definition 3 *Two refractions $\mathcal{R}[s;\mathbf{r};E] := \mathcal{R}$ and $\mathcal{R}[\widetilde{s};\widetilde{\mathbf{r}};\widetilde{E}] := \widetilde{\mathcal{R}}$ will be called* a pair of reciprocal refractions, *if s and $\widetilde{s}$ are parallel lines and $\mathbf{r}\widetilde{\mathbf{r}} = 1$.*

Reciprocal refractions deserve special interest, as they are quite common in everyday life: Rays of light passing trough a thick window are refracted reciprocally when they propagate from air to glass and from glass to air, respectively. It is well known (and immediately clear from the definition of refraction!) that a ray refracted reciprocally does not change its direction.

We will now compute a parameter representation of the diacaustic $\widetilde{d}$ of a pencil of lines $E(r_1)$ undergoing the reciprocal refraction $\widetilde{\mathcal{R}} \circ \mathcal{R}$. Basic considerations show that it has to be symmetric with respect to $\widetilde{s}$, if we take into account all possible refracted rays: One ray r through E corresponds to four rays $r_1 \ldots r_4$ after the two refractions.

We use the parameter representation (12) of d. The tangent $t(u)$ of d has then the equation

$$t(u) \ \ldots \ bS(u)x + \varepsilon aC(u)y = abS(u)C(u),$$

where $\varepsilon = 1$ if $\mathbf{r} < 1$ and $\varepsilon = -1$ otherwise. The intersection points P_1 and P_2 of t_1 and the axes of refraction s and $\widetilde{s} \ldots x = \widetilde{\xi}$, respectively, have coordinates

$$P_1(0, \varepsilon bS) \qquad \text{and} \qquad P_2\Big(\widetilde{\xi}, \frac{\varepsilon bS(aC - \widetilde{\xi})}{aC}\Big).$$

The tangent $\widetilde{t}(u) = \mathcal{R}(t(u))$ of d contains P_2 and is parallel to EP_1.[8] Its equation is

$$\widetilde{t}(u) \ \ldots \ abSCx + \varepsilon a\mathbf{e}Cy = ab(\mathbf{e} + \widetilde{\xi})SC - \mathbf{e}b\widetilde{\xi}S. \tag{14}$$

From (14) we can now deduce a parameter representation of the envelope $\widetilde{d}$ of the lines $\widetilde{t}(u)$:

$$\widetilde{d} \ \ldots \ \widetilde{\vec{d}}(u) = \frac{1}{aC^3}\begin{pmatrix} a(\widetilde{\xi} + \mathbf{e})C^3 + \varepsilon\mathbf{e}\widetilde{\xi} \\ b\widetilde{\xi}S^3 \end{pmatrix}. \tag{15}$$

Apart from a simple translation, this is just a parameter representation of the shape (13)!

The relevant part for practical purposes does not differ from the diacaustic of a simple refraction on a straight line. In this sense, reciprocal refraction is just *as simple as an ordinary refraction on a straight line.* In fact, we can even replace it by the refraction $\mathcal{R}[\widehat{s};\widehat{\mathbf{r}};\widehat{E}]$, that is determined by

$$\widehat{s} \ \ldots \ x = \widetilde{\xi} + \mathbf{e}, \qquad \widehat{\mathbf{r}} = \sqrt{\frac{|b^2 - \mathbf{e}^2|}{b^2}}, \qquad \widehat{\mathbf{e}} = \frac{\mathbf{e}b\widetilde{\xi}}{\sqrt{a^2(b^2 - \mathbf{e}^2)}}.$$

Taking into account all possible refracted rays, we get

Theorem 4 *10 The diacaustic $\widetilde{d}$ of a pair of reciprocal refractions consists of the evolutes of two congruent conic sections. For practical purposes, the reciprocal refraction is equivalent to the ordinary refraction determined by $\widehat{a} = -\varepsilon\widetilde{\xi}\mathbf{e}/a$, $\widehat{b} = \widetilde{\xi}b/a$ and axis of refraction $\widehat{s} \ \ldots \ x = \widetilde{\xi} + \mathbf{e}$. This refraction is always of the same type as the refraction belonging to the reciprocal refraction index $1/\mathbf{r}$ of the first refraction.*

[8] Here we omit the second refracted ray in order to make the calculation more lucid. In Theorem 10, we will summarize the result for all possible refracted rays.

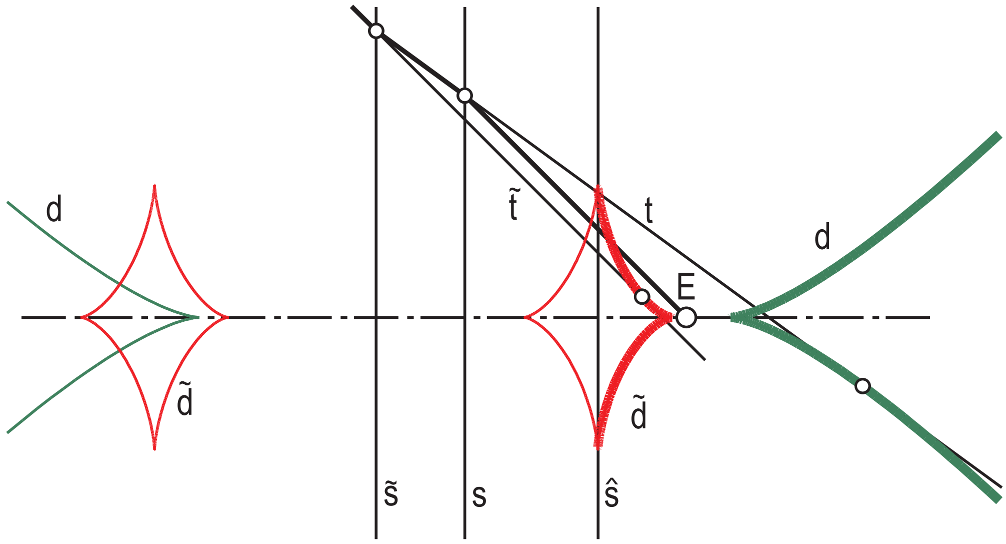

Figure 10: The diacaustic of reciprocal refraction: Relevant parts are drawn bold.

6. Refraction on a plane

The above considerations shall now be extended to Euclidean 3-space $\mathbb{E}^3$. We choose a refracting plane σ and an eye point $E \notin \sigma$. We can use a Cartesian coordinate system such that the x-axis is perpendicular to σ and E has the coordinate vector $(\mathbf{e}, 0, 0)^T$. the coordinate system is not uniquely determined and can still be scaled and rotated around the x-axis.

The refraction on σ can, of course, be reduced to the plane case. Being given a straight line r we take the plane ϱ through r that is perpendicular to σ and reflect r in ϱ on the line $\sigma \cap \varrho$.

Thus, r is again refracted into two straight lines r_1 and r_2 and Snell's law (1) holds if α_1 and α_2 denote the angles that r and the reflected rays form with the normal of σ.

Using the rotational symmetry of the system $\{E, \sigma\}$, we can immediately make use of the results of the previous chapters:

1. The calculation of the refrax of a point S needs the solution of an algebraic equation of order four and only one of the four possible solutions is relevant for practical purposes.
2. Applying the refraction $\mathcal{R}[\sigma; \mathbf{r}; E]$ to the bundle $E(r)$ yields a two parameter manifold $\mathcal{N}$ of rays that can be characterized in two ways:
   - $\mathcal{N}$ consists of all rays tangent to the *diacaustic surface* Δ that intersect the x-axis. Δ is of course the surface of revolution with the plane diacaustic d as meridian curve and axis of rotation x.
   - $\mathcal{N}$ is the *normal congruence of a surface* H *of order two*. H is an ellipsoid of revolution for $\mathbf{r} < 1$ and a hyperboloid of revolution for $\mathbf{r} > 1$.
3. The results of section 5 hold for the 3-dimensional case as well.

We will now take a closer look at the counter image Φ of a straight line, i.e. the set of all points $P \in \mathbb{E}^3$ with refraxes on a straight line $d \subset \sigma$. This investigation will be followed by a theorem on the order of the refrax of algebraic curves and a direct application in computer graphics.

It is no loss of generality to assume that

$$d \ \ldots \ \begin{cases} x = 0 \\ z = d_z \end{cases}$$

are the equations of d. It is clear that Φ is a ruled surface with double line d that has σ and the $[xz]$-plane as planes of symmetry. The x-axis is a double line of Φ as well for reasons of symmetry.

We will now assume that H is an ellipsoid of revolution.[9] It can be parameterized according to

$$\mathrm{H} \ \ldots \ \vec{h}(u,v) = \begin{pmatrix} A\cos(u) \\ B\sin(u)\sin(v) \\ B\sin(u)\cos(v) \end{pmatrix} \tag{16}$$

where A and B are the half length of the axis of the conic section (9). The normals $n(u,v)$ of H that intersect d are characterized by

$$\sin(u)\cos(v) = \frac{Bd_z}{B^2 - A^2}. \tag{17}$$

Substituting this in (16) we find that the corresponding points on H lie in the plane $z = B^2 d_z (B^2 - A^2)^{-1}$. Φ is therefore the *normal surface of a quadric surface along a planar section* c and thus an *algebraic surface of order four* (see [13]). In the line at infinity l_u of the $[xy]$-plane two generating lines of Φ coincide. Therefore each plane through l_u has a conic section in common with Φ (see Fig. 11 and Fig. 12). Summarizing all results we get

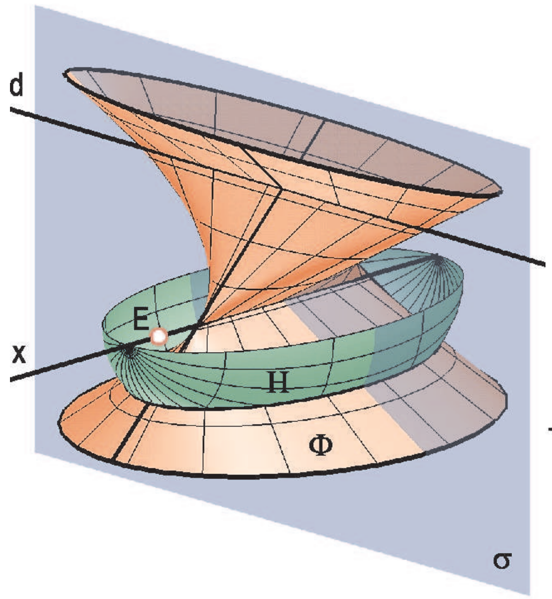

Figure 11: Ruled surface and ellipsoid

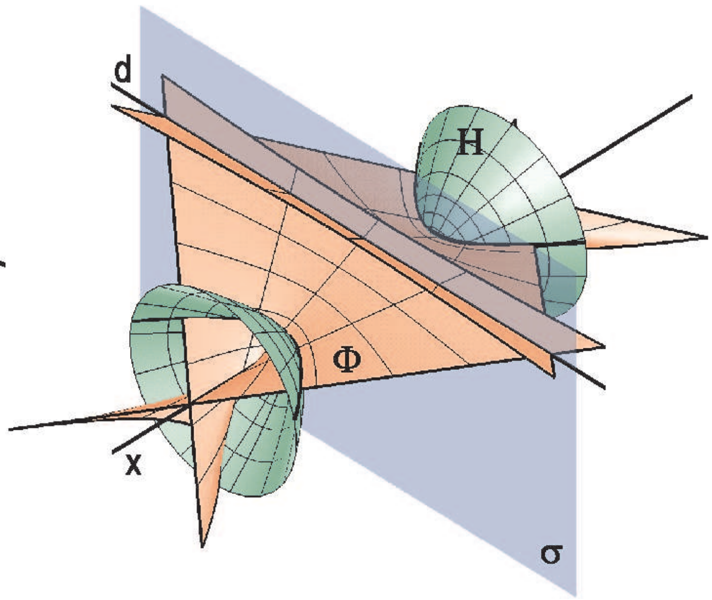

Figure 12: Ruled surface and hyperboloid

Theorem 5 *The counter image of a straight line d is a ruled surface Φ of order four. Φ has two double lines x and d two planes of symmetry σ and $[xz]$. On Φ we can find a one parameter set of conic sections in the planes parallel to x and d.*

[9] If H is a hyperboloid we will of course get analogous results; compare Fig. 12

It is now easy to prove a theorem of great theoretical interest:

Theorem 6 *The refrax of an algebraic curve k of order m is an algebraic curve k^r which in general is of order $4m$.*

Proof: We have to show that k^r and a generic straight line $l \subset \sigma$ have — in algebraic sense — $4m$ points of intersection. Each point of intersection corresponds to a point in $k \cap \Phi$ and there exist exactly $4m$ such points as Φ is of order four. □

The ruled surface Φ can be used to solve a problem of computer graphics as well. If you want to display the refrax of a filled polygon $\mathcal{P}$ you may run into troubles: Suppose that d and $\mathcal{P}$ lie in a common plane δ. d_1 and d_2 be the generators of Φ in δ. If $\mathcal{P}$ intersects one or both of these lines, its refrax $\mathcal{P}^r$ will have up to two overlappings.[10] This will cause problems with the filling algorithms (compare Fig. 13).

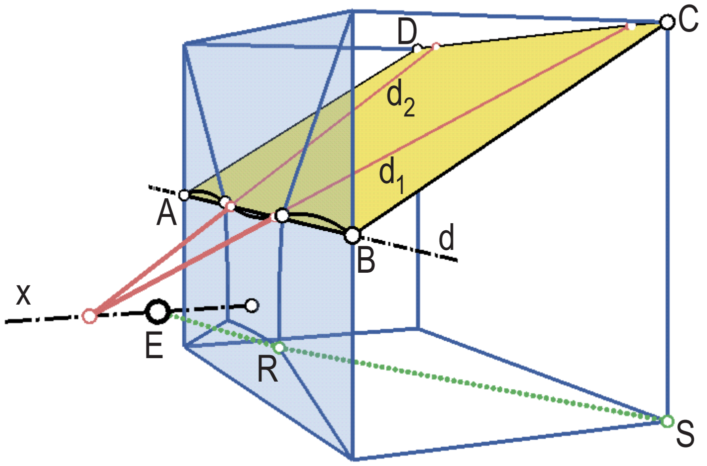

Figure 13: The refrax of the rectangle $ABCD$ has two overlappings as it is intersected by two refracted rays d_1, d_2.

To avoid this mess, you can clip $\mathcal{P}$ with the straight lines d_1 and d_2. It is easy to derive their equations from the following parameter representation of Φ

$$\frac{1-v}{A^2}\begin{pmatrix} \widehat{A}(A^2-B^2)\cos(\varphi) \\ 0 \\ 0 \end{pmatrix} + \frac{v}{B^2}\begin{pmatrix} 0 \\ \widehat{B}(B^2-A^2)\sin\varphi \\ B^2 d_z \end{pmatrix}, \qquad \varphi \in [-\pi,\pi], \quad v \in \mathbb{R}, \quad (18)$$

where $\widehat{A}$ and $\widehat{B}$ are the length of the major axes of the conic section c:

$$\widehat{A} = A\frac{\sqrt{(B^2-A^2)^2 - B^2 d_z^2}}{B^2-A^2} \qquad \text{and} \qquad \widehat{B} = B\frac{\sqrt{(B^2-A^2)^2 - B^2 d_z^2}}{B^2-A^2}.$$

Alternatively, one can use the algebraic $(2,2)$-correspondence between x and the straight line $d \subset \delta$. The point $X(\xi,0,0)^T \in x$ corresponds to the point $D(0,\eta,d_z)^T \in d$ if and only if ξ and η satisfy the relation

$$\xi^2 = \mathbf{e}^2\mathbf{r}^2 + (\eta^2 + d_z^2)(\mathbf{r}^2 - 1). \quad (19)$$

Note that (19) is valid for both cases $r < 1$ and $r > 1$.

[10] E.g., if you fix a little rectangle in an inclined position (see Fig. 13) in a box filled with water and watch it from an extreme point of view, you might be able to see both sides of the card.

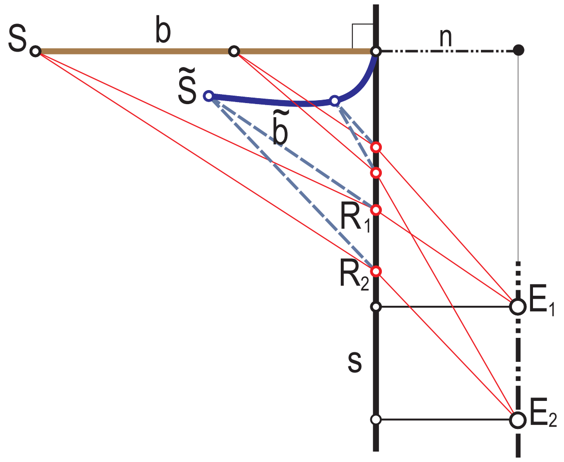

Figure 14: Reconstruction (two refraxes)

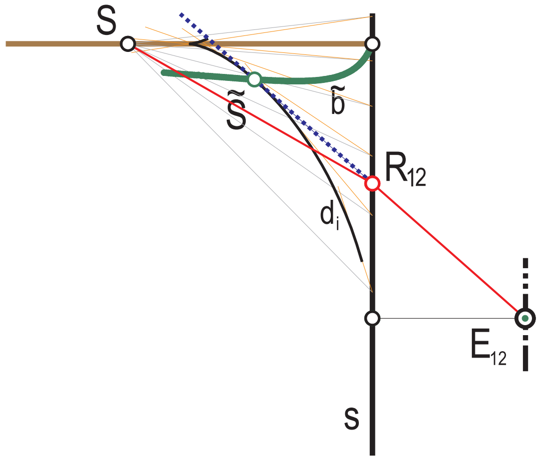

Figure 15: Passage to the limit $E_1 \to E_2$

7. Image lifting — the virtual object transformation

In this section we return to the refraction on a straight line. We want to investigate the second problem we mentioned at the beginning of Section 3 for the plane case: Given two eye points E_1 and E_2, and two refraxes R_1 and R_2 of a space point S on a straight line s. Can we say anything about the position of S?

The answer is, of course, yes (Fig. 14): The projection rays E_1R_1 and E_2R_2 intersect in a point $\widetilde{S}$. In this way, we can reconstruct geometrical primitives like straight lines b. Of course, the result can be rather complex. Even for $b = n \perp s$, e.g., the viewer sees a curved line $\widetilde{b}$ that does not look like a straight line, especially close to s (Fig. 14). In general, one can say: Objects appear both closer to the eye points and also closer to the refracting line.

The reconstruction depends on the distance of the eye points. The question is now: Whereto does $\widetilde{S}$ converge when we do a passage to the limit $E_1 \to E_2$. For this purpose, we consider a pencil of rays through S (Fig. 15). After being refracted on s inversely, they envelope a curve d_i ("inverse diacaustic"). Two neighboring tangents of d_i pass through E_1 and E_2. For $E_1 \to E_2$, these two tangents intersect in a point $\widetilde{S} \in d_i$.

This shows that the transformation of the plane $\mathbb{E}^2$ is independent of the passage to the limit $E_1 \to E_2$. Therefore we can give the following definition:

Definition 4 *The transformation $\mathbb{E}^2 \to \mathbb{E}^2$: $S \to \widetilde{S} = \mathcal{R}^2(s, \mathbf{r}, E)(S)$ denotes a plane transformation called* plane refractor, *where $\widetilde{S}$ is the tangent point of the inverse diacaustic d on the ray through E and the refrax of S.*

The refractor image of a point can be computed very efficiently with the help of formulas (5), (8) and (11). The presentation of a precise algorithm is planned for a later paper. One can argue that this transformation produces the impression of a refracted scene *plus* additional information about seeming distances. When you watch an underwater scenery you will notice extreme distortions, but still you always have the impression of being able to estimate distances. Of course these estimations are misleading and differ considerably from our daily life experience.

Our transformation is capable of explaining well-known optical effects:

Think of a person standing on a spring board above a swimming pool with constant depth (Fig. 16). The straight section line b of the bottom with a plane perpendicular to

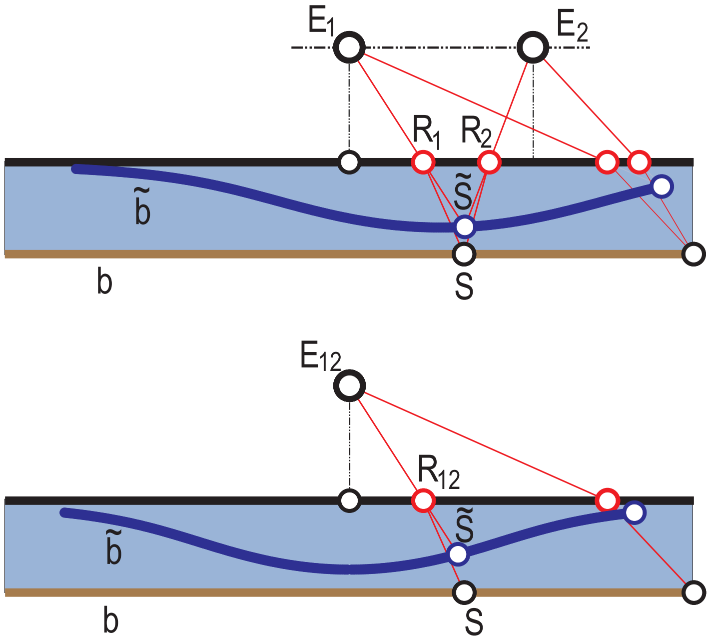

Figure 16: The bottom of a swimming pool with constant depth

the surface will have the 3D-image $\widetilde{b}$. In the upper images, the distance of the eye points $\overline{E_1E_2}$ is exaggerated, in the lower image it is "infinitesimally" small. The upper image can be interpreted as the sight through a diver's mask, where the "surface" is the glass of the mask. Therefore, one has always the impression to be above the "deepest region" of the pool when diving or snorqueling around in the pool.

8. Refraction on a circle

The refraction on a circle is much harder to deal with than the refraction on a straight line. Especially obtaining explicit formulas is a difficult task. We will therefore restrict ourselves to computing a parameter representation of the circle diacaustic.[11]

Let $c \ \ldots \ x^2+y^2 = R^2$ be a circle around O and $E(e > 0, 0)^T$ the eye point. The diacaustic d of $\mathcal{R}[c; \mathbf{r}; E]$ is then the evolute of a so called *Cartesian Oval* h [4, 20]. One way of defining a Cartesian Oval is the following:

Let F_1 and F_2 be two distinct (real or imaginary) points. Then a Cartesian Oval h is the set of all points X, satisfying

$$\tau\overline{XF_1} + \overline{XF_2} = \lambda; \qquad \tau, \lambda \in \mathbb{R}. \tag{20}$$

This definition is a generalization of the definition of a conic section through its focal property. But Cartesian Ovals are a generalization of conic sections in another respect as well:

The diacaustic of the pencil of lines $F_1(f_1)$ with respect to the refracting curve h and a suitable index of refraction is just F_2.[12]

[11] The following considerations are all due to [20]. There, the author gave all necessary details to compute a parameter representation of the diacaustic but not the parameter representation itself.

[12] That is why F_1 and F_2 are called *focal points* of h. In addition, F_1 and F_2 satisfy the Plücker-definition of a focal point as well: they are intersection points of isotropic tangents of h. Considering the facts that h is an algebraic curve of order four and that it has cusps at the circular points at infinity, we can even say that there exist nine focal points, three of them on the x-axis. Not all of them are real, of course.

In case of a refraction we have to take

$$F_1 = E^*(R^2/\mathbf{e}, 0)^T, \quad F_2 = E, \quad \tau = -\frac{\mathbf{e}}{R}, \quad \lambda = \frac{R^2 - \mathbf{e}^2}{\mathbf{r}R}. \tag{21}$$

The explicit equation of h can now be derived from (20):

$$\begin{aligned} h \quad \ldots \quad & \left(\mathbf{r}^4 - \mathbf{r}^2[(\mathbf{e}^2 - R^2)(x^2 + y^2) - \mathbf{e}^2(R^2 - 2) + R^4] + \mathbf{e}^4\right)^2 = \\ = \quad & 4R^2\mathbf{r}^2(\mathbf{r}^2 - \mathbf{e}^2)^2[(x - \mathbf{e})^2 + y^2]. \end{aligned} \tag{22}$$

If we translate the coordinate system by the vector $\overrightarrow{OE^*}$ (the new origin is then E^*) we can give a parameterization of h in polar coordinates (ϱ, θ). Using the abbreviations

$$\begin{aligned} &\alpha = \overline{EE^*} = \frac{\mathbf{e}^2 - R^2}{\mathbf{e}}, \qquad \lambda = \frac{R^2 - \mathbf{e}^2}{rR}, \qquad \tau = -\frac{\mathbf{e}}{R}, \\ &\Delta = \Delta(\theta) = \lambda^2 + \alpha^2\tau^2 - 2\alpha\lambda\tau\cos\theta - \alpha^2\sin^2\theta, \end{aligned}$$

we have

$$(\tau^2 - 1)\varrho(\theta) = \lambda\tau - \alpha\cos\theta \pm \sqrt{\Delta}. \tag{23}$$

In order to compute the evolute d of h we need the derivatives of first and second order of Δ and ϱ:

$$\begin{aligned} \Delta' &= 2\alpha\lambda\tau\sin\theta - 2\alpha_1^2\sin\theta\cos\theta, \\ \Delta'' &= 2\alpha\lambda\tau\cos\theta - 2\alpha^2\cos 2\theta, \\ (\tau^2 - 1)\varrho' &= \alpha\sin\theta \pm \frac{\Delta'}{2\sqrt{\Delta}}, \\ (\tau^2 - 1)\varrho'' &= \alpha\cos\theta \pm \frac{2\Delta\Delta''\Delta'^2}{4\Delta^{3/2}}. \end{aligned} \tag{24}$$

By substituting (24) in the well known formulas (see [1])

$$\begin{aligned} x &= \varrho\cos\theta - \frac{(\varrho^2 + \varrho'^2)(\varrho\cos\theta + \varrho'\sin\theta)}{\varrho^2 + 2\varrho'^2 - \varrho\varrho''}, \\ y &= \varrho\sin\theta - \frac{(\varrho^2 + \varrho'^2)(\varrho\sin\theta - \varrho'\cos\theta)}{\varrho^2 + 2\varrho'^2 - \varrho\varrho''}, \end{aligned}$$

we finally get a parameter representation of the diacaustic d of $\mathcal{R}[c; \mathbf{r}; E]$ in terms of $\mathbf{e}$, R and r only.[13] This parameter representation was used to draw Fig. 17.

9. Future work — a 3D-refractor map

In section 7 we presented a method of reconstructing a point S from two refraxes. In the 3D-case the analogous reconstruction of a space point fails because the projection rays through E_1 and E_2 in general do not intersect. One way of overcoming this problem is to assume the midpoint M of the common normal of the projection rays as the seeming position of S. But then another problems occurs: If E_2 converges to E_1, M does not have a well defined limiting point. I.e., the limiting point heavily depends on the limiting process $E_2 \to E_1$.

[13] Do not forget to apply the translation $x \mapsto x - R^2e^{-1}$ if you want to use the standard coordinate system with center O!

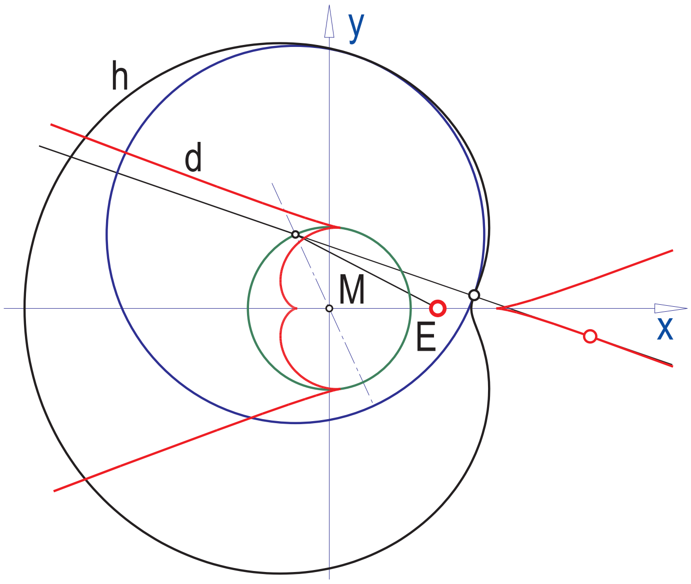

Figure 17: The diacaustic d of a circle and its involute h (Cartesian oval). Only the larger ring of h is relevant for practical purposes.

We are currently preparing a paper on this topic where we will propose a method that is capable of dealing with these difficulties. The results of the previous sections will be essential in this investigation.

The 3D-case has applications in underwater photography, underwater archaeology and computer graphics. For this reason parameter representations of the curves in Fig. 14, Fig. 15 and Fig. 16 are of interest.

References

[1] I.N. Bronstein, K.A. Semendjajew: *Taschenbuch der Mathematik.* B.G. Teubner Verlagsgesellschaft, Stuttgart, Leipzig 1991.

[2] J.W. Bruce, P.J. Giblin, C.G. Gibson: *On Caustics of Plane Curves.* Amer. Math. Monthly **88**, 651–657 (1981).

[3] A. Calyey: *A memoir upon caustics.* Philos. Trans. Roy. Soc. London **147**, 273–312 (1857).

[4] H. Ebisui: *An extension to fourth order surfaces by the oval with 3 inversion points.* Proceedings 8th ICECGDG, Austin 1998, vol. 2, 428–432.

[5] R. Freudenthal, E. Heinrich: *Kaustiken in vektorieller Behandlung.* Math. Naturw. Unterricht **XII**, 7, 307–317 (1959/60).

[6] R. Freudenthal, E. Heinrich: *Grundkurven zu vorgegebenen Kaustiken.* Math. Naturw. Unterricht **XIV**, 2, 70–76 (1961/62).

[7] J.D. Gergonne Ann. Math. **15**, 25 (1824).

[8] G. Glaeser: *Reflections on Spheres and Cylinders of Revolution.* J. Geometry Graphics **3**, no. 2, 121–139 (1999).

[9] H. Holditch: *On the n-th Caustic, by Reflexion from a Circle.* Quart. J. Math. **2**, 301–322 (1858).

[10] H. Horninger: *Zur geometrischen Theorie der Spiegelung an krummen Oberflächen.* Sitzungsber., Abt. II, österr. Akad. Wiss., Math.-Naturw. Kl. **145**, 367–385 (1936).

[11] E. Laguerre: *Sur les anticaustiques par réflexion de la parabole, les rayons incidants étant paralléles.* Nouv. Ann. math., 1883, 27.

[12] E. Müller, J.L. Krames: *Vorlesungen über Darstellende Geometrie, Bd. II: Die Zyklographie.* Franz Deuticke, Leipzig, Wien 1929.

[13] E. Müller, J.L. Krames: *Vorlesungen über Darstellende Geometrie, Bd. III: Konstruktive Behandlung der Regelflächen.* Franz Deuticke, Leipzig, Wien 1931.

[14] L. A. Quételet: Mém. Ac. Belg. **5** (1829).

[15] G. Salmon, W. Fiedler: *Analytische Geometrie der höheren ebenen Kurven.* Teubner, Leipzig 1882.

[16] J. Schwarze: *Cubic and Quartic Roots.* In A. Glassner (ed.): Graphics Gems, Academic Press, San Diego 1990, 404–407.

[17] E. Stübler: *Über Brennlinien durch Reflexion.* Zeitschr. math. Unterr. **39**, 121–144 (1908).

[18] J.L. Synge: *Geometrical Optics.* Cambridge University Press, 1937.

[19] K. Weiss: *Unterwasser-Fotografie.* Bussesche Verlagshandlung, Herford 1979.

[20] H. Wieleitner: *Spezielle ebene Kurven.* G.J. Göschen'sche Verlagshandlung, Leipzig 1908.

Received March 9, 2000; final form June 28, 2000