Journal for Geometry and Graphics
Volume VOL (YEAR), No. NO, 1–18.

Reflections on Refractions

Georg Glaeser, Hans-Peter Schröcker(Version September 5, 2001)

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 formulas and parametric equations for the refraction on planes and spheres. The calculation of the "refrax" of a space point leads to algebraic equations. The counter image of straight lines with respect to refraction on a plane are specific algebraic surfaces of degree four.
Furthermore, we develop a general theory of refractions, with reflections as a special case. The manifold of all refracted rays is the normal congruence of certain algebraic surfaces. These surfaces and their diacaustic surfaces are investigated.

Key Words: Refraction, reflection, curved perpectives, fish-eye perspectives, diacaustic, catacaustic, normal congruence, real-time rendering, underwater photography.

1. Introduction and state of the art

Refractions are to be seen everywhere in daily life (+ example). They are also used for many technical purposes, most of all for eye glasses, optical lenses, etc.
Despite of their importance, it seems that – apart from rather basic considerations – not much investigation has been done on the theoretical part of refractions. We will now briefly describe the results that are known so far.
The mathamatical interest in refraction and reflection began more than 300 years ago, when Tschirnhaus and Huygens worked on this topic. They and – little later – Johann Bernoulli were especially interested in *caustics*, 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 reflection and refraction phenomena. The catacaustic of a pencil $E(e)$ of rays with respect to a circle c, e.g., is an algebraic curve of class four. Thus, a circle (or a sphere) has theoretically four specular points.[2]

[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]).

ISSN 1433-8157/$ 2.50

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 were 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 [13], 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.[3]

Since the days of TSCHIRNHAUS and HUYGENS the caustics for many special cases were described by a number of authors ([4], [11], [14]).

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.[4] 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.*[5]

In [5], the caustics of a pencil of lines $E(e)$ with respect to a plane curve $c \ldots \vec{x} = \vec{x}(t)$ were calculated in a general form. The same authors solved the problem of finding the *"anticaustic"*, e.g., the curve producing a certain given caustic in [6]. The given formulas, however, and the according differential equations are quite complicate.

In [2], a new way of constructing the catacaustic of a pencil of lines $E(e)$ was introduced: We regard the set of all conic sections osculating the reflecting curve c and having E as one focal point. Then the catacaustic is the locus of all focal points $F \neq E$ of all conics. During the eighties the same authors (and some others as well) studied catacaustics from the completely different point of view of singularity theory ([???],[???]).

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.

[MEHR LITERATURZITATE; ABBILDUNG (KONSTRUKTION EINER DIAKAUSTIK. ZWEI KONSTRUKTIONEN EINER KATAKAUSTIK]

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

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.)

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. 1, $\mathbf{r} \approx 1.33$ was chosen for the ratio of the light speed in the atmosphere (to the right) and water (to the left).

[3] In case of a reflection h is the orthonomic of the reflecting curve with respect to E.

[4] The problem of the n-th caustic of reflection on a circle, e.g., was solved in [9].

[5] For a proof of this theorem, see [17]. Of course, an analogous theorem holds for the plane case.

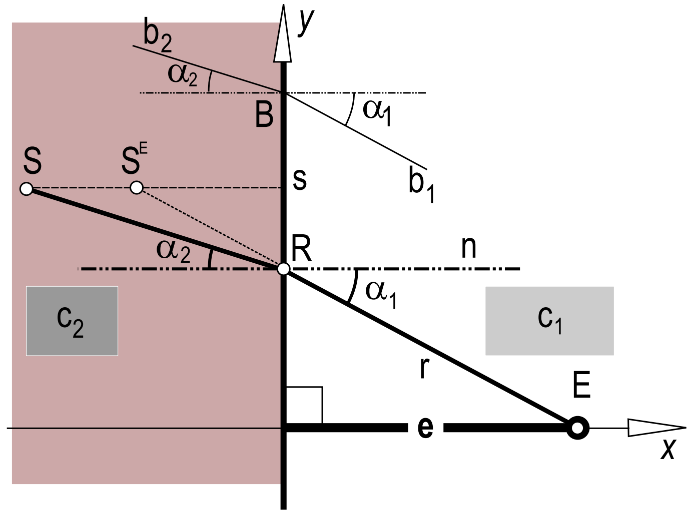

Figure 1: Refraction on a straight line

With $\mathcal{R}[s;\mathbf{r};E]$, we connect a Cartesian coordinate system as follows (Fig. 1): 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 $\alpha_1,\ \alpha_2\ \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, we have a restriction on either α_1 or α_2, depending on the ratio $\mathbf{r}$: For $\mathbf{r} > 1$, we have $|\alpha_2| \le \arcsin\frac{1}{\mathbf{r}}$, else we have $|\alpha_1| \le \arcsin\mathbf{r}$. E.g., for $\mathbf{r} \approx 0.75$ (water→air) we have $|\alpha_1| \le \alpha_1^{max} = 48.5°$. For small angles α_1, the percentage rate of reflected light is also small. When α_1 approaches the limiting angle, this percentage rate increases, until, for $|\alpha_1| \ge \alpha_1^{max}$, we have total reflection on s.

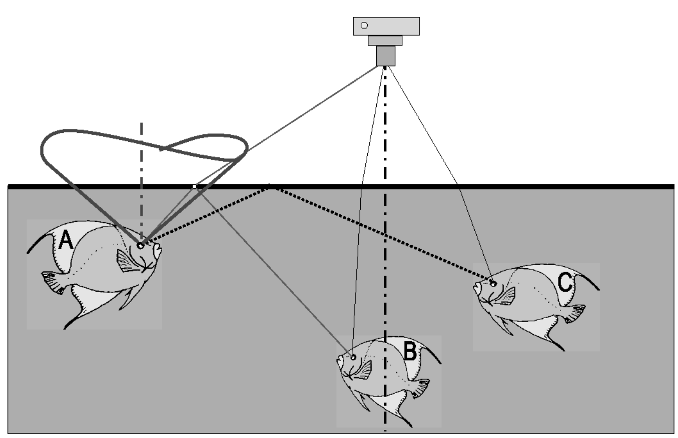

Figure 2: Who can see whom?

As a consequence, fish A in a calm pool (Fig. 2) will see

- "everyhing" outside the pool, though partly very distorted. The refracted image fills a circle c on the surface that stems from a cone a 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.

When a person outside the pool takes a picture of the pool (e.g., from the spring board), the image will show all the fishes.

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 accont such a point, since we usually oberserve 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 fishes surface, we are looking for the projection ray r through our eye that runs through S after being reflected 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 fishes: We call the photographic images created by refracting optical ultra-wideangle lenses "fisheye-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 on a refraction on a surface Φ $\big(R = \mathcal{R}[\Phi; \mathbf{r}; E](S)\big)$ 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 lenght") has a stationary value with respect to infinitesimal variations of the path (see, e.g., [17]):

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

We introduce the variable $y = r_y$. Then the "total-time function" or "optical length function" (Fig. 3)

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

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 proof

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 Polynom* (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 Equation (5) then fulfill $t'(y) = u'(y) + v'(y) = 0$, or $\tilde{t}'(y) = u'(y) - v'(y) = 0$, respectively. Only those solutions that fulfill $t'(y) = 0$ (Equation (4)) are practical solutions. We will now show that exactly only one of the roots y_i fulfills (4), wheras the residual ones fulfill $\tilde{t}' = 0$ (Fig. 3).
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 $\tilde{t}(y)$ are outside this interval. This is exactly the contents of the following Lemma 1 with $x_0 = 0$, $x_1 = s_y$. □

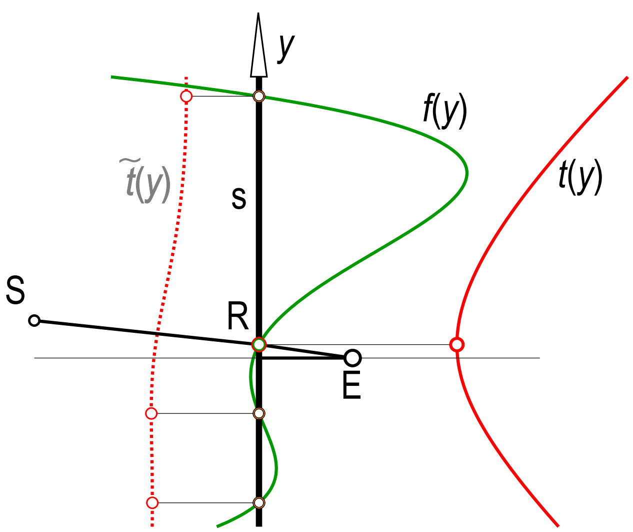

Figure 3: The four roots of $f(y)$

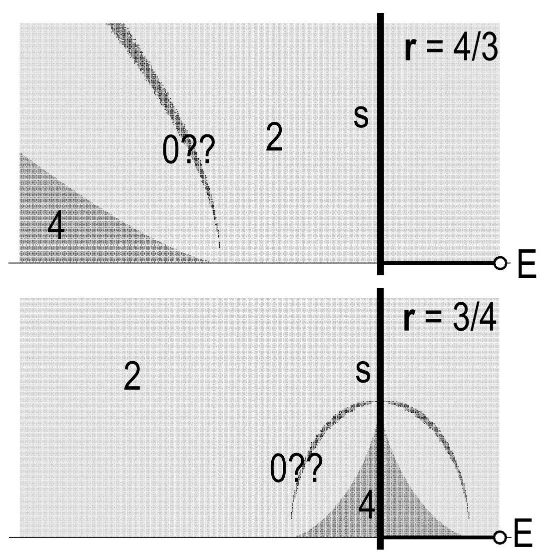

Figure 4: Number of roots

Lemma 1. Let $I \subset \mathbb{R}$ be a real intervall and u, v two strictly convex functions in $\mathcal{C}^1[I, \mathbb{R}]$. 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 $\tilde{t} := u - v$ 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 monotonious 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 $\tilde{t}$. 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. 4 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 Equation (4) when we declare an ε that is too small for $|t'(y)| < \varepsilon$ (in Fig. 4, $\varepsilon = 10^{-11}$ was chosen; with $\varepsilon = 10^{-6}$, the verification was always OK). We will explain this behaviour in Section 4; it is closely connected with the three residual roots of equation (5).
Anyway, the fast criterion $0 \leq y \leq s_y$ (or $s_y \leq y \leq 0$, 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 Equation (5). The up to four real solutions of the polynom (5) can be calculated by means of well known formulas ([15]).[6]
In the next section, we will take a closer look at the diacaustic of $\mathcal{R}[s; \mathbf{r}; E]$.

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

Actually, Equation (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 $E(r_1)$ 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. 5). For α_1 and α_2 we then have

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

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

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{c_4x_n^4 + c_3x_n^3 + c_2x_n^2 + c_1x_n + c_0}{4c_4x_n^3 + 3c_3x_n^2 + 2c_2x_n + c_1}. \tag{6}$$

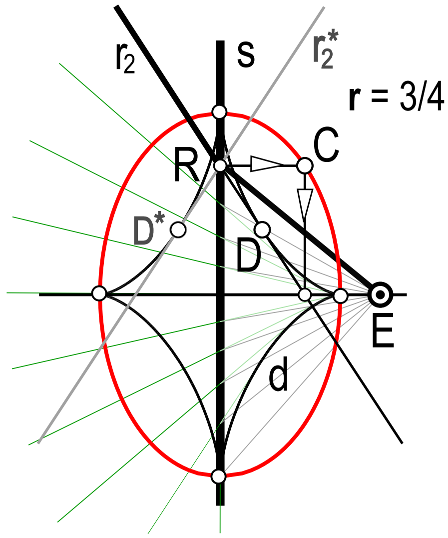

Figure 5: The characteristic ellipse

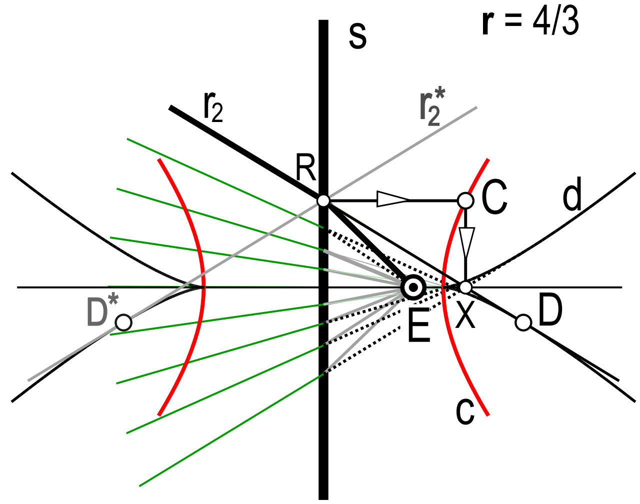

Figure 6: The characteristic hyperbola

Together with Equation (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 5, 6). c is an ellipse if $\mathbf{r} < 1$ (Fig. 5), a pair of parallel lines $y = \pm\mathbf{e}$ if $\mathbf{r} = 1$, and a hyperbola if $\mathbf{r} > 1$ (Fig. 6). 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^*.

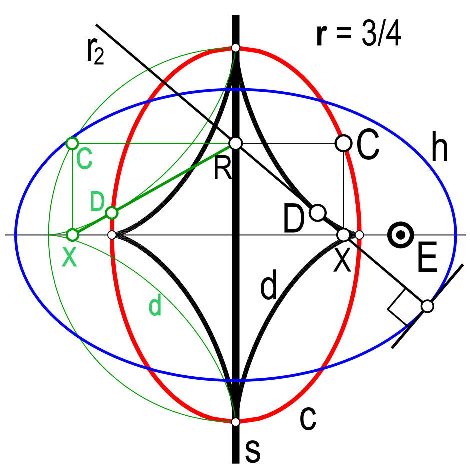

Figure 7: Diacaustic and involute ellipse

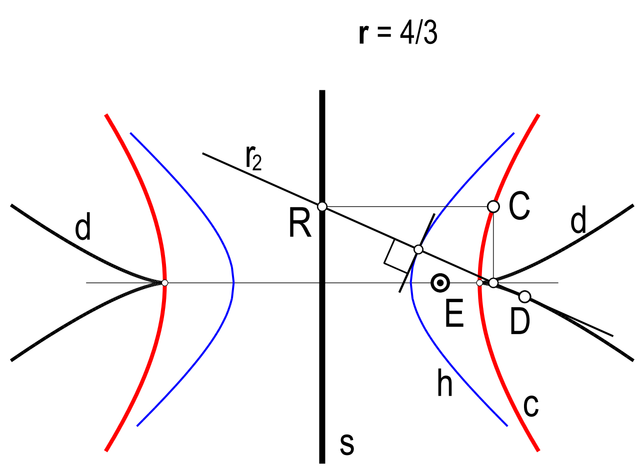

Figure 8: Diacaustic and involute hyperbola

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 7, 8). 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. 7). 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 ([19]). 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 [14]).

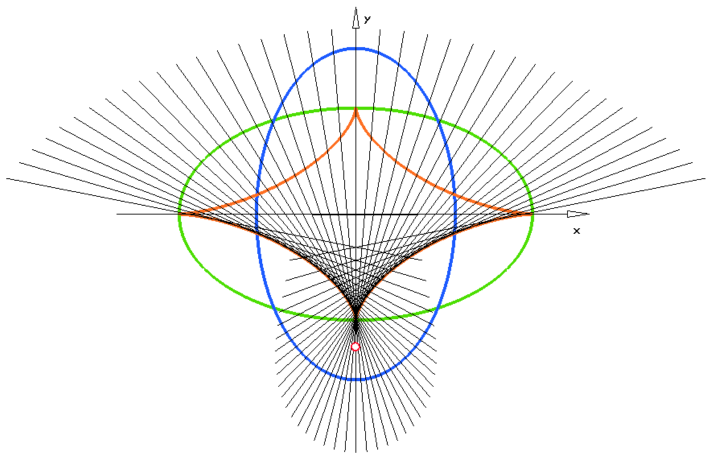

Figure 9: diacaustic and conics

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. 4. 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.[7] 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. 4 are just the characteristic conics of the refraction.

[7] The evolutes of conic sections are algebraic curves of order 6 and class 4, i.e., we have most four real tangents.

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

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

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

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

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 $E_1(t) \mathrel{\hat{=}} \vec{e}_1(t)$ to the point $E_2(t) \mathrel{\hat{=}} \vec{e}_2(t)$. Therefore, the evolute of an ellipse and an hyperbola are 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 proof, but quite remarkable, 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., $\mathcal{R}[s; \frac{4}{5}; E(0,1)] \Rightarrow a = \frac{4}{5},\ b = \frac{4}{3}$) by creating tables, we can calculate *all* other refractions *in real time* by transforming the scenery using a simple collination.
A well known parameter representation of d (see [1]) is

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

where

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

Theorem 3 gives us 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 (13) in the next chapter.

5. Reciprocal refractions

We are going to study a pair of refractions now: $\mathcal{R}[s;\mathbf{r};E] := \mathcal{R}$ and $\mathcal{R}[\tilde{s};\tilde{\mathbf{r}};\tilde{E}] := \tilde{\mathcal{R}}$, where s and $\tilde{s}$ are parallel lines and $\mathbf{r}\tilde{\mathbf{r}} = 1$. $\mathcal{R}$ and $\tilde{\mathcal{R}}$ will be called *a pair of reciprocal refractions* in the following (see Fig. 10).

Reciprocal refractions deserve special interest, as they are quite common in everyday life: Rays of Light passing trough a thick window are refracted recoprocally when they propagate from air to glass and from glass to air, respectively.

It is well known (and immediatly clear from the definition of refraction!) that a ray refracted reciprocally does not change its direction (see ???).

We will now compute a parameter representation of the diacaustic $\tilde{d}$ of $\tilde{\mathcal{R}}\circ\mathcal{R}\big(E(r_1)\big)$. A basic considerations shows, that it has to be symmetric with respect to 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.

The tangent $t(u)$ of d has the equation

$$t(u)\ldots bSx + aCy = abSC.$$

The intersection points P_1 and P_2 of t_1 and the axes of refraction s_1 and $s_2\ldots x = \tilde{\xi}$, respectively, have the coordinates

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

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

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

From (14) we can now easily deduce a parameter representation of $\tilde{d}$:

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

Apart from the translation $x \mapsto x \mp e - \tilde{\xi}$, 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 stright line.[9] 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 $\mathcal{R}[\hat{s};\hat{\mathbf{r}};\hat{E}]$, that is determined by

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

Taking into account all possible refracted rays, we get

[8] Here we omit the second refracted ray in order to make the calculation more lucid. In Theorem 4, we will summarize the result for all possible refracted rays.

[9] This relevant part belongs to the upper line in (15). In case of $r < 1$, we have $u \in [-\pi/2, \pi/2]$, in case of $r > 1$ we have $u \in (-\infty, \infty)$.

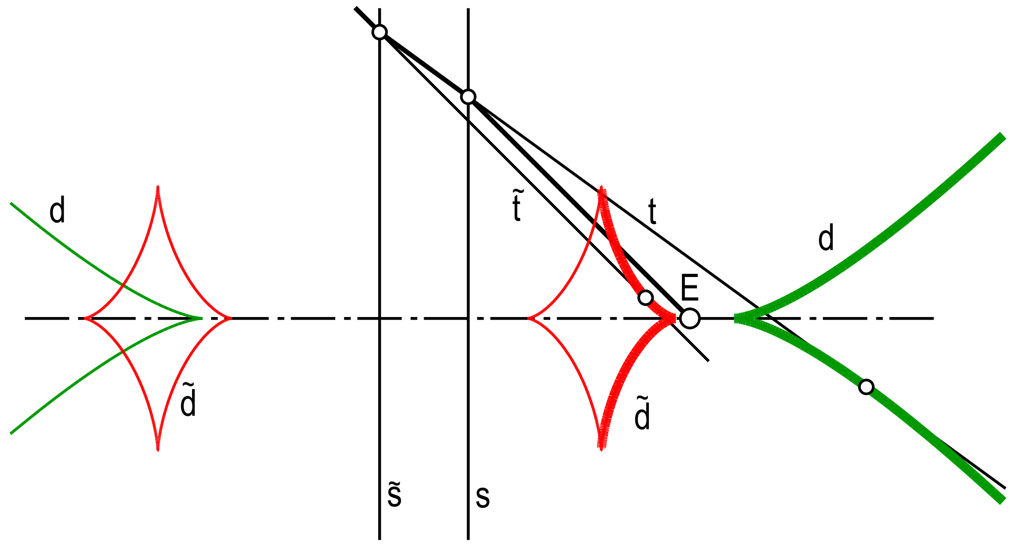

Figure 10: The diacaustic of reciprocal refraction: Relevant parts are drawn bold.

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

6. Refraction on a plane

The above considerations shall now be extended into Euclidean 3-space $\mathbb{E}^3$. We choose a plane σ (the *refracting plane*) and a point $E \notin \sigma$ (the *eye point*). 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 (calculation of refraxes, diacaustic surfaces, reciprocal refractions etc.). Instead of doing this in detail, we will proof a theorem on the refraction image of an arbitrary algebraic curve $c \subset \mathbb{E}^3$.[10] At first we will proof the following interesting lemma:

Lemma 2. The points being refracted onto a given straight line $d \subset \sigma$ are exactly the points on some algebraic ruled surface Φ of order four. Φ has double lines d and x (the normal of σ through E) and is symmetric with respect to the planes σ and φ (the plane perpendicular to d trough x).

[10]The *refracton image* of c is defined as the set c^r of all refraxes of points of c.

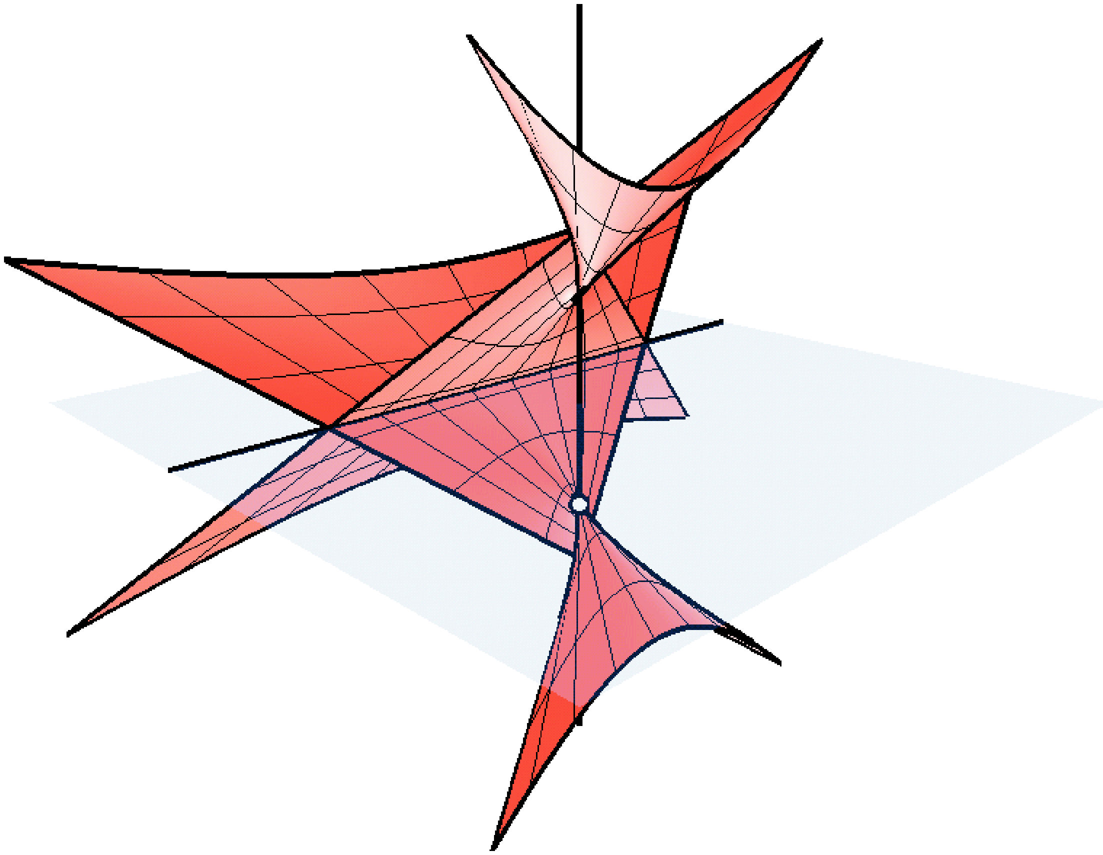

Figure 11: Ruled surfaces; $\mathbf{r} = 0.8$

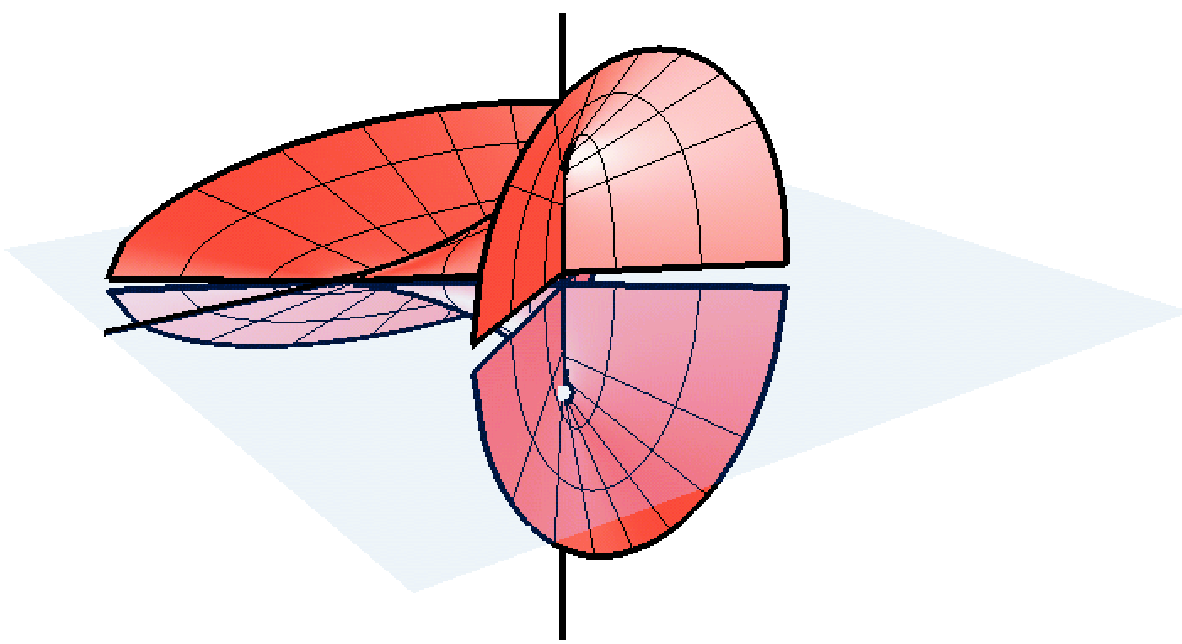

Figure 12: Ruled surfaces; $\mathbf{r} = 1.2$

Proof. It is clear that the points being refracted to d lie on a ruled surface Φ with double line d that has σ and φ as planes of symmetry. As each generating line of Φ intersects x and Φ

cannot be a torse, the planes of the pencil $x(\tau)$ are all tangent to Φ. A fixed plane $\tau_0 \in t(\tau)$ and Φ have the two reflexion rays t_0, t_1 and the axis x in common. Thus the points $T_0 := t_0 \cap t$ and $T_1 := t_1 \cap t$ are the two points of tangency in τ_0. t is met by two other generating lines g_0 and g_1 in T_0 and T_1, respectively, and therefore has to be a double line.
Let us now take a closer look at the $(2,2)$-correspondency between x and d induced by the generating lines of Φ. By rotating Σ around x we can arrange that d intersects the z-axis of our coordinate system orthogonally in the point $(0,0,b)^T$ (see Fig. **??**).
By $X(u)$ and $D(v)$ we denote the points $(u,0,0)^T$ and $(v,0,b)^T$ respectively ($u, v \in \mathbb{R}$). If a straight line through $D(v)$ is refracted to a line trough $X(u)$ the corresponding angles $\alpha_1(v)$ and $\alpha_2(v)$ must fulfill (1). Furthermore it is easy to see, that the equation

$$\tan^2 \alpha_1(v) = \frac{v^2 + b^2}{\mathbf{e}^2}$$

holds. By applying (1) we find

$$\sin^2 \alpha_2(v) = \frac{v^2 + b^2}{\mathbf{r}^2(v^2 + b^2 + \mathbf{e}^2)}, \quad \cos^2 \alpha_2(v) = \frac{\mathbf{e}^2\mathbf{r}^2 + (v^2 + b^2)(1 - \mathbf{r}^2)}{\mathbf{r}^2(v^2 + b^2 + \mathbf{e}^2)}$$

and can therefore calculate the parameter u of the corresponding point $Z(u)$ on z:

$$u^2 = (v^2 + b^2)\cot^2 \alpha_2(v) = \mathbf{e}^2\mathbf{r}^2 + (v^2 + b^2)(1 - \mathbf{r}^2). \tag{16}$$

(16) is the equation of an *algebraic* (2,2)-correspondency, showing that the ruled surface Φ is an *algebraic surface of order four* (see [12]). □

Now we ready to proof the theorem we mentioned above:

Theorem 5: *The refraction image of an algebraic curve c of order m is an algebraic curve c^r which in general is of order $4m$.*

Proof. We have to show, that c^r and an arbitrary straight line $l \subset \varepsilon$ have – in algebraic sense – $4m$ points of intersection. Each point of intersection corresponds to a point in $c \cap \Phi$, where Φ is the ruled surface that is refracted to l. There exist exactly $4m$ such points as Φ is, according to lemma 2, an algebraic surface of order four. □

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 chapter 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. 13): The projection rays E_1R_1 and E_2R_2 intersect in a point $\tilde{S}$. 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 that does not look like a straight line, especially close to s (Fig. 13). 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: Wherto does $\tilde{S}$ converge when we do a passage to the limit $E_1 \to E_2$. For this purpose, we consider

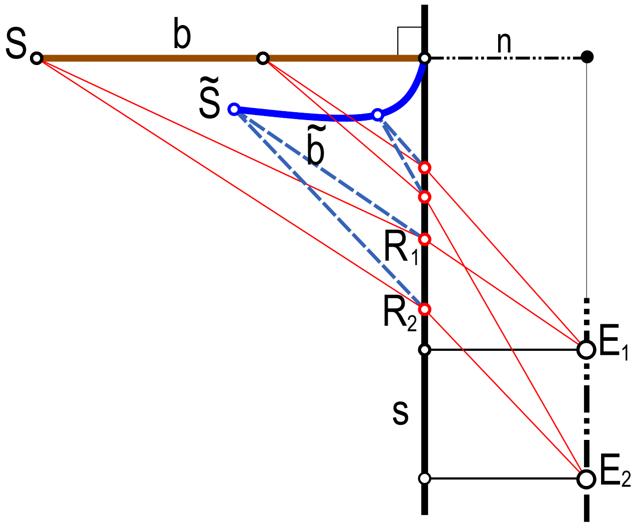

Figure 13: Reconstruction (two refraxes)

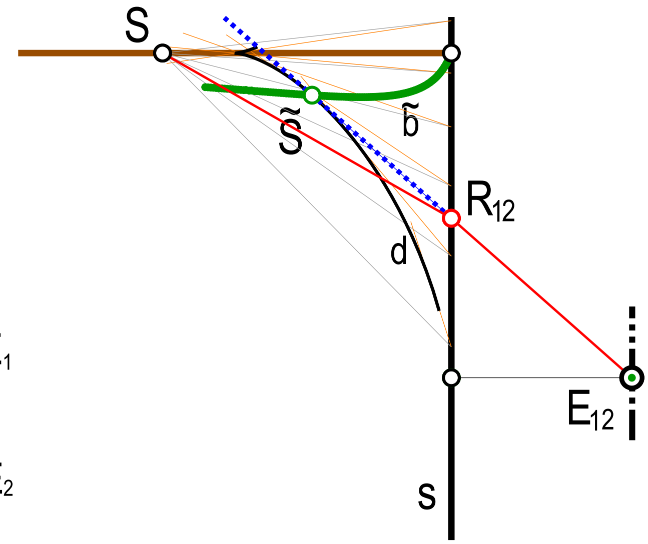

Figure 14: Passage to the limit $E_1 \to E_2$

a pencil of rays through S (Fig. 14). 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 $\tilde{S} \in d_i$.
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 3. The transformation $\mathbb{E}^2 \to \mathbb{E}^2$: $S \to \tilde{S} = \mathcal{R}^2(s, \mathbf{r}, E)(S)$ denotes a plane transformation called *plane refractor*, where $\tilde{S}$ is the tangent point of the inverse diacaustic d on the ray through E and the refrax of S.

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. 15). The straight section line b of the bottom with a plane perpendicular to the surface will have the 3D-image $\tilde{b}$. In the upper images, the distance of the eye points $\overline{E_1E_2}$ is exagerated[11], in the lower image it is "infinitesimally" small.
To give another example, think of a person standing on a spring board above a swimming pool with constant depth (Fig. 15). The straight section line b of the bottom with a plane perpendicular to the surface will have the 3D-image $\tilde{b}$. In the upper images, the distance of the eye points $\overline{E_1E_2}$ is exagerated, 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. Therfore, one has always the impression to be above the "deepest region" of the pool when diving or snorqueling around in the pool.

[11]The upper image can be interpreted as the sight through a diver's mask, where the "surface" is the glass of the mask. Therfore, one has always the impression to be above the "deepest region" of the pool when diving or snorqueling around in the pool. This effect is called "image lifting" ("Bildhebung") in [18] (p. 25), where $\tilde{S}$ is chosen exactly above S, which is only approximately true for smaller angles α_1, and not at all true for points close to σ (compare Fig. 13).

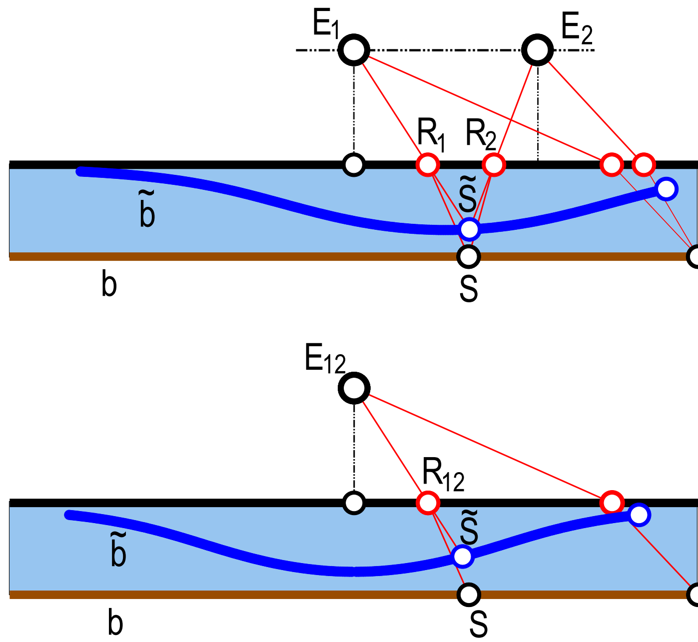

Figure 15: The bottom of a swimming pool with constant depth

8. Refraction on a circle

The refraction at a circle is much harder to deal with than the refraction at a straight line. Especially obtaining explicit formulas is a difficult task. We will therfore restrict ourselves to computing a parameter representation of the circle diacaustic.[12]

Let $c \dots x^2 + y^2 = R^2$ 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 than the evolute of a so called *Cartesian Oval* o. 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 o is the set of all points X, satisfying

$$\tau \overline{XF_1} + \overline{XF_2} = \lambda; \qquad \tau, \lambda \in \mathbb{R}. \tag{17}$$

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

The diacaustic of the the pencil of lines $F_1(f_1)$ with respect to the refracting curve o and a suitable index of refraction is just F_2.[13]

In case of a refraction we have to take

$$F_1 = E^*(\tfrac{R^2}{\mathbf{e}}, 0)^T, \quad F_2 = E, \quad \tau = -\frac{\mathbf{e}}{R}, \quad \lambda = \frac{R^2 - \mathbf{e}^2}{\mathbf{r}R}. \tag{18}$$

[12] The following considerations are all due to [19].

[13] That is why F_1 and F_2 are called *focal points* of o. Furthermore, F_1 and F_2 are the intersection points of isotropic tangents of o and satisfy the PLUECKER-Definition of a focal point. Considering the facts that o is an algebraic curve of order four and that it has cusps at the circular points at infinity, we can say that there exist even nine focal points, three of them on the x-axis.

The explicit equation of o can now easily be derived from (17):

$$o \ldots \left(\mathbf{r}^4 - \mathbf{r}^2\Big[(\mathbf{e}^2 - R^2)(x^2+y^2) - \mathbf{e}^2(R^2-2) + R^4\Big] + \mathbf{e}^4\right)^2 = \\ = 4R^2\mathbf{r}^2(\mathbf{r}^2-\mathbf{e}^2)^2\Big[(x-\mathbf{e})^2 + y^2\Big]. \tag{19}$$

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

$$\alpha_1 = \overline{EE^*} = \frac{e^2 - R^2}{e}, \qquad \lambda = \frac{R^2 - e^2}{rR}, \qquad \tau = -\frac{e}{R}$$

and

$$\Delta = \Delta(\theta) = \lambda^2 + \alpha_1^2\tau^2 - 2\alpha_1\lambda\tau\cos\theta - \alpha_1^2\sin^2\theta,$$

we have

$$(\tau^2 - 1)\varrho(\theta) = \lambda\tau - \alpha_1\cos\theta \pm \sqrt{\Delta}. \tag{20}$$

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

$$\begin{aligned} \Delta' &= 2\alpha_1\lambda\tau\sin\theta - 2\alpha_1^2\sin\theta\cos\theta, \\ \Delta'' &= 2\alpha_1\lambda\tau\cos\theta - 2\alpha_1^2\cos 2\theta, \\ (\tau^2-1)\varrho' &= \alpha_1\sin\theta \pm \frac{\Delta'}{2\sqrt{\Delta}}, \\ (\tau^2-1)\varrho'' &= \alpha_1\cos\theta \pm \frac{2\Delta\Delta'' - \Delta'^2}{4\Delta^{3/2}}. \end{aligned} \tag{21}$$

By substituting (21) 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 e, R and r only.[14] This parameter representation was used to draw Fig. 16.

Contents

1 Introduction and state of the art **1**

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

3 Refracting projecting rays through space points **4**

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

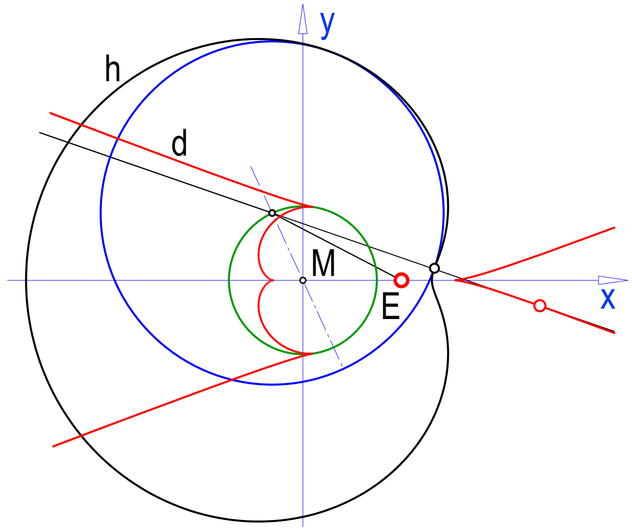

Figure 16: The diacaustic of a circle and its involute (Cartesian oval).

4 Diacaustic and characteristic conic of $\mathcal{R}[s;\mathbf{r};E]$ 6

5 Reciprocal refractions 10

6 Refraction on a plane 11

7 Image lifting – the virtual object transformation 13

8 Refraction on a circle 15

References

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

[2] J.W. Bruce, P.J. Giblin, C.G. Gibson: *On Caustics of Plane Curves.* Am. Math. Month. **88**, 1981, pp. 651–657.

[3] P. Burger, D. Gillies: *Interactive Computer Graphics.* Addison-Wesley Publishing Company, 1989.

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

[5] R. Freudenthal, E. Heinrich: *Kaustiken in vektorieller Behandlung.* Mathematisch Naturwissenschaftlicher Unterricht XII, 7, 1959/60, pp. 307 – 317.

[6] R. Freudenthal, E. Heinrich: *Grundkurven zu vorgegebenen Kaustiken.* Mathematisch Naturwissenschaftlicher Unterricht XIV, 2, 1961/62, pp. 70 – 76.

[7] J. D. Gergonne: *???.* Ann. Math. **15**, 1824,25.

[8] G. Glaeser: *Reflections on Spheres and Cylinders of Revolution.* Journal for Geometry and Graphics (JGG) **2** (2), Heidemann Verlag Berlin, 1999, pp. 1–19.

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

[10] H. Horninger: *Zur geometrischen Theorie der Spiegelung an krummen Oberflächen.* Akad.Wiss.Wien **145**, 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.III: Konstruktive Behandlung der Regelflächen.* Leipzig, Wien: Franz Deuticke, 1931.

[13] L. A. Quételet: *???.* M'em. Ac. Belg. **5**, 1829.

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

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

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

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

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

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

Received