*Hassan Azad, *Abdallah Laradji and **Asghar Qadir
*Department of Mathematical Sciences
King Fahd University of Petroleum & Minerals
Dhahran 31261, Saudi Arabia

**Department of Mathematics
Quaid-i-Azam University
Islamabad, Pakistan

1 Introduction

In standard textbooks (e.g., [7, p. 877]) as well as in Encylopaedia Britannica [6, Vol. 13, p.611], one finds variants of the following statement:

In spherical mirrors, rays parallel and very close to the principal axis are all reflected through a single point.

This statement is actually false, and can only be understood in the following sense: It is true only up to a linear approximation, and in fact, as soon as a mirror focuses parallel rays, it must be a parabolic mirror.

This note makes these ideas precise by providing in Section 2 a novel approach to reflecting properties of surfaces. There, inward and outward pointing vectors are introduced, leading to a straightforward derivation of reflecting properties of surfaces of revolution. In this same section, a formula is given for the reflected ray from an arbitrary smooth surface whose incident ray is parallel to a given direction. Its usefulness is demonstrated in Section 3 by applying it to spheres and paraboloids, where we show that, whereas parabolic mirrors are focusing surfaces, the only rays focused by the sphere have to originate at the center, where they are all reflected back through the center. The same method works for all surfaces of revolution generated by conic sections. (We refer the reader to, e.g., [7, p. 877] and [12] for erroneous explanations that spheres are focusing surfaces for a family of parallel rays.) In Section 4, we prove the main result which gives a characterization of surfaces that focus a pencil of rays from infinity or a finite point. Let us note that even for conic sections, the usual verification that they are focusing curves is not entirely straightforward as evidenced, for example, by the discussion in the archives of NRICH MATH [13]. We have written this note with teachers of undergraduate calculus in mind, hoping that our discussion is simple enough to be explained to their students.

The main result in this note is the following

Proposition 1 *If all rays from a finite point or a point at infinity focus, or appear to focus, after reflection from a smooth algebraic surface, then this surface must be a plane, a paraboloid of revolution, an ellipsoid of revolution or a sheet of a hyperboloid of revolution.*

Moreover, a pencil of incident rays on a spherical mirror never focuses unless the rays originate at the center of the sphere.

By a *smooth (or non-singular) algebraic surface* we mean a connected component S of a surface in $\mathbb{R}^3$ defined by a polynomial equation $E(x, y, z) = 0$, with non-zero gradient on S. A *pencil of rays* is simply a set of rays originating from a finite point or at a point at infinity (in this latter case, the rays are parallel). For a more advanced discussion of smoothness see [9, p. 95].

Some words about the origin of this paper are in order. It arose in an attempt by one of us to explain the focusing properties of the sphere, which are stated persistently in all modern books on physics, although it was already known to Diocles that this was not correct: see Hogendijk [8, p. 192] for an illuminating discussion and extensive references. Since the sphere does not have such properties, it was natural to ask for surfaces with these reflecting properties. After writing the first draft of this paper we found extensive work of Daniel Drucker and Phil Locke on this theme: see [3, 4, 5]. The main technical difference is the use of ideas of Section 2 of this paper which leads to substantial simplifications in the arguments. We subsequently discovered the truly classical nature of this circle of ideas. In Descartes [2, pp. 124–148], there is even a discussion of refracting properties of certain quartics, which were subsequently called *cartesians* by Salmon [10, p. 99]. Salmon, referring to Quetelet, also sketches a proof that the only curves which are refracting are these cartesians. As our results deal with surfaces and were independenly arrived at, they use contemporary ideas and standards of rigour and are sufficiently different that we decided to submit them for publication. We hope that this paper and its sequel [1] on refracting surfaces might encourage others to look for hidden gems in these classical sources and recast them in a contemporary framework.

2 Incident and Reflected Rays

Let $\Omega \subset \mathbb{R}^3$ be a region with a smooth boundary. This means that there is a differentiable function E such that $\Omega = \{P \in \mathbb{R}^3 : E(P) < 0\}$ and $\partial\Omega$ − the boundary of Ω − is the set of all points P where $E(P) = 0$. We assume that $\partial\Omega$ is nonempty and E has a nonzero gradient on the boundary of Ω.

Definition 2 *Given a point $P \in \partial\Omega$, we say that a unit vector $\vec{e}$ points outward if for sufficiently small $\delta > 0$, the points $P + \delta\vec{e}$ are outside $\overline{\Omega}$. Equivalently, if $\nabla E(P) \cdot \vec{e} > 0$, then $\vec{e}$ is an outward pointing vector at P. (The equivalence of these statements follows from a second degree Taylor polynomial.)*

Inward pointing vectors are defined similarly. In particular, the inward normal at a boundary point P is $\overrightarrow{n_P} = \dfrac{-\nabla E(P)}{|\nabla E(P)|}$. An incident ray at a boundary point P is a ray of the form $P - t\vec{e}\ (t > 0)$ for some outward pointing vector $\vec{e}$ at P: (see Fig. 1). So, if Q is a point of Ω on this ray, then $\overrightarrow{QP}$ is a positive multiple of $\vec{e}$.

The corresponding reflected ray at P is a ray of the form $P + t\vec{f}\ (t > 0)$ for an inward pointing unit vector $\vec{f}$ at P which is uniquely determined by the following requirements:

If $\vec{e}$ is normal at P, then the reflected ray coincides with the incident ray. If $\vec{e}$ is not normal at P, then the reflected ray at P lies in the plane at P determined by $\vec{e}$ and $\overrightarrow{n_P}$,

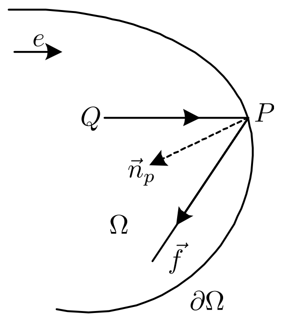

Fig 1. *A paraxial ray (parallel to the chosen vector $\vec{e}$) is depicted by $\overrightarrow{QP}$ and the unit normal at P by $\vec{n}_p$. The region Ω is shaded here and $\partial\Omega$ is its boundary. The reflected ray along $\vec{f}$ must make the same angle with $\vec{n}_p$ as $\overrightarrow{QP}$.*

whose direction vector $\vec{f}$ is not a a multiple of $\vec{e}$ and $\vec{f}$ satisfies Snell's (or Descartes') law of reflection:

$$\vec{f} \cdot \overrightarrow{n_P} = -\vec{e} \cdot \overrightarrow{n_P}.$$

A computation shows that, in all cases,

$$\vec{f} = \vec{e} - 2(\vec{e} \cdot \overrightarrow{n_P})\overrightarrow{n_P}$$

We summarize this as follows:

Proposition 3 *If $\vec{e}$ is an outward pointing vector at a boundary point P of Ω and $P - t\vec{e}$ $(t \geq 0)$ is an incident ray at P, then its reflected ray at P is*

$$P + t(\vec{e} - 2(\vec{e} \cdot \overrightarrow{n_P})\overrightarrow{n_P}) \qquad (t \geq 0),$$

where $\overrightarrow{n_P} = \dfrac{-\nabla E(P)}{|\nabla E(P)|}$ is the inner normal at P.

3 The Sphere and the Paraboloid

We apply the formalism of Section 2 to study focusing properties of spheres and paraboloids. The same method works for determining the reflected ray of a ray incident on a given surface.

A) **Spheres**: Let $E(x, y, z) = x^2 + y^2 + z^2 - R^2$. The domain defined by $E < 0$ has the sphere as its boundary. So $\nabla E = \langle 2x, 2y, 2z \rangle$ is an outward normal and $\vec{e} = \langle 0, 0, 1 \rangle$ points outward at $P = (x, y, z)$ if $z > 0$. For such a point P, a ray parallel to $\vec{e}$ and incident at P is reflected to the ray $P + t(\vec{e} - 2(\vec{e} \cdot \vec{n})\vec{n})$ $(t \geq 0)$, where $\vec{n} = \langle -x/R, -y/R, -z/R \rangle$ is the unit inward normal at P. Therefore, the reflected ray at $P(x, y, z)$ is

$$\left(x - 2t\frac{xz}{R^2}, y - 2t\frac{yz}{R^2}, z + t\left(1 - 2\frac{z^2}{R^2}\right)\right).$$

This ray intersects the z-axis at the point given by $t = R^2/2z$; i.e. at the point $(0, 0, R^2/2z)$.

Note that if all rays parallel to $\vec{e} = \langle 0, 0, 1\rangle$ and near the z-axis focus at a point, it must necessarily be on the z-axis. Therefore, more than two parallel rays never focus after reflection from a spherical mirror (cf. [6] and [7])].

B) **Paraboloids**: A paraboloid of revolution has equation $z = c(x^2+y^2)$. Let $E(x, y, z) = c(x^2+y^2)-z$. We assume that $c > 0$. The domain given by $E < 0$ has the paraboloid as its boundary, and as $\nabla E = \langle 2cx, 2cy, -1\rangle$, the vector $\vec{e} = \langle 0, 0, -1\rangle$ is an outward pointing vector and $\overrightarrow{n} = -\dfrac{\nabla E}{|\nabla E|} = \dfrac{\langle -2cx, -2cy, 1\rangle}{\sqrt{1+4cz}}$ is an inward normal vector to the paraboloid. Therefore, if $P(x, y, z)$ is a point on the paraboloid, then a ray parallel to $\vec{e} = \langle 0, 0, -1\rangle$ and incident at P is reflected to the ray $P + t(\vec{e} - 2(\vec{e}\cdot\vec{n})\vec{n})$. This works out to be the ray

$$\left(x - \frac{4cxt}{1+4cz}, y - \frac{4cyt}{1+4cz}, z + \frac{(1-4cz)t}{1+4cz}\right) \quad (t \geq 0).$$

This ray passes through the z-axis for $t = \dfrac{1+4cz}{4d}$ ($t > 0$ because $P(x, y, z)$ is on the paraboloid), and so it passes through $\left(0, 0, \dfrac{1}{4c}\right)$.

Notice that if all rays parallel to $\vec{e} = \langle 0, 0, -1\rangle$ focus, then they must focus on the z-axis, because the z-axis itself is one of these incident rays. Therefore, all the rays parallel to $\vec{e} = \langle 0, 0, -1\rangle$ focus at the point $\left(0, 0, \dfrac{1}{4c}\right)$, which is inside the paraboloid [cf. 7].

The computations for rays originating from foci of ellipsoids of revolution and two-sheeted hyperboloids of revolution are similar.

4 Focusing Surfaces

We prove Proposition 1 in several steps. A sketch of the argument is as follows: The assumption that rays from infinity or a finite point focus after reflection from a surface implies that the surface is a surface of revolution, i.e. invariant under rotations with respect to some fixed axis. This reduces the problem to a two-dimensional situation.

In the case of parallel incident rays, the focus lies on the axis of rotation (see Fig. 2). The normal at a point P of the reflecting surface intersects the axis of rotation at a point Q which is equidistant from the incident and the reflected ray. This gives the differential equation for the profile curve.

In the case of a pencil of rays emerging from a point F_1 and focusing to a point F_2 after reflection, the axis of rotation is the line joining F_1 and F_2. The normal at a point P of the reflecting surface intersects the axis of rotation at a point Q which is equidistant

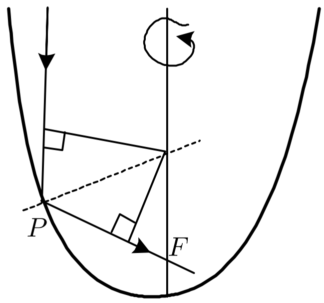

Fig 2.

from the incident and the reflected rays. This again gives the differential equation for the profile curve.

For the remainder of this section, we assume that the reflecting surface is a smooth algebraic surface, and rays from a finite point or from a point at infinity all focus or appear to focus, after reflection, to a finite point.

Step 1: *The surface is a surface of revolution.* Assume first that all rays parallel to a given direction focus after reflection from the surface S to a point F_0. We may assume that this direction is given by $\vec{e} = \langle 1, 0, 0 \rangle$ and the focus F_0 has coordinates $(c, 0, 0)$. So, for any point of S which lies on the x-axis, the incident, reflected and normal rays are all along the x-axis. For a point P of S which is not on the x-axis, the incident, reflected and normal rays are coplanar and the x-axis passes through this plane. Since P is not on the x-axis, the normal $\overrightarrow{n_P}$ is not parallel to $\vec{e}$, and so it intersects the x-axis, say at the point $(a, 0, 0)$. Therefore, if $P = (x, y, z)$, then $(x - a, y, z)$ is a non-zero multiple of $\nabla E(P) = \left\langle \frac{\partial E(P)}{\partial x}, \frac{\partial E(P)}{\partial y}, \frac{\partial E(P)}{\partial z} \right\rangle$. Hence $y = \lambda \frac{\partial E(P)}{\partial y}$, $z = \lambda \frac{\partial E(P)}{\partial z}$ for some λ, i.e. $y \frac{\partial E(P)}{\partial z} - z \frac{\partial E(P)}{\partial y} = 0$. Using cylindrical coordinates with $y = r \cos\theta$, $z = r \sin\theta$, we obtain $\frac{\partial E(P)}{\partial \theta} = \frac{\partial E(P)}{\partial y} \frac{\partial y}{\partial \theta} + \frac{\partial E(P)}{\partial z} \frac{\partial z}{\partial \theta} = 0$. This shows that the surface S given by $E(x, y, z) = 0$ is a surface of revolution about the x-axis. (In more technical terms, this means that the vector field $y \frac{\partial}{\partial z} - z \frac{\partial}{\partial y}$ is tangent to S with defining equation $E = 0$. This field is generated by rotations whose axis is along the x-axis, and therefore, S is a surface of revolution.)

Now assume that rays originating at a point F_1 focus to a point F_2 after reflection from the surface with defining equation $E = 0$. We may assume that $F_1 = (-c, 0, 0)$ and $F_2 = (c, 0, 0)$. If the normal ray at P is parallel to the focal line, then $\frac{\partial E}{\partial y}(P) = 0$, $\frac{\partial E}{\partial z}(P) = 0$; otherwise, the normal ray at P is not parallel to the line $F_1 F_2$ and so it intersects it at some point $(a, 0, 0)$. Hence, we again have $y \frac{\partial E(P)}{\partial z} - z \frac{\partial E(P)}{\partial y} = 0$ for all points $P(x, y, z)$ on the surface. This implies that the surface S is invariant under rotations whose axis is the x-axis. In particular, S is obtained by rotating its trace in a plane containing the x-axis. Without loss of generality, we may take this plane to be the xy-plane.

Step 2: *The trace in the xy-plane is a smooth curve, except for possible singularities along the focal line.* The trace $\mathcal{C}$ in the xy-plane is given by the equation $E(x, y, 0) = 0$. Assume that there is a point $P(x, y, 0)$ on $\mathcal{C}$ where $\frac{\partial E}{\partial y}(P) = 0$. By the above paragraph, $y\frac{\partial E(P)}{\partial z} - z\frac{\partial E(P)}{\partial y} = 0$, so that $\frac{\partial E(P)}{\partial z} = 0$ or $y = 0$. Since S is a smooth algebraic surface, $\frac{\partial E}{\partial x}(P) \neq 0$ when $y \neq 0$. Therefore, the singular points, if any, of the trace are all along the axis of rotation. As far as reflections are concerned, these points play no role.

Step 3: *The trace on the xy-plane is part of a conic section or of the y-axis.* First note that we have chosen coordinates so that the foci - or focus in the case of parallel incident rays- lie on the x-axis, which is also the axis of rotation. Since the incident, reflected and normal ray all lie in the same plane and the angle of incidence is the same as the angle of reflection, the normal at a point P of the reflecting surface intersects the axis of rotation at a point Q which is equidistant from the incident and reflected ray.

Let $P(x, y)$ be a point on the curve $\mathcal{C}$ such that the normal $\overrightarrow{n_P}$ to $\mathcal{C}$ at P is not parallel to the x-axis, and let $Q(a, 0)$ be the point where $\overrightarrow{n_P}$ intersects the x-axis, so that

$$(x - a)dx + ydy = 0. \tag{1}$$

In the case of parallel incident rays, we obtain by the equidistance property above that $PF = QF$, where $F(c, 0)$ is the focus of the reflected rays. Thus $(x - c)^2 + y^2 = (a - c)^2$. Put $u^2 = (x - c)^2 + y^2$, so that $udu = (x - c)dx + ydy = (a - c)dx$, which gives $(udu)^2 = ((a - c)dx)^2 = (udx)^2$. Since n_p is not parallel to the x-axis, $y \neq 0$ and therefore $u \neq 0$. Hence $du = \pm dx$, and $(x - c)^2 + y^2 = (K \pm x)^2$ for some constant K, i.e. $y^2 = Ax + B$, where $A = 2(c \pm K)$, $B = K^2 - c^2$. If $A = 0$ then $B = 0$ and $\mathcal{C}$ would be along the x-axis, which is plainly impossible. Hence $\mathcal{C}$ is a parabola.

In the case of incident rays originating at a finite point $F'(-c, 0)$ and reflected by $\mathcal{C}$ through the point $F(c, 0)$, the angles QPF' and FPQ are equal, and hence the sine law for the triangles QPF' and FPQ yields $PF \cdot QF' = PF' \cdot QF$, i.e.

$$(a + c)^2((x - c)^2 + y^2) = (a - c)^2((x + c)^2 + y^2). \tag{2}$$

Put $u^2 = (x - c)^2 + y^2$ and $v^2 = (x + c)^2 + y^2$, so that $udu = (x - c)dx + ydy$ and $vdv = (x + c)dx + ydy$. By (1) $udu = (a - c)dx$ and $vdv = (a + c)dx$, and by (2) $(a + c)^2u^2 = (a - c)^2v^2$, so that $u^2v^2((du)^2 - (dv)^2) = 0$. Since $y \neq 0$, $du = \pm dv$, and so, for some constant K, $u = K \pm v$. We therefore obtain $(x - c)^2 + y^2 = K^2 \pm 2Kv + (x + c)^2 + y^2$, i.e.

$$(4K^2 - 16c^2)x^2 + 4K^2y^2 = K^4 - 4K^2c^2. \tag{3}$$

If $c = 0$ then F, F' and Q coincide with the origin and $xdx + ydy = 0$, which implies $x^2 + y^2 = r^2$ for some constant r, and $\mathcal{C}$ is part of a circle. If $K = 0$ and $c \neq 0$, then $x = 0$ and $\mathcal{C}$ would be on the y-axis, which is impossible, because of the reflecting property of the curve. If $K \neq 0$, then $4K^2 - 16c^2 \neq 0$ (otherwise $y = 0$), and $\mathcal{C}$ is an ellipse or a branch of a hyperbola.

Assume now that the normal $\overrightarrow{n_P}$ to $\mathcal{C}$ at P is parallel to the x-axis. In the case of parallel incident rays, the reflected ray at P must also be parallel to the x-axis, and since it contains the focus $F(c,0)$, P itself must be on the x-axis. Since S is a smooth algebraic surface and is generated by revolving $\mathcal{C}$ about the x-axis, the intersection of $\mathcal{C}$ with the x-axis is finite, so that P,by continuity, satisfies Equation (3). If the incident ray originates at $F'(-c,0)$ then F is on the line containing the reflected ray and $PF = PF'$, so that P is on the y-axis. In this case, either $\mathcal{C}$ intersects the y-axis at infinitely many points, and then $\mathcal{C}$ is part of the y-axis, or at a finite number of points and then, by continuity again, P satisfies Equation (3).

Remarks. 1. The disc with the Poincaré metric is a classical example of a non-euclidean geometry which has many features in common with ordinary geometry. (See [11].) For example, any two points can be joined by a unique curve of shortest length, so one can define ellipses, etc. as in Euclidean geometry. An interesting undergraduate project could be to investigate whether these curves have similar focusing properties. There are several computer programs – for instance [11] – that one can use for experimenting with these ideas.

2. At a more advanced level, we note that the argument in Step 1 above carries over to the general case of focusing hypersurfaces in Euclidean n-dimensional space because the focusing property gives invariance under vector fields which generate the Lie algebra of $\mathrm{SO}(n-1)$. The set-up of Section 2 generalizes to simply connected homogeneous spaces of non-positive curvature by replacing rays by geodesic rays since any two points can be joined by a unique geodesic and the exponential map is bijective.

References

[1] H. Azad and A. Laradji, Refracting surfaces, in preparation.

[2] R. Descartes, Geometry, Second Book, Dover Publications (1954).

[3] D. Drucker, Euclidean hypersurfaces with reflection properties, Geom. Dedicata 33 (1990), 325-329.

[4] D. Drucker, Reflection properties of curves and surfaces, Math. Mag. 65 (1992), no. 3, 147–157.

[5] D. Drucker and P. Locke, A natural classification of curves and surfaces with reflection properties, Math. Mag. 69 (1996), no. 4, 249-256.

[6] Encyclopaedia Britannica, Encyclopaedia Britannica Educational Corporat, 15th edition (1991).

[7] D. Halliday, R. Resnick and J. Walker, Fundamentals of Physics Extended, 5th Edition, John Wiley & Sons (1997).

[8] J. Hogendijk, The burning mirrors of Diocles: reflections on the methodology and purpose of the history of pre-modern science, Early Sci. Med.7 (2002), no. 3, 181-197.

[9] M. Reid, Undergraduate Algebraic Geometry, Cambridge University Press, Cambridge, 1988.

[10] G. Salmon, Higher Plane Curves, 3rd Edition, Chelsea Publishing Company (1960).

[11] http://www.geom.uiuc.edu/docs/forum/

[12] http://www.glenbrook.k12.il.us/gbssci/phys/Class/refln/reflntoc.html

[13] http://nrich.maths.org/askedNRICH/edited/4989.html