Computational Methods and Function Theory (2019) 19:135–155
https://doi.org/10.1007/s40315-018-0257-z

The Ptolemy–Alhazen Problem and Spherical Mirror Reflection

Masayo Fujimura[1] · **Parisa Hariri**[2] · **Marcelina Mocanu**[3] · **Matti Vuorinen**[2]

Received: 29 August 2018 / Revised: 6 September 2018 / Accepted: 6 September 2018 /
Published online: 15 December 2018

Abstract
An ancient optics problem of Ptolemy, studied later by Alhazen, is discussed. This problem deals with reflection of light in spherical mirrors. Mathematically, this reduces to the solution of a quartic equation, which we solve and analyze using a symbolic computation software. Similar problems have been recently studied in connection with ray-tracing, catadioptric optics, scattering of electromagnetic waves, and mathematical billiards, but we were led to this problem in our study of the so-called triangular ratio metric.

Keywords Triangular ratio metric · Ptolemy–Alhazen problem · Reflection of light

Mathematics Subject Classification 30C20 · 30C15 · 51M99

1 Introduction

The Greek mathematician Ptolemy (ca. 100–170) formulated a problem concerning reflection of light at a spherical mirror surface: *Given a light source and a spherical*

Communicated by Stephan Ruscheweyh.

✉ Parisa Hariri
parisa.hariri@utu.fi

Masayo Fujimura
masayo@nda.ac.jp

Marcelina Mocanu
mmocanu@ub.ro

Matti Vuorinen
vuorinen@utu.fi

1 Department of Mathematics, National Defense Academy of Japan, Yokosuka, Japan

2 Department of Mathematics and Statistics, University of Turku, Turku, Finland

3 Department of Mathematics and Informatics, Vasile Alecsandri University of Bacau, Bacău, Romania

mirror, find the point on the mirror where the light will be reflected to the eye of an observer.

Alhazen (ca. 965–1040) was a scientist who lived in Iraq, Spain, and Egypt and extensively studied several branches of science. For instance, he wrote seven books about optics and studied, e.g., Ptolemy's problem as well as many other problems of optics and is considered to be one of the greatest researchers of optics before Kepler [2]. Often the above problem is known as Alhazen's problem [10, p. 1010]. At the end of this introduction, we will point out various applications and earlier results connected with the Ptolemy–Alhazen problem.

We will consider the two-dimensional version of the problem and present an algebraic solution for it. The solution reduces to a quartic equation which we solve with symbolic computation software.

Let $\mathbb{D}$ be the unit disk $\{z \in \mathbb{C} : |z| < 1\}$, and suppose that the circumference $\partial\mathbb{D} = \{z \in \mathbb{C} : |z| = 1\}$ is a reflecting curve. This two-dimensional problem reads: Given two points $z_1, z_2 \in \mathbb{D}$, find $u \in \partial\mathbb{D}$ such that

$$\measuredangle(z_1, u, 0) = \measuredangle(0, u, z_2). \tag{1.1}$$

Here, $\measuredangle(z, u, w)$ denotes the radian measure in $(-\pi, \pi]$ of the oriented angle with initial side $[u, z]$ and final side $[u, w]$. This equality condition for the angles says that the angles of incidence and reflection are equal, a light ray from z_1 to u is reflected at u and goes through the point z_2. Recall that, according to Fermat's principle, light travels between two points along a path of extremal time, as compared to other nearby paths. One proves that $u = e^{it_0}$, $t_0 \in \mathbb{R}$ satisfies (1.1) if and only if t_0 is a critical point of the function $t \mapsto \left|z_1 - e^{it}\right| + \left|z_2 - e^{it}\right|$, $t \in \mathbb{R}$. In particular, condition (1.1) is satisfied by the extremum points (a minimum point and a maximum point, at least) of the function $u \mapsto |z_1 - u| + |z_2 - u|$, $u \in \partial\mathbb{D}$.

We call this the *interior problem*—there is a natural counterpart of this problem for the case when both points are in the exterior of the closed unit disk, called the *exterior problem*. Indeed, this exterior problem corresponds to Ptolemy's questions about light source, spherical mirror, and observer. As we will see below, the interior problem is equivalent to finding the maximal ellipse with foci at z_1, z_2 contained in the unit disk, and the point of reflection $u \in \partial\mathbb{D}$ is the tangent point of the ellipse with the circumference. Algebraically, this leads to the solution of a quartic equation as we will see below.

We met this problem in a different context, in the study of the triangular ratio metric s_G of a given domain $G \subset \mathbb{R}^2$ defined as follows for $z_1, z_2 \in G$ [6,12]

$$s_G(z_1, z_2) = \sup_{z \in \partial G} \frac{|z_1 - z_2|}{|z_1 - z| + |z - z_2|}. \tag{1.2}$$

By compactness, this supremum is attained at some point $z_0 \in \partial G$. If G is convex, it is simple to see that z_0 is the point of contact of the boundary with an ellipse, with foci z_1, z_2, contained in G. Now for the case $G = \mathbb{D}$ and $z_1, z_2 \in \mathbb{D}$, if the extremal point is $z_0 \in \partial\mathbb{D}$, the connection between the triangular ratio distance

$$s_{\mathbb{D}}(z_1, z_2) = \frac{|z_1 - z_2|}{|z_1 - z_0| + |z_2 - z_0|}$$

and the Ptolemy–Alhazen interior problem is clear: $u{=}z_0$ satisfies (1.1). Note that (1.1) is just a reformulation of a basic property of the ellipse with foci z_1, z_2 : the normal to the ellipse (which in this case is the radius of the unit circle terminating at the point u) bisects the angle formed by segments joining the foci z_1, z_2 with the point u. During the past decade, the s_G metric has been studied in several papers e.g. by P. Hästö [13,14]; the interested reader is referred to [12] and the references there.

We study the Ptolemy–Alhazen interior problem and in our main result, Theorem 1.1, we give an equation of degree four that yields the reflection point on the unit circle. Standard symbolic computation software can then be used to find this point numerically. We also study the Ptolemy–Alhazen exterior problem.

Theorem 1.1 *The point u in (1.1) is given as a solution of the equation*

$$\overline{z_1 z_2} u^4 - (\overline{z_1} + \overline{z_2}) u^3 + (z_1 + z_2) u - z_1 z_2 = 0. \tag{1.3}$$

It should be noted that the Eq. (1.3) may have roots in the complex plane that are not on the unit circle, and of the roots on the unit circle, we must choose one root u , that minimizes the sum $|z_1 - u| + |z_2 - u|$. We call this root *the minimizing root* of (1.3).

Corollary 1.2 *For $z_1, z_2 \in \mathbb{D}$ we have*

$$s_{\mathbb{D}}(z_1, z_2) = \frac{|z_1 - z_2|}{|z_1 - u| + |z_2 - u|}$$

where $u \in \partial\mathbb{D}$ is the minimizing root of (1.3).

As we will see below, the minimizing root need not be unique.

We have used Risa/Asir symbolic computation software [20] in the proofs of our results. We give a short Mathematica code for the computation of $s_{\mathbb{D}}(z_1, z_2)$.

Theorem 1.1 is applicable not merely to light signals but whenever the angles of incidence and reflection of a wave or signal are equal, for instance, in the case of electromagnetic signals like radar signals or acoustic waves. H. Bach [4] has made numerical studies of Alhazen's ray-tracing problem related to circles and ellipses. A.R. Miller and E. Vegh [18] have studied the exterior Ptolemy–Alhazen problem and computed the grazing angle of specular reflection (the complement of the equal angles of incidence and of reflection) using a quartic equation, which is not the same as (1.3). They did not consider the problem of finding the point of incidence in the case of specular reflection, which is solved through Eq. (1.3).

Mathematical theory of billiards also leads to similar studies: see for instance the paper by M. Drexler and M.J. Gander [9]. The Ptolemy–Alhazen problem also occurs in computer graphics and catadioptric optics [1]. The well-known lithograph of M. C.

Escher named “Hand with reflecting sphere” demonstrates nicely the idea of reflection from a spherical mirror.

2 Algebraic solution to the Ptolemy–Alhazen problem

In this section, we prove Theorem 1.1 and give an algorithm for computing $s_{\mathbb{D}}(z_1, z_2)$ for $z_1, z_2 \in \mathbb{D}$.

Problem 2.1 *For $z_1, z_2 \in \mathbb{D}$, find the point $u \in \partial\mathbb{D}$ such that the sum $|z_1 - u| + |z_2 - u|$ is minimal.*

The point u is given as the point of tangency of an ellipse $|z - z_1| + |z - z_2| = r$ with the unit circle.

Remark 2.2 For $z_1, z_2 \in \mathbb{D}$, if $u \in \partial\mathbb{D}$ is the point of tangency of an ellipse $|z - z_1| + |z - z_2| = r$ and the unit circle, then r is given by

$$r = |2 - \overline{u}z_1 - u\overline{z_2}|.$$

In fact, from the “reflective property” $\measuredangle(z_1, u, 0) = \measuredangle(0, u, z_2)$ of an ellipse, the following holds

$$\arg \frac{u}{u - z_1} = \arg \frac{u - z_2}{u} = -\arg \frac{\overline{u} - \overline{z_2}}{\overline{u}}, \tag{2.1}$$

and

$$\arg(\overline{u}(u - z_1)) = \arg(u(\overline{u} - \overline{z_2})). \tag{2.2}$$

Since the point u is on the ellipse $|z - z_1| + |z - z_2| = r$ and satisfies $u\overline{u} = 1$, we have

$$\begin{aligned} r &= |u - z_1| + |u - z_2| = |\overline{u}(u - z_1)| + |u(\overline{u} - \overline{z_2})| \\ &= |\overline{u}(u - z_1) + u(\overline{u} - \overline{z_2})| = |2 - \overline{u}z_1 - u\overline{z_2}|. \end{aligned}$$

2.1 Proof of Theorem 1.1

From the Eq. (2.1), we have

$$\arg\left(\frac{u - z_1}{u} \cdot \frac{u - z_2}{u}\right) = 0.$$

This implies $\dfrac{(u - z_1)(u - z_2)}{u^2}$ is real and its complex conjugate is also real. Hence,

$$\frac{(u - z_1)(u - z_2)}{u^2} = \frac{(\overline{u} - \overline{z_1})(\overline{u} - \overline{z_2})}{\overline{u^2}}$$

holds. Since u satisfies $u\overline{u} = 1$, we have the assertion. □

Remark 2.3 The solution of (1.3) includes all the tangent points of the ellipse $|z - z_1| + |z - z_2| = |2 - \overline{u}z_1 - u\overline{z_2}|$ and the unit circle. (See Figs. 1, 2). Figure 2 displays a situation where all the roots of the quartic equation have unit modulus. However, this is not always the case for the Eq. (1.3). E.g., if $z_1 = 0.5 + (0.1 \cdot k)i$, $k = 1, .., 5,\ z_2 = 0.5$, the Eq. (1.3) has two roots of modulus equal to 1 and two roots off the unit circle, see Fig. 3. Miller and Vegh [18] computed the grazing angle of specular reflection using a quartic self-inversive polynomial equation, which is not the same as (1.3). Note that all the roots of their equation have modulus equal to one. They have also studied the Ptolemy–Alhazen problem using a quartic equation, that is different from our equation and, moreover, all the roots of their equation have modulus equal to one Fig. 3.

We say that a polynomial $P(z)$ is *self-inversive* if $P(1/\overline{u}) = 0$ whenever $u \neq 0$ and $P(u) = 0$. It is easily seen that the quartic polynomial in (1.3) is self-inversive. Note that the points u and $1/\overline{u}$ are obtained from each other by the inversion transformation $w \mapsto 1/\overline{w}$.

It is clear from the compactness of the unit circle, that the function $|z_1 - z| + |z_2 - z|$ attains its maximum and minimum on the unit circle. However, as a property of the Eq. (1.3) itself, the following results can also be derived.

Lemma 2.4 *The Eq.* (1.3) *always has at least two roots of modulus equal to* 1.

Proof Consider first the case, when $z_1 z_2 = 0$. In this case, the Eq. (1.3) has two roots u, $|u| = 1$, with $u^2 = z_1/\overline{z_1} \in \partial\mathbb{D}$ if $z_2 = 0, z_1 \neq 0$. (The case $z_1 = z_2 = 0$ is trivial.) Suppose that the equation has no root on the unit circle $\partial\mathbb{D}$.

By the invariance property pointed out above, if $u_0 \in \mathbb{C} \setminus (\{0\} \cup \partial\mathbb{D})$ is a root of (1.3), then $1/\overline{u_0}$ also is a root of (1.3). Hence, the number of roots off the unit circle is even and the number of roots on the unit circle must also be even. We will now show that this even number is either 2 or 4.

Let $a, b, \alpha, \beta \in \mathbb{R}, 0 < a < 1, 0 < b < 1$, and let

$$ae^{i\alpha},\ \frac{1}{a}e^{i\alpha},\ be^{i\beta},\ \frac{1}{b}e^{i\beta}$$

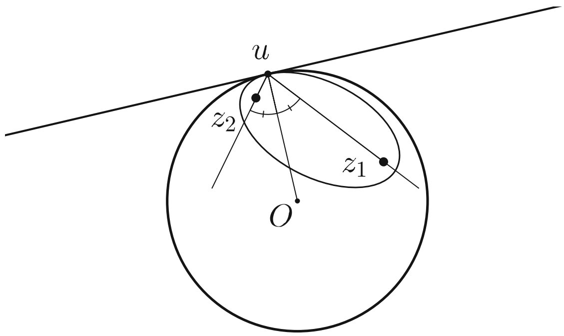

Fig. 1 Light reflection on a circular arc: The angles of incidence and reflection are equal. Ptolemy–Alhazen interior problem: Given z_1 and z_2, find u. The maximal ellipse contained in the unit disk with foci z_1 and z_2 meets the unit circle at u

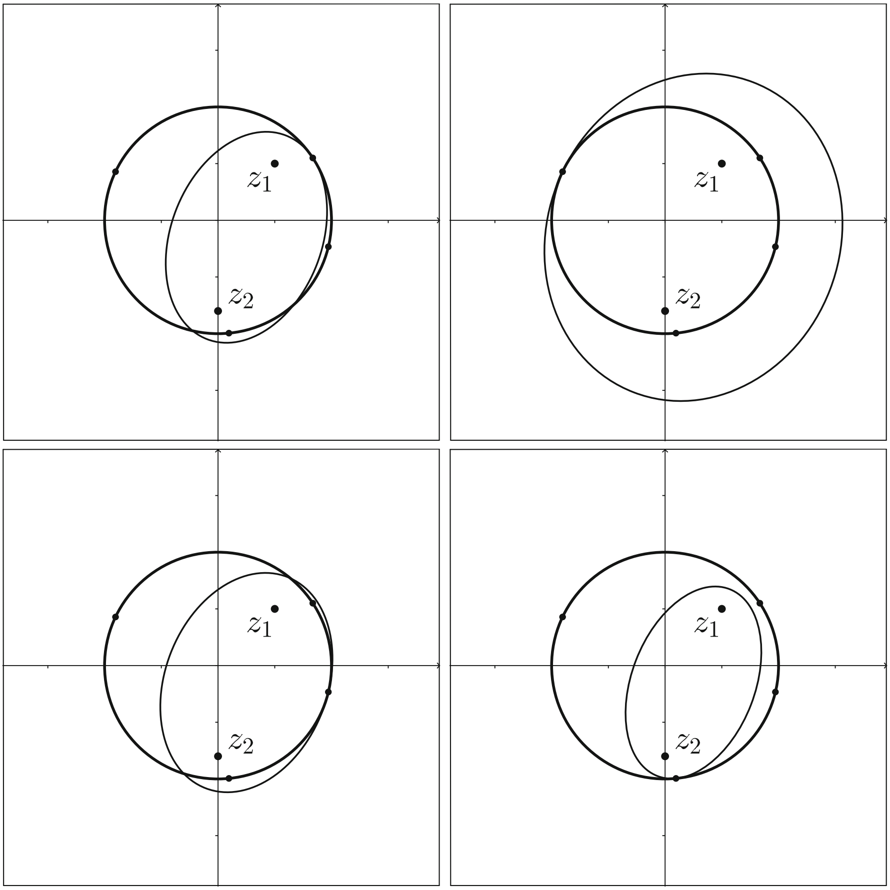

Fig. 2 This figure indicates the four solutions of (1.3) (dots on the unit circle) and the ellipse that corresponds to each u, for $z_1 = 0.5 + 0.5i$, $z_2 = -0.8i$. The figure on the lower right shows the point u that gives the minimum

be the four roots of the Eq. (1.3). Then, the equation

$$z_1 z_2 (u - a e^{i\alpha}) \left(u - \frac{1}{a} e^{i\alpha}\right) (u - b e^{i\beta}) \left(u - \frac{1}{b} e^{i\beta}\right) = 0 \tag{2.3}$$

coincides with (1.3). Therefore, the coefficient of degree two of (2.3) vanishes, and we have

$$e^{i2\alpha} + e^{i2\beta} = -\left(a + \frac{1}{a}\right)\left(b + \frac{1}{b}\right) e^{i(\alpha+\beta)}. \tag{2.4}$$

The absolute value of the left hand side of (2.4) satisfies

$$|e^{i2\alpha} + e^{i2\beta}| \leq 2. \tag{2.5}$$

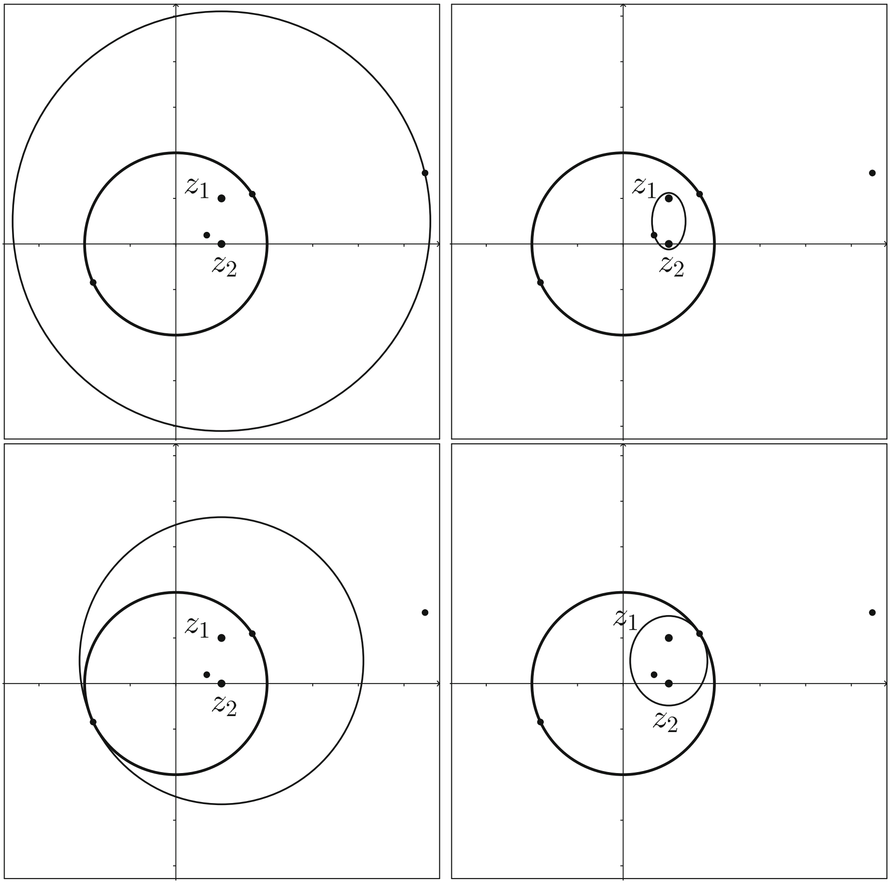

Fig. 3 For $z_1 = 0.5 + 0.5i$ and $z_2 = 0.5$, there are only two solutions of (1.3) on the unit circle. The figure on the lower right shows the point u that gives the minimum

On the other hand, the absolute value of the right hand side of (2.4) satisfies

$$\left|\left(a + \frac{1}{a}\right)\left(b + \frac{1}{b}\right)e^{i2(\alpha+\beta)}\right| = \left|a + \frac{1}{a}\right|\left|b + \frac{1}{b}\right| > 4, \tag{2.6}$$

because the function $f(x) = x + \frac{1}{x}$ is monotonically decreasing on $0 < x \leq 1$ and $f(1) = 2$. The inequalities (2.5) and (2.6) imply that the equality (2.4) never holds. Hence (1.3) has roots of modulus equals to 1. $\square$

Remark 2.5 We consider here several special cases of the Eq. (1.3) and for some special cases, we give the corresponding formula for the $s_{\mathbb{D}}$ metric which readily follows from Corollary 1.2.

Case 1. $z_1 \neq 0 = z_2$ (cubic equation). The Eq. (1.3) is now $(-\overline{z_1})\, u^3 + z_1 u = 0$ and has the roots $u_1 = 0$, $u_{2,3} = \pm\frac{z_1}{|z_1|}$ and for $z \in \mathbb{D}$

$$s_{\mathbb{D}}(0, z) = \frac{|z|}{2 - |z|}\,.$$

Case 2. $z_1 + z_2 = 0$, $z_1 \neq 0$. The Eq. (1.3) reduces now to:

$$\left(-\overline{z_1}^2\right) u^4 + z_1^2 = 0 \Leftrightarrow u^4 = \left(\frac{z_1}{\overline{z_1}}\right)^2 \Leftrightarrow u^4 = \left(\frac{z_1}{|z_1|}\right)^4.$$

The roots are: $u_{1,2} = \pm\frac{z_1}{|z_1|}$, $u_{3,4} = \pm i\frac{z_1}{|z_1|}$ (four distinct roots of modulus 1) and for $z \in \mathbb{D}$

$$s_{\mathbb{D}}(z, -z) = |z|\,.$$

Case 3. $z_1 = z_2 \neq 0$. Clearly $s_{\mathbb{D}}(z, z) = 0$. Denote $z := z_1 = z_2$. The Eq. (1.3) reduces now to:

$$\overline{z}^2u^4 - 2\overline{z}u^3 + 2zu - z^2 = (\overline{z}u^2 - z)(\overline{z}u^2 - 2u + z) = 0\,.$$

Then, we see that $u_{1,2} = \pm\frac{z}{|z|}$ are roots. The other roots are:

1) If $|z| < 1$, then $u_{3,4} = \frac{1}{\overline{z}}\left(1 \pm \sqrt{1 - |z|^2}\right)$ (with $|u_3| > 1$, $|u_4| < 1$)
2) If $|z| > 1$, then $u_{3,4} = \frac{1}{\overline{z}}\left(1 \pm i\sqrt{|z|^2 - 1}\right)$ (with $|u_3| = |u_4| = 1$).

Case 4. $|z_1| = |z_2| \neq 0$.
Denote $\rho = |z_1| = |z_2|$. Using a rotation around the origin and a change of orientation, we may assume that $\arg z_2 = -\arg z_1 =: \alpha$, where $0 \leq \alpha \leq \frac{\pi}{2}$. The Eq. (1.3) reads now: $\rho^2u^4 - 2\rho\,(\cos\alpha)\, u^3 + 2\rho\,(\cos\alpha)\, u - \rho^2 = 0$

$$\rho^2u^4 - 2\rho\,(\cos\alpha)\, u^3 + 2\rho\,(\cos\alpha)\, u - \rho^2 = \rho^2\left(u^2 - 1\right)\left(u^2 - \frac{2\cos\alpha}{\rho}u + 1\right)$$

The roots are: $u_{1,2} = \pm 1$ and

1) If $0 < \rho < \cos\alpha$, then $u_{3,4} = \frac{\cos\alpha}{\rho} \pm \sqrt{\left(\frac{\cos\alpha}{\rho}\right)^2 - 1}$ (here $|u_3| > 1$, $|u_4| < 1$)
2) If $\rho \geq \cos\alpha$, then $u_{3,4} = \frac{\cos\alpha}{\rho} \pm i\sqrt{1 - \left(\frac{\cos\alpha}{\rho}\right)^2}$ (here $|u_3| = |u_4| = 1$).

Note that Case 4 includes Cases 2 and 3 (for $\alpha = \frac{\pi}{2}$, respectively, $\alpha = 0$).

Case 5. $z_1 = tz_2$ ($t \in \mathbb{R}$, $z_2 \neq 0$). This case is generalization of cases $z_1 = 0 \neq z_2$, $z_1 + z_2 = 0$, $z_1 \neq 0$ and $z_1 = z_2 \neq 0$.
Denote $P\,(u) = \overline{z_1 z_2}u^4 - (\overline{z_1} + \overline{z_2})\, u^3 + (z_1 + z_2)\, u - z_1 z_2$.

Denoting $z_2 = z$ we have:

$$P(u) = t\overline{z}^2 u^4 - (1+t)\,\overline{z}u^3 + (1+t)\,zu - tz^2$$
$$= t\overline{z}^2\left(u^4 - \frac{z^4}{|z|^4}\right) - (1+t)\,\overline{z}u\left(u^2 - \frac{z^2}{|z|^2}\right).$$
$$P(u) = \overline{z}\left(u - \frac{z}{|z|}\right)\left(u + \frac{z}{|z|}\right)\left(t\overline{z}u^2 - (1+t)\,u + tz\right)$$

For $t = 0$ the roots of P are $0, \pm\frac{z}{|z|}$.
Let $t \neq 0$. Besides $\pm\frac{z}{|z|}$ there are two roots, which have modulus 1 if and only if $|z| \geq \left|\frac{1+t}{2t}\right|$.

2.2 Exterior Problem

Given $z_1, z_2 \in \mathbb{C} \setminus \overline{\mathbb{D}}$, find the point $u \in \partial\mathbb{D}$ such that the sum $|z_1 - u| + |z_2 - u|$ is minimal.

Lemma 2.6 *If the segment* $[z_1, z_2]$ *does not intersect with* $\partial\mathbb{D}$*, the point* u *is given as a solution of the equation*

$$\overline{z_1 z_2}u^4 - (\overline{z_1} + \overline{z_2})u^3 + (z_1 + z_2)u - z_1 z_2 = 0.$$

Remark 2.7 The above equation coincides with the Eq. (1.3) for the "interior problem", since Theorem 1.1 could be proved without using the assumption $z_1, z_2 \in \mathbb{D}$.

Remark 2.8 The equation of the line joining two points z_1 and z_2 is given by

$$\frac{z_1 - z}{z_2 - z} = \frac{\overline{z_1} - \overline{z}}{\overline{z_2} - \overline{z}}. \tag{2.7}$$

Then, the distance from the origin to this line is

$$\frac{|\overline{z_1}z_2 - z_1\overline{z_2}|}{2|z_1 - z_2|}.$$

Therefore, if two points z_1, z_2 satisfy $\dfrac{|\overline{z_1}z_2 - z_1\overline{z_2}|}{2|z_1 - z_2|} \leq 1$, the line (2.7) intersects with the unit circle, and the triangular ratio metric $s_{\mathbb{C}\setminus\overline{\mathbb{D}}}(z_1, z_2) = 1$.

Lemma 2.9 *The boundary of* $B_s(z, t) = \{w \in \mathbb{D} : s_{\mathbb{D}}(z, w) < t\}$ *is included in an algebraic curve.*

Proof Without loss of generality, we may assume that the center point $z =: c$ is on the positive real axis. Then,

$$\begin{aligned} s_{\mathbb{D}}(c,w) &= \sup_{\zeta\in\partial\mathbb{D}} \frac{|c-w|}{|c-\zeta|+|\zeta-w|} \\ &= \frac{|c-w|}{|2-\overline{u}c-u\overline{w}|} \quad \text{(from Remark 2.2)}, \end{aligned} \tag{2.8}$$

where u is a minimizing root of the equation

$$U_c(w) = c\overline{w}u^4 - (c+\overline{w})u^3 + (c+w)u - cw = 0. \tag{2.9}$$

Moreover, $B_s(0,t) = \{|w| < \frac{2t}{1+t}\}$ (resp. $B_s(c,0) = \{c\}$) holds for $c = 0$ (resp. $t = 0$), and $B_s(c,t) = \{0\}$ holds if and only if $c = 0$ and $t = 0$. Therefore, we may assume that $c \neq 0$, $t \neq 0$ and $w \not\equiv 0$.

Now, consider the following system of equations $s_{\mathbb{D}}(c,w) = t$ and $U_c(w) = 0$, i.e,

$$S_{c,t}(w) = t^2|2-\overline{u}c-u\overline{w}|^2 - |c-w|^2 = 0 \quad \text{and} \quad U_c(w) = 0. \tag{2.10}$$

The above two equations have a common root if and only if both of the polynomials $S_{c,t}(w)$ and $U_c(w)$ have non-zero leading coefficient with respect to u variable and the resultant satisfies $\text{resultant}_u(S_{c,t}, U_c) = 0$. Using the "resultant" command of the Risa/Asir software, we have

$$\text{resultant}_u(S_{c,t}, U_c) = cw\overline{w}\cdot\mathcal{B}_{c,t}(w),$$

where

$$\begin{aligned} &\mathcal{B}_{c,t}(w) \\ &\quad= (\overline{w}c-1)(wc-1)\big((c^2+w\overline{w}-2)^2 - 4(\overline{w}c-1)(wc-1)\big)^2 t^8 \\ &\qquad - (c-w)(c-\overline{w})\big(4\overline{w}wc^8 - 3(w+\overline{w})c^7 - 2(2\overline{w}^2w^2+2\overline{w}w-1)c^6 \\ &\qquad - (w+\overline{w})(13w\overline{w}+2)c^5 - 2(2\overline{w}^3w^3 - (36\overline{w}^2+10)w^2 - 27\overline{w}w \\ &\qquad - 10\overline{w}^2 - 4)c^4 - (w+\overline{w})(13\overline{w}^2w^2+92\overline{w}w+32)c^3 \\ &\qquad + 2(w\overline{w}(2\overline{w}^3w^3 - 2\overline{w}^2w^2 + 27\overline{w}w + 48) + 2(5w\overline{w}+2)(w^2+\overline{w}^2))c^2 \\ &\qquad - w\overline{w}(w+\overline{w})(3\overline{w}^2w^2+2\overline{w}w+32)c + 2w^2\overline{w}^2(w\overline{w}+4)\big)t^6 \\ &\qquad + (c-w)^2(c-\overline{w})^2\big(6\overline{w}wc^6 - 3(w+\overline{w})c^5 + (4\overline{w}^2w^2+16\overline{w}w+1)c^4 \\ &\qquad - 2(w+\overline{w})(13w\overline{w}+5)c^3 + (6\overline{w}^3w^3 + (16\overline{w}^2+1)w^2 + 52\overline{w}w + \overline{w}^2)c^2 \\ &\qquad - w\overline{w}(w+\overline{w})(3w\overline{w}+10)c + \overline{w}^2w^2\big)t^4 \\ &\qquad - c(c-w)^3(c-\overline{w})^3\big(4w\overline{w}c(c^2+w\overline{w}+3) - (c^2+w\overline{w})(w+\overline{w})\big)t^2 \\ &\qquad + c^2w\overline{w}(c-w)^4(c-\overline{w})^4 . \end{aligned}$$

Moreover, we can check that

$$\mathcal{B}_{c,0}(w) = |w|^2c^2|c-w|^8$$

and

$$\mathcal{B}_{0,t}(w) = |w|^4 t^4 \big((t-1)^2|w|^2 - 4t^2\big)\big((t+1)^2|w|^2 - 4t^2\big).$$

Hence, the boundary of $B_s(c,t)$ is included in the algebraic curve defined by the equation $\mathcal{B}_{c,t}(w) = 0$. □

Remark 2.10 The algebraic curve $\{w : \mathcal{B}(w) = 0\}$ does not coincide with the boundary $\partial B_s(c,t)$. There is an "extra" part of the curve since the Eq. (2.9) contains extraneous solutions.

The analytic formula in Corollary 1.2 for the triangular ratio metric $s_{\mathbb{D}}(z_1, z_2)$ is not very practical. Therefore, we next give an algorithm based on Theorem 1.1 for the evaluation of the numerical values.

Algorithm. We next give a Mathematica algorithm for computing $s_{\mathbb{D}}(x, y)$ for given points $x, y \in \mathbb{D}$. Figure 4 was drawn with the help of this algorithm.

```
sD[x_, y_] := Module[{u, sol, mySol, tmp = 2*Sqrt[2]},
sol = Solve[ Conjugate[ x*y] u^4 -  Conjugate[x + y] u^3 +
   (x + y) u - x*y == 0, {u}];
mySol = u /. sol;
Do[If[Abs[Abs[mySol[[i]] ] - 1] < 10^(-12),
  tmp = Min[tmp,
   Abs[mySol[[i]] - x] + Abs[mySol[[i]] - y]]],
     {i, 1, Length[mySol]}];
 Abs[x - y]/tmp] ;
```

One can also use numerical methods to compute $s_{\mathbb{D}}$, see [6].

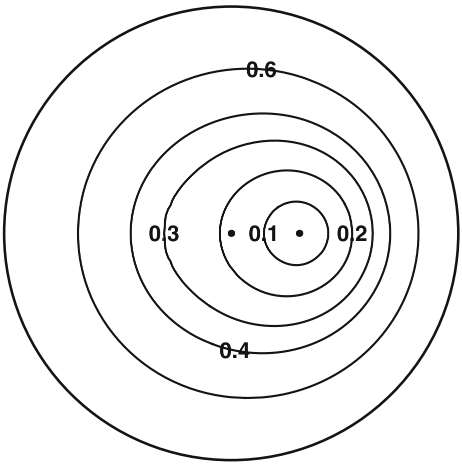

Fig. 4 Level sets $\{x + iy : s_{\mathbb{D}}(0.3, x + iy) = t\}$ for $t = 0.1, 0.2, 0.3, 0.4, 0.6$ and the unit circle. By Lemma 2.9, these level sets are contained in an algebraic curve. These level sets are drawn with the help of the Mathematica algorithm below

3 Geometric Approach to the Ptolemy–Alhazen Problem

In this section, the unimodular roots of Eq. (1.3) are characterized as points of intersection of a conic section and the unit circle, then n such roots are studied, where $n = 4$ in the case of the exterior problem and $n = 2$ in the case of the interior problem. We describe the construction of the conic section mentioned above. Except in the cases where $0, z_1, z_2$ are collinear or $|z_1| = |z_2|$, the construction cannot be carried out as ruler-and-compass construction. Neumann [19] proved that Alhazen's interior problem for points z_1, z_2 is solvable by ruler and compass only for $(\mathrm{Re} z_1, \mathrm{Im} z_1, \mathrm{Re} z_2, \mathrm{Im} z_2)$ belonging to a null subset of $\mathbb{R}^4$, in the sense of Lebesgue measure.

We characterize algebraically condition (1.1) without assuming that $z_1, z_2 \in \mathbb{D}$, or $z_1, z_2 \in \mathbb{C} \setminus \overline{\mathbb{D}}$, or $u \in \partial\mathbb{D}$.

Lemma 3.1 *Let $z_1, z_2 \in \mathbb{C}$ and $u \in \mathbb{C}^* \setminus \{z_k : k = 1, 2\}$. The following are equivalent:*

(i) $\measuredangle(z_1, u, 0) = \measuredangle(0, u, z_2)$.
(ii) $\frac{u^2}{(u-z_1)(u-z_2)} = \frac{\overline{u}^2}{(\overline{u}-\overline{z_1})(\overline{u}-\overline{z_2})}$ *and* $\frac{u^2}{(u-z_1)(u-z_2)} + \frac{\overline{u}^2}{(\overline{u}-\overline{z_1})(\overline{u}-\overline{z_2})} > 0$;
(iii)
$$\overline{z_1 z_2} u^2 - (\overline{z_1} + \overline{z_2})\, \overline{u} u^2 + (z_1 + z_2)\, \overline{u}^2 u - z_1 z_2 \overline{u}^2 = 0 \tag{3.1}$$

and

$$\overline{z_1 z_2} u^2 - (\overline{z_1} + \overline{z_2})\, \overline{u} u^2 - (z_1 + z_2)\, \overline{u}^2 u + z_1 z_2 \overline{u}^2 + 2u^2 \overline{u}^2 > 0. \tag{3.2}$$

Proof Let $u \in \mathbb{C}^* \setminus \{z_k : k = 1, 2\}$. Clearly, $\measuredangle(z_1, u, 0) = \arg \frac{u}{u-z_1}$ and $\measuredangle(0, u, z_2) = \arg \frac{u-z_2}{u}$. Denoting $v := \frac{u}{u-z_1} : \frac{u-z_2}{u}$, we see that $\measuredangle(z_1, u, 0) = \measuredangle(0, u, z_2)$ if and only if v satisfies both $v = \overline{v}$ and $v + \overline{v} > 0$, i.e. if and only if (ii) holds.

We have $v = \overline{v}$ (respectively, $v + \overline{v} > 0$) if and only if (3.1) (respectively, (3.2)) holds, therefore (ii) and (iii) are equivalent.

In the special case $z_1 = z_2 = 0$ ($z_1 = z_2 \neq 0$) (i), (ii) and (iii) are satisfied whenever $u \in \mathbb{C}^*$ (respectively, if and only if $u = \lambda z_1$ for some real number $\lambda \neq 0, 1$). □

Remark 3.2 Let $u \in \mathbb{C}^* \setminus \{z_k : k = 1, 2\}$. If

$$\frac{u^2}{(u-z_1)(u-z_2)} = \frac{\overline{u}^2}{(\overline{u}-\overline{z_1})(\overline{u}-\overline{z_2})} \text{ and } \frac{u^2}{(u-z_1)(u-z_2)} + \frac{\overline{u}^2}{(\overline{u}-\overline{z_1})(\overline{u}-\overline{z_2})} < 0,$$

then $|\measuredangle(z_1, u, 0) - \measuredangle(0, u, z_2)| = \pi$. The converse also holds.

Consider the interior problem, with $z_1, z_2 \in \mathbb{D}$ and $u \in \partial\mathbb{D}$. The unit circle is exterior to the circles of diameters $[0, z_1]$, $[0, z_2]$. An elementary geometric argument shows that $-\frac{\pi}{2} < \measuredangle(z_1, u, 0) < \frac{\pi}{2}$ and $-\frac{\pi}{2} < \measuredangle(0, u, z_2) < \frac{\pi}{2}$, therefore $|\measuredangle(z_1, u, 0) - \measuredangle(0, u, z_2)| \neq \pi$. In this case (3.1) implies $\measuredangle(z_1, u, 0) = \measuredangle(0, u, z_2)$.

The Eq. (3.1) defines a curve passing through $0, z_1$ and z_2, that is a cubic if $z_1 + z_2 \neq 0$, respectively, a conic section if $z_1 + z_2 = 0$ with $z_1, z_2 \in \mathbb{C}^*$. Then, under the

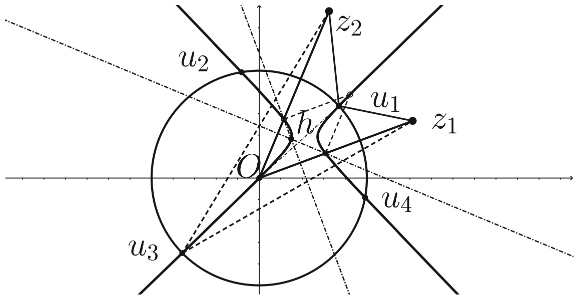

Fig. 5 The exterior problem. Intersection of the conic (3.3) with the unit circle

inversion with respect to the unit circle, the image of the curve given by (3.1) has the equation

$$\overline{z_1 z_2} u^2 - (\overline{z_1} + \overline{z_2})\, u + (z_1 + z_2)\, \overline{u} - z_1 z_2 \overline{u}^2 = 0. \tag{3.3}$$

This is a conic section, that degenerates to a line if $z_1 z_2 = 0$ with z_1, z_2 not both zero. The points of intersection of the unit circle with the conic section (3.3) are shown in Fig. 5.

Remark 3.3 If $u \in \partial\mathbb{D}$, then (3.1) (respectively, (3.3)) holds if and only if

$$\overline{z_1 z_2} u^2 - (\overline{z_1} + \overline{z_2})\, u + (z_1 + z_2)\, \frac{1}{u} - z_1 z_2 \frac{1}{u^2} = 0.$$

The Eqs. (3.3), (3.1) and (1.3) have the same unimodular roots.

Lemma 3.4 *Let $z_1, z_2 \in \mathbb{C}^*$. The conic section Γ given by (3.3) has the center $c = \frac{1}{2}\left(\frac{1}{\overline{z_1}} + \frac{1}{\overline{z_2}}\right)$ and it passes through 0, $\frac{1}{\overline{z_1}}$, $\frac{1}{\overline{z_2}}$, $\frac{1}{\overline{z_1}} + \frac{1}{\overline{z_2}}$. If $|z_1| = |z_2|$ or $|\arg z_1 - \arg z_2| \in \{0, \pi\}$, then Γ consists of the parallels d_1, d_2 through c to the bisectors (interior, respectively, exterior) of the angle $\measuredangle(z_1, 0, z_2)$. In the other cases, Γ is an equilateral hyperbola having the asymptotes d_1 and d_2.*

Proof The Eq. (3.3) is equivalent to

$$\operatorname{Im}\left(\overline{z_1 z_2} u \left(\frac{1}{\overline{z_1}} + \frac{1}{\overline{z_2}} - u\right)\right) = 0. \tag{3.4}$$

The curve Γ passes through the points 0 and $2c = \frac{1}{\overline{z_1}} + \frac{1}{\overline{z_2}}$. If u satisfies (3.4), then $2c - u$ also satisfies (3.4); therefore, Γ has the center c. Since z_1 and z_2 are on the cubic curve given by (3.1), Γ passes through $\frac{1}{\overline{z_1}}$ and $\frac{1}{\overline{z_2}}$. The conic section Γ is a pair of lines if and only if Γ passes through its center. For $u = \frac{1}{2}\left(\frac{1}{\overline{z_1}} + \frac{1}{\overline{z_2}}\right)$ we have

$$\operatorname{Im}\left(\overline{z_1 z_2} u \left(\frac{1}{\overline{z_1}} + \frac{1}{\overline{z_2}} - u\right)\right) = \frac{1}{4} \operatorname{Im}\left(\frac{\overline{z_1}}{\overline{z_2}} + \frac{\overline{z_2}}{\overline{z_1}}\right),$$

therefore Γ is a pair of lines if and only if $\frac{\overline{z_1}}{z_2}+\frac{\overline{z_2}}{z_1}\in\mathbb{R}$. The following conditions are equivalent:

(1) $\frac{\overline{z_1}}{z_2}+\frac{\overline{z_2}}{z_1}\in\mathbb{R}$;
(2) $\frac{z_2}{z_1}\in\mathbb{R}$ or $\left|\frac{z_2}{z_1}\right|=1$;
(3) $|\arg z_1-\arg z_2|\in\{0,\pi\}$ or $|z_1|=|z_2|$.

Denote $u=x+iy$. Using a rotation around the origin and a reflection we may assume that $\arg z_2=-\arg z_1=:\alpha$, where $0\le\alpha\le\frac{\pi}{2}$. In this case, the equation of Γ is

$$\left(x-\frac{|z_1|+|z_2|}{2\,|z_1z_2|}\cos\alpha\right)\left(y-\frac{|z_2|-|z_1|}{2\,|z_1z_2|}\sin\alpha\right)=\frac{|z_2|^2-|z_1|^2}{8\,|z_1z_2|^2}\sin 2\alpha\,. \tag{3.5}$$

The Eq. (3.5) shows that Γ is the pair of lines d_1, d_2 if $|z_1|=|z_2|$ or $\sin 2\alpha=0$; otherwise, Γ is an equilateral hyperbola having the asymptotes d_1 and d_2. □

Lemma 3.5 (Sylvester's theorem) *In any triangle with vertices z_1,z_2,z_3, the orthocenter z_H and the circumcenter z_C satisfy the identity $z_H+2z_C=z_1+z_2+z_3$.*

Proof Let z_G be the centroid of the triangle. It is well known that $z_G=\frac{z_1+z_2+z_3}{3}$. By Euler's straightline theorem, $z_H-z_G=2(z_G-z_C)$. Then $z_H+2z_C=3z_G=z_1+z_2+z_3$. □

Lemma 3.6 *Let $z_1,z_2\in\mathbb{C}^*$. The orthocenter of the triangle with vertices $0,\frac{1}{\overline{z_1}},\frac{1}{\overline{z_2}}$ belongs to the conic section given by Eq. (3.3).*

Proof Consider a triangle with vertices z_1,z_2,z_3 and denote by z_H and z_C the orthocenter and the circumcenter, respectively. By Sylvester's theorem, Lemma 3.5, $z_H=z_1+z_2+z_3-2z_C$.

But

$$z_C=\det\begin{pmatrix}1&1&1\\ z_1&z_2&z_3\\ |z_1|^2&|z_2|^2&|z_3|^2\end{pmatrix}:\det\begin{pmatrix}1&1&1\\ z_1&z_2&z_3\\ \overline{z_1}&\overline{z_2}&\overline{z_3}\end{pmatrix}.$$

If $z_3=0$, then $z_C=\frac{z_1z_2(\overline{z_2}-\overline{z_1})}{z_1\overline{z_2}-\overline{z_1}z_2}$, hence

$$z_H=\frac{(z_1-z_2)\,(z_1\overline{z_2}+\overline{z_1}z_2)}{z_1\overline{z_2}-\overline{z_1}z_2}.$$

Let h be the orthocenter of the triangle with vertices $0,\frac{1}{\overline{z_1}},\frac{1}{\overline{z_2}}$. The above formula implies

$$h=\frac{\overline{z_2}-\overline{z_1}}{\overline{z_1z_2}}\,\frac{z_1\overline{z_2}+\overline{z_1}z_2}{z_1\overline{z_2}-\overline{z_1}z_2}. \tag{3.6}$$

Let $f(u):=\overline{z_1z_2}u^2-(\overline{z_1}+\overline{z_2})\,u+(z_1+z_2)\,\overline{u}-z_1z_2\overline{u}^2$. Then $f(u)=2i\,\mathrm{Im}\left(\overline{z_1z_2}u^2-(\overline{z_1}+\overline{z_2})\,u\right)$. Since $\overline{z_1z_2}h-(\overline{z_1}+\overline{z_2})=\frac{2\overline{z_1z_2}(z_2-z_1)}{z_1\overline{z_2}-\overline{z_1}z_2}$, it follows that

$$\overline{z_1z_2}h^2-(\overline{z_1}+\overline{z_2})\,h=\frac{-16\,|z_2-z_1|^2}{|z_1\overline{z_2}-\overline{z_1}z_2|^4}\mathrm{Re}\,(z_1\overline{z_2})\,\mathrm{Im}^2\,(z_1\overline{z_2})$$

is a real number, hence $f(h)=0$. □

Let $z_1, z_2 \in \mathbb{C}^*$ be such that $|z_1| \neq |z_2|$ and $|\arg z_1 - \arg z_2| \notin \{0, \pi\}$. Let h be given by (3.6). Note that $h - \left(\frac{1}{z_1} + \frac{1}{z_2}\right) = \frac{2(z_2 - z_1)}{z_1\overline{z_2} - \overline{z_1}z_2} \neq 0$. If $h \notin \left\{0, \frac{1}{z_1}, \frac{1}{z_2}\right\}$ then the hyperbola Γ passing through the five points $0, \frac{1}{z_1}, \frac{1}{z_2}, \frac{1}{z_1} + \frac{1}{z_2}, h$ can be constructed using a mathematical software.

In the cases where $h \in \left\{0, \frac{1}{z_1}, \frac{1}{z_2}\right\}$, we choose a vertex of the hyperbola Γ as the fifth point needed to construct Γ. The vertices of the equilateral hyperbola Γ are the intersections of Γ with the line passing through the center of the hyperbola, with the slope $m = 1$ if $|z_1| > |z_2|$, respectively, $m = -1$ if $|z_1| < |z_2|$. Let $\alpha := \frac{\arg z_2 - \arg z_1}{2}$. Using (3.5) it follows that the distance d between a vertex and the center of Γ is $d = \frac{\sqrt{\left||z_1|^2 - |z_2|^2\right|}}{2|z_1 z_2|}\sqrt{\sin 2\alpha}$.

If $h = 0$ we have $\alpha = \frac{\pi}{4}$ and $d = \frac{1}{2}\sqrt{\left|\left|\frac{1}{z_2}\right|^2 - \left|\frac{1}{z_1}\right|^2\right|}$. Assume that $h = \frac{1}{z_1}$, the case $h = \frac{1}{z_2}$ being similar. Then $|z_2| = |z_1| \cos 2\alpha < |z_1|$ and $\left||z_1|^2 - |z_2|^2\right| = |z_1 - z_2|^2$, therefore $d = \frac{1}{2}\left|\frac{1}{z_2} - \frac{1}{z_1}\right| \sqrt{\sin 2\alpha}$. Let z_3 be the orthogonal projection of $\frac{1}{z_1}$ on the line joining $\frac{1}{z_2}$ to the origin. Then $d = \frac{1}{2}\sqrt{\left|\frac{1}{z_2} - \frac{1}{z_1}\right| \cdot \left|\frac{1}{z_2} - z_3\right|}$. We see that a vertex of Γ can be constructed with ruler and compass if $h \in \left\{0, \frac{1}{z_1}, \frac{1}{z_2}\right\}$.

Remark 3.7 Being symmetric with respect to the center of Γ, $\frac{1}{z_1}$ and $\frac{1}{z_2}$ belong to distinct branches of Γ, each branch being divided by $\frac{1}{z_1}$ or $\frac{1}{z_2}$ into two arcs. If $z_k \in \mathbb{C}\backslash\overline{\mathbb{D}}$, $k \in \{1, 2\}$, then each of these arcs joins $\frac{1}{z_k}$, that is in the unit disk, with some point exterior to the unit disk; therefore, it intersects the unit circle. It follows that, in the case of the exterior problem, Γ intersects the unit circle at four distinct points.

In the following, we identify the points of intersection of the conic section Γ given by (3.3) with the unit circle. After finding the points $u \in \partial\mathbb{D} \cap \Gamma$, it is easy to select among these the points u for which (1.1) holds, respectively, for which $|u - z_1| + |u - z_2|$ attains its minimum or its maximum on $\partial\mathbb{D}$.

First assume that Γ is a pair of lines d_1, d_2, parallel to the interior bisector and to the exterior bisector of the angle $\measuredangle(z_1, 0, z_2)$, respectively. Let $\alpha = \frac{1}{2}|\arg z_2 - \arg z_1|$. Then $\alpha \in \left\{0, \frac{\pi}{2}\right\}$ or $|z_1| = |z_2|$. The distances from the origin to d_1 and d_2 are $\delta_1 = \frac{||z_2| - |z_1||}{2|z_1 z_2|}\sin\alpha$ and $\delta_2 = \frac{|z_1| + |z_2|}{2|z_1 z_2|}\cos\alpha$. Then, Γ intersects the unit circle at four distinct points in the following cases: (i) $z_1, z_2 \in \mathbb{C} \setminus \mathbb{D}$; (ii) $z_1, z_2 \in \mathbb{D}$ with $\frac{1}{2}\left|\frac{1}{|z_1|} - \frac{1}{|z_2|}\right| < 1$ or with $|z_1| = |z_2| > \cos\alpha$. In the other cases for $z_1, z_2 \in \mathbb{D}$ the intersection of Γ with the unit circle consists of two distinct points.

Proposition 3.8 *If the conic section* Γ *given by* (3.3) *is a hyperbola, then the intersection of* Γ *with the unit circle consists of*

(i) *four distinct points if* $z_1, z_2 \in \mathbb{C} \setminus \mathbb{D}$*, one in the interior of each angle determined by the lines that pass through the origin and* z_1*, respectively,* z_2*;*

(ii) *at least two distinct points if* $z_1, z_2 \in \mathbb{D}$*, one in the interior of the angle determined by the rays passing starting at the origin and passing through* z_1*, respectively,* z_2 *and the other in the interior of the opposite angle.*

Proof The intersection of Γ with the unit circle consists of the points $u = e^{it}$, $t \in (-\pi, \pi]$ satisfying

$$\operatorname{Im}\left(\overline{z_1 z_2} e^{i2t} - (z_1 + z_2)\, e^{-it}\right) = 0.$$

Let $z_1, z_2 \in \mathbb{C}^*$. There are at most four points of intersection of Γ and the unit circle, since these are the roots of the quartic Eq. (1.3).

Using a rotation around the origin and a change of orientation, we may assume that $\arg z_2 = -\arg z_1 =: \alpha$, where $0 \le \alpha \le \frac{\pi}{2}$. The above equation is equivalent to

$$g\,(t) := |z_1 z_2| \sin 2t - |z_1| \sin (t + \alpha) - |z_2| \sin (t - \alpha) = 0. \tag{3.7}$$

We have

$$\begin{aligned}
&g\,(-\pi) = g\,(\pi) = -g\,(0) = (|z_1| - |z_2|) \sin \alpha,\\
&g\,(\alpha - \pi) = |z_1|\,(|z_2| + 1) \sin 2\alpha\,, \quad g\,(-\alpha) = |z_2|\,(1 - |z_1|) \sin 2\alpha,\\
&g\,(\alpha) = |z_1|\,(|z_2| - 1) \sin 2\alpha\,, \quad g\,(\pi - \alpha) = -\,|z_2|\,(|z_1| + 1) \sin 2\alpha.
\end{aligned}$$

Consider the cases where Γ is a hyperbola, i.e., $0 < \alpha < \frac{\pi}{2}$. Clearly, $-\pi < \alpha - \pi < -\alpha < 0 < \alpha < \pi - \alpha < \pi$. We have $g\,(\pi - \alpha) < 0 < g\,(\alpha - \pi)$, while $g\,(-\pi) = g\,(\pi) = -g\,(0)$ has the same sign as $|z_1| - |z_2|$.

(i) Assume that $z_1, z_2 \in \mathbb{C} \setminus \mathbb{D}$. Then $g\,(-\alpha) < 0$ and $g\,(\alpha) > 0$.
If $|z_1| < |z_2|$, then $g\,(-\pi) < 0 < g\,(\alpha - \pi) > 0$, $g\,(-\alpha) < 0 < g\,(0)$ and $g(\alpha) > 0 > g\,(\pi - \alpha)$. Since g is continuous on $\mathbb{R}$, Eq. (3.7) has at least one root in each of the open intervals $(-\pi, \alpha - \pi)$, $(\alpha - \pi, -\alpha)$, $(-\alpha, 0)$ and $(\alpha, \pi - \alpha)$.
If $|z_2| < |z_1|$, then $g\,(\alpha - \pi) > 0 > g\,(-\alpha)$, $g\,(0) < 0 < g(\alpha)$ and $g\,(\pi - \alpha) < 0 < g\,(\pi)$. The Eq. (3.7) has at least one root in each of the open intervals $(\alpha - \pi, -\alpha)$, $(0, \alpha)$, $(\alpha, \pi - \alpha)$ and $(\pi - \alpha, \pi)$.

(ii) Now assume that $z_1, z_2 \in \mathbb{D}$. Then $g\,(-\alpha) > 0$ and $g\,(\alpha) < 0$.
If $|z_1| < |z_2|$, then $g\,(-\pi) < 0 < g\,(\alpha - \pi)$ and $g\,(0) > 0 > g(\alpha)$. Since g is continuous on $\mathbb{R}$, Eq. (3.7) has at least one root in each of the open intervals $(-\pi, \alpha - \pi)$ and $(0, \alpha)$.
If $|z_1| > |z_2|$, then $g\,(0) > 0 > g(\alpha)$ and $g\,(\pi - \alpha) < 0 < g\,(\pi)$. The Eq. (3.7) has at least one root in each of the open intervals $(0, \alpha)$ and $(\pi - \alpha, \pi)$.

□

Corollary 3.9 *The Eq.* (1.3) *has four distinct unimodular roots in the case of the exterior problem and has at least two distinct unimodular roots in the case of the interior problem.*

4 Remarks on the Roots of the Equation (1.3)

In this section, we study the number of the unimodular roots of the Eq. (1.3) (i.e., the roots lying on the unit circle) and their multiplicities. Denote $P(u) = \overline{z_1 z_2} u^4 -$

$(\overline{z_1}+\overline{z_2})u^3+(z_1+z_2)u-z_1z_2$. If either $z_1=0$ or $z_2=0$ then the cubic Eq. (1.3) $P(u)=0$ has a root $u=0$ and two simple roots on the unit circle.

We will assume in the following that $z_1\neq 0$ and $z_2\neq 0$. As we observed in Sect. 2, the quartic polynomial P is self-inversive. Then, P has an even number of zeros on the unit circle, each zero being counted as many times as its multiplicity. According to Lemma 2.4, P has at least two unimodular zeros, distinct or not, that is P has four or two unimodular zeros. There is a rich literature dealing with the location of zeros of a complex self-inversive polynomial with respect to the unit circle [5,7,8,15–17].

Lemma 4.1 *$P(u)=\overline{z_1z_2}u^4-(\overline{z_1}+\overline{z_2})u^3+(z_1+z_2)u-z_1z_2$ cannot have two double zeros on the unit circle.*

Proof Assume that P has two double zeros a and b on the unit circle, $P(u)=\overline{z_1z_2}(z-a)^2(z-b)^2$ $(a,b\in\partial\mathbb{D},\ a\neq b)$. Since the coefficient of u^2 in $P(u)$ vanishes,

$$a^2+4ab+b^2=\big(a+(2-\sqrt{3})b\big)\big(a+(2+\sqrt{3})b\big)=0\,.$$

This contradicts the assumption $|a|=|b|=1$. □

Similarly, we rule out another case.

Lemma 4.2 *For $P(u)=\overline{z_1z_2}u^4-(\overline{z_1}+\overline{z_2})u^3+(z_1+z_2)u-z_1z_2$ it is not possible to have a double zero on the unit circle and two zeros not on the unit circle.*

Proof Assume that P has a double zero a with $|a|=1$ and the zeros $b\neq\frac{1}{\overline{b}}$. Then $P(u)=\overline{z_1z_2}(z-a)^2(z-b)\left(z-\frac{1}{\overline{b}}\right)$. The coefficient of u^2 in $P(u)$ vanishes,

$$a^2+\frac{b}{\overline{b}}+2a\left(b+\frac{1}{\overline{b}}\right)=0.$$

We have

$$\left|b+\frac{1}{\overline{b}}\right|^2=\left(b+\frac{1}{\overline{b}}\right)\left(\overline{b}+\frac{1}{b}\right)=2+|b|^2+\frac{1}{|b|^2}>4.$$

Then $2\geq\left|a^2+\frac{b}{\overline{b}}\right|=\left|2a\left(b+\frac{1}{\overline{b}}\right)\right|>4$, a contradiction. □

Lemma 4.3 *If $P(u)=\overline{z_1z_2}u^4-(\overline{z_1}+\overline{z_2})u^3+(z_1+z_2)u-z_1z_2$ has a triple zero a and a simple zero b, then $b=-a$, with a and b lying on the unit circle and $|z_1+z_2|=2\,|z_1z_2|$.*

Proof Assume that P has a triple zero a and a simple zero b, $P(u)=\overline{z_1z_2}(z-a)^3(z-b)$, where $a,b\in\mathbb{C},\ a\neq b$. Since P is self-inversive, $|a|=|b|=1$ and $b=\frac{1}{\overline{a}}=-a$. Also, the fact that the coefficient of u^2 in $P(u)$ vanishes already implies $a(a+b)=0$. But $\overline{z_1z_2}a^2b=-z_1z_2\neq 0$, therefore $b=-a$. Considering the coefficient of u^3 in $P(u)=\overline{z_1z_2}(u-a)^3(u+a)$, it follows that $2a\overline{z_1z_2}=\overline{z_1}+\overline{z_2}$, hence $|z_1+z_2|=2\,|z_1z_2|$. □

Example 4.4 Find the relation between z_1, z_2 such that $P(u) = \overline{z_1 z_2} u^4 - (\overline{z_1} + \overline{z_2})u^3 + (z_1 + z_2)u - z_1 z_2$ has the triple zero 1 and the simple zero (-1).

Suppose

$$P(u) = \overline{z_1 z_2}(u-1)^3(u+1) = 0\,. \tag{4.1}$$

From the constant term of (1.3) and (4.1), we have $z_1 z_2 \in \mathbb{R}$. Similarly, from the coefficient of u in (1.3) and (4.1), we have

$$z_1 + z_2 - 2z_1 z_2 = 0\,.$$

Therefore z_1 and z_2 coincide with the two solutions of $w^2 - 2pw + p = 0$, where $p = z_1 z_2 \in \mathbb{R}$ (in particular $-1 < p < 1$ for the interior problem).

In the case where $0 < p < 1$, z_1 and z_2 are complex conjugates to each other since discriminant$(w^2 - 2pw + p, w) = 4(p^2 - p) < 0$. Hence, $P(u) = z_1 \overline{z_1}(u-1)^3(u+1) = 0$, and we have

$$(2\overline{z_1} - 1)z_1 - \overline{z_1} = 2\left|z_1 - \frac{1}{2}\right| - \frac{1}{2} = 0.$$

Therefore, for z_1 on the circle $|z - \frac{1}{2}| = \frac{1}{2}$ and $z_2 = \overline{z_1}$, $P(u) = 0$ has exactly two roots 1 and -1. This case was studied in [11, Thm. 3.1]. In fact, for $z_1 = a + bi$ with $a^2 - a + b^2 = 0$, $P(u) = a(u-1)^3(u+1) = 0$.

In the case where $-1 < p < 0$, the quadratic equation $w^2 - 2pw + p = 0$ has two real roots and we have

$$P(u) = z_1 z_2 (u-1)^3(u+1).$$

Moreover, we can parametrize two foci as follows, $z_1 = t,\ z_2 = \frac{t}{2t-1}$ $(-1 < t < \sqrt{2} - 1)$.

It remains to study the following cases:

Case 1. P has four simple unimodular zeros.
Case 2. P has two simple unimodular zeros and two zeros that are not unimodular.
Case 3. P has a double unimodular zero and two simple unimodular zeros.

Proposition 4.5 *Assume that $z_1, z_2 \in \mathbb{C}^*$. Let $P(u) = \overline{z_1 z_2} u^4 - (\overline{z_1} + \overline{z_2})u^3 + (z_1 + z_2)u - z_1 z_2$. Then*

a) *P has four simple unimodular zeros if $|z_1 + z_2| < |z_1 z_2|$ and*
b) *P has exactly two unimodular zeros, that are simple, if $|z_1 + z_2| > 2\,|z_1 z_2|$.*
c) *If P has four simple unimodular zeros, then $|z_1 + z_2| < 2\,|z_1 z_2|$.*
d) *If P has exactly two unimodular zeros, that are simple, then $|z_1 + z_2| > |z_1 z_2|$.*

Proof Let f be a complex polynomial. The location of the zeros of the derivative f' of f is connected with the location of the zeros of f. Gauss–Lucas theorem [17, Thm. 6.1] shows that the zeros of the derivative f' lie within the convex hull of the set of zeros of f. In particular, if all the zeros of f lie on the unit circle, then all the zeros of f' lie in the closed unit disk (and f is self-inversive). The converse holds by a theorem of Cohn [8] stating that a complex polynomial has all its zeros on the unit circle if and only if the polynomial is self-inversive and its derivative has all its zeros in the closed unit disk.

In our case $P'(u) = 4\overline{z_1 z_2}u^3 - 3(\overline{z_1} + \overline{z_2})u^2 + (z_1 + z_2)$ and $P''(u) = 12\overline{z_1 z_2}u^2 - 6(\overline{z_1} + \overline{z_2})u$.

a) Assume that $|z_1 + z_2| < |z_1 z_2|$. Then for $u \in \partial\mathbb{D}$ we have

$$\left|4\overline{z_1 z_2}u^3\right| = 4|z_1 z_2| > 4|z_1 + z_2| \geq \left|-3(\overline{z_1} + \overline{z_2})u^2 + (z_1 + z_2)\right|$$

It follows by Rouché's theorem [22, 3.10] that the derivative P' has all its zeros in the unit disk. By Cohn's theorem cited above, P has all its four zeros on the unit circle $\partial\mathbb{D}$.

Assume that the polynomial f is self-inversive. By [7, Thm.1], the following are equivalent:

(i) all the zeros of f are simple and unimodular;

(ii) there exist a polynomial g having all its zeros in the unit disk $|z| < 1$, a non-negative integer m and a real number θ such that $f(z) = z^m g(z) + e^{i\theta} g^*(z)$ for all $z \in \mathbb{C}$. Here $g^*(z) := z^n \overline{g}\left(\frac{1}{z}\right)$, where $n = \deg Q$.

In our case, $P(u) = u^m Q(z) + e^{i\theta} Q^*(u)$ for $m = 3$, $\theta = \pi$ and $Q(u) = \overline{z_1 z_2}u^3 + (z_1 + z_2)$. The roots of Q have modulus $\sqrt[3]{\frac{|z_1+z_2|}{|z_1 z_2|}} < 1$. The implication $(ii) \Rightarrow (i)$ from [7, Thm. 1] shows that P has four simple zeros on the unit circle.

b) Now assume that $|z_1 + z_2| > 2|z_1 z_2|$. For $u \in \partial\mathbb{D}$ we have

$$\left|-3(\overline{z_1} + \overline{z_2})u^2\right| = 3|z_1 + z_2| > 4|z_1 z_2| + |z_1 + z_2| \geq \left|4\overline{z_1 z_2}u^3 + (z_1 + z_2)\right|$$

and it follows using Rouché's theorem that P' has exactly two zeros in the closed unit disk. Cohn's theorem shows that P cannot have all its zeros on $\partial\mathbb{D}$. By Lemma 2.4, P has at least two unimodular zeros; therefore, P has exactly two unimodular zeros. By Lemma 4.2, these unimodular zeros are simple.

An alternative way to prove that P has exactly two unimodular zeros is indicated below. Assume by contrary that P has four unimodular zeros. Using the Gauss–Lucas theorem two times, it follows that each of the derivatives $P'$and P'' has all its zeros in the closed unit disc $|z| \leq 1$. The zeros of P'' are 0 and $\frac{\overline{z_1}+\overline{z_2}}{2\overline{z_1 z_2}}$. Then, under the assumption $|z_1 + z_2| > 2|z_1 z_2|$, the second derivative P'' has a zero in $|z| > 1$, which is a contradiction.

c) Assume that P has four simple unimodular zeros. Then, P' has all its zeros in the closed unit disk. Given a self-inversive polynomial f, it is proved in [5, Lem.] that each unimodular zero of the derivative f' is also a zero of f. If P' has a unimodular zero a, then $P(a) = 0$; therefore, a is a zero of P of multiplicity at least 2, a contradiction. It follows that P' has all its zeros in the unit disk. By Gauss–Lucas theorem, the second derivative P'' also has all its zeros in the unit disk; therefore, $|z_1 + z_2| < 2|z_1 z_2|$.

d) Now suppose that P has exactly two simple unimodular zeros, a and b. Let c and $\frac{1}{\overline{c}}$ the other zeros of P, with $|c| < 1$. Then $P(u) = \overline{z_1 z_2}(u - a)(u - b)(u - c)$

$\left(u - \frac{1}{c}\right)$. The coefficient of u^2 in $P(u)$ vanishes; therefore,

$$ab + \frac{c}{\overline{c}} + (a+b)\left(c + \frac{1}{\overline{c}}\right) = 0,$$

and $a + b = -\frac{ab}{c+\frac{1}{\overline{c}}} - \frac{c}{|c|^2+1}$. Because $\left|\frac{ab}{c+\frac{1}{\overline{c}}}\right| = \frac{1}{\left|c+\frac{1}{\overline{c}}\right|} < \frac{1}{2}$ and $\frac{|c|}{|c|^2+1} < \frac{1}{2}$, we get $|a+b| < 1$. Considering the coefficient of u^3 in $P(u)$ we obtain $\frac{\overline{z_1}+\overline{z_2}}{\overline{z_1 z_2}} = a + b + c + \frac{1}{\overline{c}}$. Then $\frac{|z_1+z_2|}{|z_1 z_2|} \geq \left|\left|c + \frac{1}{\overline{c}}\right| - |a+b|\right| > 1$.

□

Example 4.6 Let $z_1 = (1+t)\,e^{i\alpha}$ and $z_2 = (1+t)e^{i(\alpha+t)}$, where $t > 0$ and $\alpha \in (-\pi, \pi]$. By Corollary 3.9, the Eq. (1.3) has four simple unimodular roots in this case. On the other hand, $\frac{|z_1+z_2|}{|z_1 z_2|} = (1+t)\left(1 + e^{-it}\right) \to 2$ as $t \to 0$, therefore the constant 2 in Proposition 4.5 c) cannot be replaced by a smaller constant.

We give a direct proof for the following consequence of Proposition 4.5.

Corollary 4.7 *If $P(u) = \overline{z_1 z_2} u^4 - (\overline{z_1} + \overline{z_2})u^3 + (z_1 + z_2)u - z_1 z_2$ has one double zero and two simple zeros on the unit circle, then $|z_1 z_2| \leq |z_1 + z_2| \leq 2\,|z_1 z_2|$.*

Proof Assume that P has one double unimodular zero a and two simple unimodular zeros b, c. Then $P(u) = \overline{z_1 z_2}\,(z-a)^2\,(z-b)\,(z-c)$.

The coefficient of u^2 in $P(u)$ vanishes,

$$a^2 + bc + 2a\,(b+c) = 0.$$

Considering the coefficient of u^3 in $P(u)$ we obtain $\frac{\overline{z_1}+\overline{z_2}}{\overline{z_1 z_2}} = 2a+b+c = 2a - \frac{a^2+bc}{2a} = \frac{3a^2-bc}{2a} = \frac{3}{2}a - \frac{bc}{2a}$. Then $\frac{|z_1+z_2|}{|z_1 z_2|} \leq \left|\frac{3}{2}a\right| + \left|-\frac{bc}{2a}\right| = 2$ and $\frac{|z_1+z_2|}{|z_1 z_2|} \geq \left|\left|\frac{3}{2}a\right| - \left|-\frac{bc}{2a}\right|\right| = 1$. □

Acknowledgements Open access funding provided by University of Turku (UTU) including Turku University Central Hospital. This research was begun during the Romanian-Finnish Seminar in Bucharest, Romania, June 20–24, 2016, where the authors P.H., M.M., and M.V. met. During a workshop at the Tohoku University, Sendai, Japan, in August 2016 organized by Prof. T. Sugawa, M.F., P.H., and M.V. met and had several discussions about the topic of this paper. P.H. and M.V. are indebted to Prof. Sugawa for his kind and hospitable arrangements during their visit. This work was partially supported by JSPS KAKENHI Grant Number 15K04943. The second author was supported by University of Turku Foundation and CIMO. The authors are indebted to Prof. G.D. Anderson for a number of remarks on this paper and to the referee for several useful remarks and comments.

References

1. Agrawal, A., Taguchi, Y., Ramalingam, S.: Beyond Alhazen's problem: analytical projection model for noncentral catadioptric cameras with quadric mirrors, (2011). http://www.merl.com. Accessed Jan 2017
2. Alhazen: https://en.wikipedia.org/wiki/Alhazen, 2016-06-28. Accessed Jan 2017
3. Andreescu, T., Andrica, D.: Complex Numbers...From A to Z. 2nd edn. pp. xviii+391, Birkhäuser/Springer, New York (2014). ISBN: 978-0-8176-8414-3; 978-0-8176-8415-0
4. Bach, H.: Some ray tracing problems related to circles and ellipses, tech. report, Rome air development center, Air Force Systems Command Griffiss Air Force Base, NY 13441–5700, 50 pp (1989)
5. Bonsall, F.F., Marden, M.: Zeros of self-inversive polynomials. Proc. Amer. Math. Soc. **3**, 471–475 (1952)
6. Chen, J., Hariri, P., Klén, R., Vuorinen, M.: Lipschitz conditions, triangular ratio metric, and quasiconformal maps. Ann. Acad. Sci. Fenn. **40**, 683–709 (2015). https://doi.org/10.5186/aasfm.2015.4039
7. Choo, Y., Kim, Y.J.: On the zeros of self-inversive polynomials. Int. J. Math. Analysis **7**, 187–193 (2013)
8. Cohn, A.: Über die Anzahl der Wurzeln einer algebraischen Gleichung in einem Kreise. Math. Zeit. **14**, 110–148 (1922)
9. Drexler, M., Gander, M.J.: Circular billiard (English summary). SIAM Rev. **40**, 315–323 (1998)
10. Gowers, T., Barrow-Green, J., Leader, I. (eds.): The Princeton companion to mathematics, Princeton University Press, Princeton, NJ, (2008). xxii+1034 pp. ISBN: 978-0-691-11880-2
11. Hariri, P., Klén, R., Vuorinen, M., Zhang, X.: Some remarks on the Cassinian metric, Publ. Math. Debrecen, 90 (2017), pp. 269–285. https://doi.org/10.5486/PMD.2017.7386, arXiv:1504.01923 [math.MG]
12. Hariri, P., Vuorinen, M., Zhang, X.: Inequalities and bilipschitz conditions for triangular ratio metric, Rocky Mountain J. Math., 47 (2017), pp. 1121–1148. arXiv:1411.2747 [math.MG] 21pp
13. Hästö, P.A.: A new weighted metric: the relative metric I. J. Math. Anal. Appl. **274**, 38–58 (2002)
14. Hästö, P.A.: A new weighted metric: the relative metric. II. J. Math. Anal. Appl. **301**, 336–353 (2005)
15. Lalín, M.N., Smyth, C.J.: Unimodularity of zeros of self-inversive polynomials. Acta Math. Hungar. **138**, 85–101 (2013)
16. Lalín, M.N., Smyth, C.J.: Addendum to: unimodularity of zeros of self-inversive polynomials. Acta Math. Hungar. **147**, 255–257 (2015)
17. Marden, M.: Geometry of Polynomials, 2nd ed. Mathematical Surveys 3, American Mathematical Society (Providence, R.I.), (1966)
18. Miller, A.R., Vegh, E.: Exact result for the grazing angle of specular reflection from a sphere. SIAM Rev. **35**, 472–480 (1993)
19. Neumann, P.M.: Reflections on reflection in a spherical mirror. Amer. Math. Monthly **105**, 523–528 (1998)
20. Noro, M., Shimoyama, T., Takeshima, T.: Risa/Asir symbolic computation system, http://www.math.kobe-u.ac.jp/Asir/
21. Pathak, H.K., Agarwal, R.P., Cho, Y.J.: Functions of a Complex Variable. pp. xxiv+718, CRC Press, Boca Raton, FL (2016). ISBN: 978-1-4987-2015-1
22. Sansone, G., Gerretsen, J.: Lectures on the theory of functions of a complex variable. I. Holomorphic functions, pp. xii+488 P. Noordhoff, Groningen (1960)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.