MATHEMATICAL ORIGAMI: ANOTHER VIEW OF ALHAZEN'S OPTICAL PROBLEM

Roger C. Alperin

1. Fields and Constructions

We can solve some elementary problems from geometry using origami foldings. Below are the axioms which guide the allowable constructible folds and points in $\mathbb{C}$, the field of complex numbers, starting from the labelled points 0 and 1 (see [1] for more details and references) .

(1) The line connecting two constructible points can be folded.
(2) The point of coincidence of two fold lines is a constructible point.
(3) The perpendicular bisector of the segment connecting two constructible points can be folded.
(4) The line bisecting any given constructed angle can be folded.
(5) Given a fold line l and constructed points P, Q, then whenever possible, the line through Q, which reflects P onto l, can be folded.
(6) Given fold lines l, m and constructed points P, Q, then whenever possible, any line which simultaneously reflects P onto l and Q onto m, can be folded.

The first three are the Thalian constructions which ensure that we have a field after we have a non-real complex number z. Some properties of these constructions include: reflections can be folded; the set of slopes of fold lines correspond to the points constructed on the imaginary axis together with ∞; the points which can be constructed contain the field $\mathbb{Q}(z)$ and is contained in $\mathbb{Q}(z, i)$; starting with $z = \frac{1+i\sqrt{3}}{2}$, the point i can not be constructed.

Adding the fourth axiom gives the Pythagorean field of points constructible by straightedge and dividers as discussed by Hilbert. Some properties of this field include: the (real) field of Pythagorean constructible numbers is characterized as the smallest field containing the rational numbers and is closed under $\sqrt{a^2 + b^2}$ for (a, b) a constructed point; furthermore, any algebraic conjugate of a real Pythagorean is real and has degree which is a power of 2; the number $\sqrt{1 + \sqrt{2}}$ does not belong to the field; any regular polygon which can be constructed by ruler and compass can be constructed with these axioms.

Adding the fifth axiom gives the Euclidean field of points constructible by straightedge and compass; it enables the construction of the tangents (through Q) to a parabola with focus P and directrix l. The Euclidean field is characteized as the smallest field containing the rational numbers and closed under all square roots.

Allowing axiom (6) yields the construction of the simultaneous tangents to two parabolas. Since this is related to pencils of conics we may also refer to the origami numbers as constructible by straightedge, compass and pencils. The axioms (1)-(6) are the Origami construction axioms for the complex origami numbers, $\mathcal{O}$. The Origami constructions (1)-(6) enable us to construct a real solution to a cubic equation with real coefficients in this field $\mathcal{O}$. To see this, consider the conics $(y - \frac{a}{2})^2 = 2bx, \quad 2y = x^2$. These conics have foci and directrices that are constructible using field operations involving a and b. Consider a simultaneous tangent, a line with slope μ meeting these curves at the respective points (x_0, y_0), (x_1, y_1). By differentiation we find that the slope μ of a common tangent of these two parabolas satisfies $\mu^3 + a\mu + b = 0$, and hence we can solve any cubic equation with specified real constructible $a, b \in \mathcal{O}$ for its real roots. Using the resolvant cubic of a quartic equation we can also solve fourth degree equations over $\mathcal{O}$. The field of Origami constructible numbers is characterized as the smallest subfield of the complex numbers which contains the rational numbers and is closed under all square roots, cube roots and complex conjugation.

This Origami field of numbers $\mathcal{O}$ is the same field Viete studied systematically at the beginning of 'algebraic geometry' in 1600. Of course Viete was in a sense rediscovering what was already long ago known via *neusis* constructions, to Archimedes and Apollonius, 250 BC, some of which was written down by Pappus over five hundred years later, 325 AD. One of Pappus' aims, it appears to me, was to prove all conic constructibles are the same as *neusis* constructions. This was later picked up by Alhazen in the 11^{th} century. Alhazen was interested in reconstruction of the lost works of Apollonius related to these geometrical constructions.

Moreover, the constructions using a *neusis* or marked ruler are the same constructions as using intersections of conics; both are equivalent to the origami constructions described by axioms (1)-(6). Our renewed interest in Alhazen's problem arose because of the this equivalence ([1]).

2. Harmonic Origami Numbers

In fact we can introduce yet another intermediate field of numbers and constructions using the Pythagorean axioms (1-4) together with

T. Given P and Q and a line l through P then we can simultaneously fold Q onto l and P onto the perpendicular bisector of PQ.

This gives the field of Harmonic numbers. Some of the properties of this field include: the real subfield of harmonic constructible numbers is the smallest field closed under $\sqrt{a^2+b^2}$ for a, b in the field and also closed under adjunction of any real number satisfying an irreducible cubic having three real roots with real harmonic number coefficients; trisections can be done using axiom T; $\sqrt[3]{2}$ is not harmonic constructible. Call an integer of the form 2^a3^b, a harmonic integer, so named by Phillip de Vitry (14^{th} century) in studying relations to music. Any regular polygon with n sides can be constructed using these axioms iff $\Phi(n)$ is harmonic iff n is a product of a harmonic integer and distinct primes p, so that $p-1$ is harmonic. Any regular polygon with n sides can be constructed iff $\Phi(n)$ is a harmonic integer; the real constructible harmonic points are characterised as the real Origami numbers which have Galois closure whose degree is an harmonic integer. We shall leave the details of these remarks for a later time.

3. Some Elementary Problems from Geometry

To lead up to Alhazen's problem we start with an easy geometry problem, the river crossing.

3.1. **In going from town A to town B one must cross the river L (of fixed width d) at a point z. Locate z so as to minimize the distance (by land) from A to B.**
The solution can be folded easily; fold so that the the river disappears, down the middle and then fold the line AB; this line gives the path and the bridge is the perpendicular at the river. It is clear that if we have any path from A to the river and across followed by the path to B that when we fold the river away that the two parts of the path give a length which can be shortened if it is made into a straight path.

A similar problem arises when we go from A to B on the same side of L.

3.2. **In going from town A to town B on the same side of the river, one must first stop at the river L at a point z. Locate z so as to minimize the distance from A to B.**

One can first reflect B across the line L to B_0 and then the straight line AB_0 meets the river at $z = C$ giving the stopping point for shortest distance. Morever, now it is also apparent that the two angles at C made with L are equal.

After giving the solution to these, I suggest to my students that it would be interesting to formulate and solve these problems for a circular rather than a straight river. I had in mind one point inside an inner circle and the other outside the outer concentric circle. The bridge is to be along the diameters.

3.3. **In going from town A to town B one must cross the annular river (or moat) of fixed width d at a point z. Locate z so as to minimize the distance (by land) from A to B.**
In this circular or annular problem after we locate the points $z = C$ on the outer circle and the corresponding point D on the inner circle which are the ends of the bridge then we can fold away the section between them. What remains now are two segments which give the path from A to B. In order to minimize this then it should be a straight path as in Figure 3.1. Thus we need only construct AB originally and this gives the point C. Another potential solution occurs at the point where AB meets the inner circle.

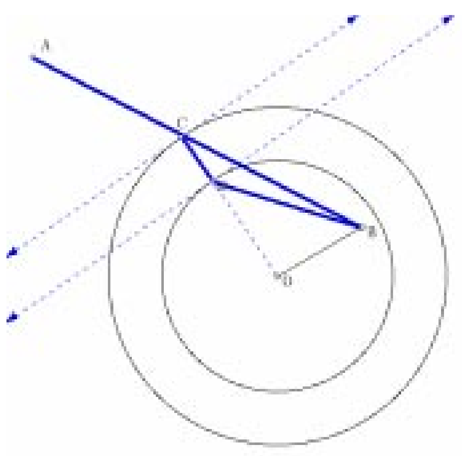

FIGURE 3.1. *Bridge over Annular River*

The last problem, I have learned, is equivalent to the classical Alhazen's Problem. Suprisingly several others have recently taken a recent look at this ancient piece of beautiful work. It is the circular analogue of the second problem. We formulate the problem as follows.

3.4. **Given points A, B exterior to a circle Γ. Locate $z = C$ on Γ so that the distance $AC + CB$ is minimized.**
At a point $z = C \in \Gamma$, we construct the tangent to Γ. Reflect B to the other side giving B_1. The sum of the two segment lengths AC and CB can be shortened by making $AC \cup CB_1$ straight. Thus the shortest distance occurs when $\angle ACB$ is bisected by the diameter through C. A similar angle restriction gives the shortest distance when then the two points are first inside the circle. (We leave to the interested reader other questions of this sort involving the other conics.)

When the points are outside the circle this is 'Alhazen's problem', a problem of optics possibly first formulated by Ptolemy. The famous al-Hasen Ibn al-Haytham (*latinized* Alhazen) lived in the 11th century (Basra, Persia and Cairo, Egypt) and wrote a most influential text on optics, later used by Renaissance scientists. According to [7], Alhazen's problem can be reduced to the construction by a *neusis*, op. cit. Lemma 1, p. 310.

Huygen's solution (1672) of Alhazen's Problem is equivalent to ibn al-Haytham's original solution [7]. Huygen's was displeased with Alhazen's solution by a classical *neusis*. The problem was solved by Huygens, reducing the problem to the intersection of the circle with an equilateral hyperbola; the solutions can then realized as solutions of a real fourth degree equation. A recent construction of this hyperbola is given in [5]. With the given points inside the circle this is the circular billiard problem considered also in [8]. One can use inversion in the given circle to convert one formulation to the other. Yet another formulation [3], asks when the chords from a point on the circle to the given points are equal.

It is important to realize that geometrical constructions of the early Renaissance eventually led to formulas for the roots of cubics and quartics involving radicals, by del Ferro in the early 16^{th} century and popularized by Cardano in the latter part of that century.

4. Alhazen's Problem-Huygen's Solution

We restate Alhazens's problem in terms of angles.

4.1. **Given points A, B exterior to a circle Γ. Locate $z = C$ on Γ so that the $\angle ACB$ is bisected by the diameter through C.**
We regard the given points $a = A$, $b = B$ as complex numbers and the circle Γ of radius 1 centered at the origin O. Using the *argument* of a complex number (or polar coordinates) as in [6, 8], $arg(\frac{a-z}{b-z})$ is the measure of the angle $\angle azb$. Thus we want z so that $\angle azO = \angle Ozb$ or equivalently $arg(\frac{a-z}{O-z}) = arg(\frac{O-z}{b-z})$; we obtain a real number by

evaluating $\frac{a-z}{z}(\frac{z}{b-z})^{-1}$. Using the fact that this number is equal to its conjugate and simplfying with $z\overline{z} = 1$ we find that

$$(ab)\overline{z}^2 - (\overline{ab})z^2 = (a+b)\overline{z} - (\overline{a+b})z.$$

This gives an equation relating the imaginary parts of two complex numbers,

$$\mathcal{I}m((ab)\overline{z}^2) = \mathcal{I}m((a+b)\overline{z}).$$

This is easily seen to be satisfied by $z \in \{O, \frac{1}{\overline{a}}, \frac{1}{\overline{b}}, \frac{1}{\overline{a}} + \frac{1}{\overline{b}}\}$. Let $p = Im(ab)$, $q = Re(ab)$, $r = Re(a+b)$, $s = Im(a+b)$. We can write this equation in terms of real coordinates using $z = x + y \cdot i$ to obtain

$$p(x^2 - y^2) - 2qxy = sx - ry.$$

This is the equation of an hyperbola. Thus, the solution to Alhazen's problem if it exists, occurs at one of the intersections of this hyperbola and the unit circle centered at the origin.

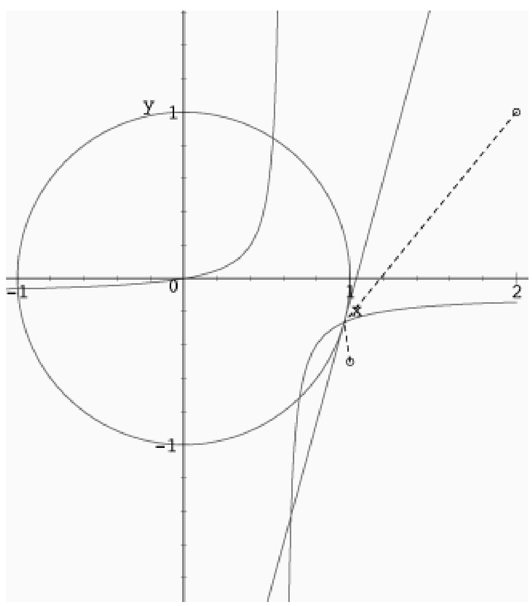

FIGURE 4.1. *Huygens' Solution: A=(2,1), B=(1,$-\frac{1}{2}$)*

Now rotate the plane about the origin by the angle so that the positive x-axis bisects the angle $\angle a0b$, hence $b = k\overline{a}$ for some $k > 0$; consequently $p = Im(ab) = 0$, thus eliminating the terms x^2, y^2 from the equation of the hyperbola above. The equation of the equilateral hyperbola now simplifies to $(x - \frac{r}{2q})(y + \frac{s}{2q}) = -\frac{rs}{4q^2}$. It follows that the center of this hyperbola is at $(\frac{r}{2q}, \frac{-s}{2q}) = (\frac{Re(a+b)}{2ab}, \frac{-Im(a+b)}{2ab}) = \frac{1}{2}(\frac{1}{\overline{a}} + \frac{1}{\overline{b}})$.

A simple example for $a = 2+i, k = \frac{1}{2}$ is displayed in Figure 4.1. The point $z \approx .9643353545 - .2646834413 \cdot i$ gives the solution to Alhazen's problem where the x coordinate solves $x^2 + (\frac{-x}{2(5x-3)})^2 - 1$ and the y coordinate is $-\sqrt{1-x^2}$. The argument $arg(\frac{a-z}{b-z})$ gives 132.06736 degrees which is twice $arg(\frac{a-z}{z})$.

The solution to Alhazen's problem is the same as the simultaneous solution to

$$A_1: \ x^2 + y^2 = 1, \ A_2: \ 2qxy + sx - ry = 0.$$

These simultaneous solutions to A_1 and A_2 also lie on any curve in the pencil $A_2 - \lambda A_1$; the pencil has matrix $\begin{pmatrix} -\lambda & q & \frac{s}{2} \\ q & -\lambda & -\frac{r}{2} \\ \frac{s}{2} & -\frac{r}{2} & \lambda \end{pmatrix}$. The determinant of this matrix pencil gives a cubic polynomial, $p(\lambda) = \lambda^3 + \frac{1}{4}(s^2 + r^2 - 4q^2)\lambda - \frac{1}{2}rqs$. Now solve the reduced cubic $p(\lambda) = 0$ using the technique allowed by axiom (6). Finding the roots to this cubic is similar to the technique of solving the resolvant cubic of a quartic. Each root gives a degenerate conic in the pencil; a pair of lines meeting at a diagonal point. One needs to use square roots to obtain the equations of these lines. By solving this cubic and the quadratics we obtain the (at most) six lines in the complete quadrilateral, and therefore the four points of intersection of the conics and the three diagonal points as in Figure 4.2 for the example above.

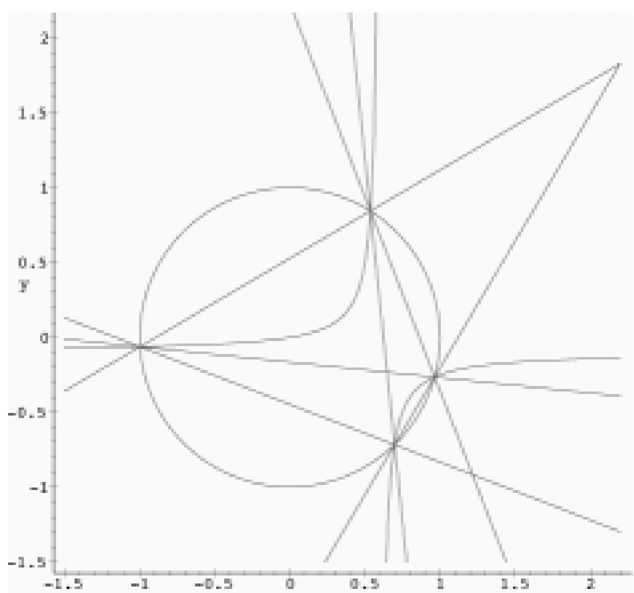

FIGURE 4.2. Complete Quadrilateral

We can also replace the Alhazen-Huygens solution as simultaneous intersection of conics by a single real quartic. We describe in the next section how pencils and origami may be used to simplify and solve the quartic.

5. An Origamic Solution to the Quartic

We assume that we are given a general real quartic. Complete the quartic to eliminate the coefficient of x^3 to obtain the quartic, say $x^4 - 2cx^2 - dx - e = 0$ for certain real c, d, e. (If $d = 0$ then we can solve the quartic using square roots so we assume $d \neq 0$.) Put $y = x^2 - c$; then the quartics' solutions are the same as the simultaneous solutions to two parabolas

$$P_1 :\ y = x^2 - c,\ P_2 :\ y^2 = dx + e + c^2.$$

The duals of these two parabolas give two conics. The dual equations are

$$C_1 :\ X^2 + 4cY^2 - 4Y = 0,\ C_2 :\ 4(e + c^2)X^2 + d^2Y^2 - 4dX = 0.$$

Duality Principles.

(1) The common tangents to C_1 and C_2 correspond to the common points of P_1 and P_2.
(2) The common tangents to P_1 and P_2 correspond to the common points of C_1 and C_2.

When a tangent line's equation is say $aX + bY + 1 = 0$, then a corresponding common point on the dual is (a, b) and conversely.

Now we consider the pencil generated by C_1 and C_2; it is $C_2 - uC_1$; when $u = 4(e+c^2)$ we obtain the parabola C_3. Next consider $C_3 - vC_2$; when $v = \frac{d^2-4cu}{d^2}$ we obtain an independant parabola C_4.

In this case, we have

$$C_3 :\ d^2vY^2 + 4uY = 4dX,\ C_4 :\ dvX^2 + 16cX = 4dY,$$

and their duals

$$P_3 :\ (ux + dy)^2 = d^3vx,\ P_4 :\ (dx + 4cy)^2 = d^2vy,$$

are parabolas since the equations have double points at infinity.

By folding tangents to C_3 and C_4 using axiom (6) we can recover the points of intersection on P_3 and P_4 by duality. We obtain the common points of P_1 and P_2 by use of the associated complete quadrilateral which has the same diagonal points as the pencil P_3 and P_4. We shall elaborate on this connection in further detail at another time.

6. Geometry and Axiomatic Origami

One interseting problem is to give elegant constructions using a given set of axioms. For example show how to solve problems from Euclidean geometry with the Euclidean axioms or Pythagorean axioms. With the Origami axioms we can fold the common intersections or tangents to two conics. Is there an elegant solution?

6.1. **Fold the intersections of a conic and a line. Fold the tangents to a conic from given point. Fold the common tangents to two conics. Fold the common points of two conics.**
One needs to give geometrical data about a conic. Specifying a conic requires five pieces of information, either points on the curve or tangent lines to the curve; this can also be replaced by focus or directrix information. For example the circle is either given by three points (and 2 circular points at infinity) or three tangents, or the center and tangents. A parabola can be given by four points (and tangent line at infinity), or a focus and directrix; hyperbola or ellipse can be specified by data from foci, center, diameter, tangents, points or asymptotes.

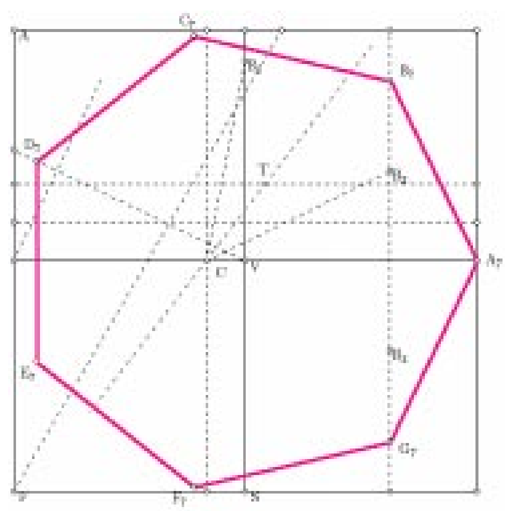

Figure 6.1. Folding the Regular Heptagon using Harmonics

6.2. **Given a square sheet of paper fold a regular n gon of largest area possible using one of the axiom systems Pythagorean, Harmonic, Euclidean, or Origami?**
To fold a heptagon using harmonic constructions. Take a piece of paper say 6 units square. Fold the corner A to point on CV using a fold through the corner F as in the diagram of Figure 6.1. This fold

meets the central vertical line at a point; the midpoint B_1 of this point with the top edge is constructed. The distance of B_1 from V is $3\sqrt{3}$ and so the hypotenuse with the edge CV of length 1 has length $\sqrt{28}$. Now trisect the angle B_1CV to get the point T as in Abe's trisection (it is the trisection with the base CB_1). Reflect B_1 to B_2 across CT. Reflect B_2 across CV to get B_3. Now fold A_7 onto the line B_2B_3 passing through V. This gives the second point of the heptagon B_7. Now reflect this across VA_7 to get the vertex G_7 of the heptagon. Fold B_7A_7 across the line VB_1 and bisect the angle with CV; construct the vertex D_7 on this bisector. Reflect across CV to get vertex E_7. Bisect the angle $\angle E_7VG_7$; construct F_7 on that bisector. Reflect across CV to get the final vertex.

As one can show, if a polygon can be folded in the square then so can the optimal one. An analysis of the cosine of the angle involved in tipping the polygon to its optimal position shows that it belongs to the field, [2]. Can *you* fold the optimal heptagon using only harmonic constructions?

References

[1] Roger C. Alperin, *A Mathematical Theory of Origami Constructions and Numbers*, New York J. Math., 6, 119-133 (2000) http://nyjm.albany.edu:8000/j/2000/6-8.html

[2] David Dureisseix, *Searching for the Optimal Polygon*, webprint (October 1997).

[3] Jack M. Elkin, *A Deceptively Easy Problem*, Math. Teacher, 194-199 (March 1965)

[4] W. Gander and D. Gruntz, *The Billiard Problem*, Maple Worksheet Shareware (July 1992) gander@inf.ethz.ch

[5] Norbert Hungerbühler, *Geometrical Aspects of the Circular Billiard*, El. Math. 47, 114-117 (1992)

[6] Peter M. Neumann, *Reflections on Reflection in a Spherical Mirror*, Am. Math. Monthly, 105, No. 6, 523-528 (1998)

[7] A. I. Sabra, *Ibn al-Haytham's Lemmas for Solving 'Alhazen's Problem'*, Archive for History of Exact Sciences,, 26, 299-324 (1982)

[8] Jörg Waldvogel, *The Problem of the Circular Billiard*, El. Math. 47, 108-113 (1992)

Department of Mathematics and Computer Science, San Jose State University, San Jose, CA 95192, email: alperin@mathcs.sjsu.edu