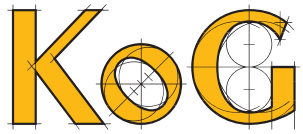
KoG

No. 18. (2014)
ISSN 1331-1611

SCIENTIFIC-PROFESSIONAL JOURNAL OF CROATIAN SOCIETY
FOR GEOMETRY AND GRAPHICS



Official publication of the Croatian Society for Geometry and Graphics publishes scientific and professional papers from the fields of geometry, applied geometry and computer graphics.

Founder and Publisher

Croatian Society for Geometry and Graphics

Design

Miroslav Ambruš-Kiš

Layout

Sonja Gorjanc, Ema Jurkin

Cover Illustration

F. V. Holi, *"Old fashioned Croatian joint"*, photography

Print

SAND d.o.o., Zagreb

URL address

http://www.hdgg.hr/kog
http://hrcak.srce.hr

Edition

150

Published annually

Guide for authors

Please, see the last page

KoG is cited in: Mathematical Reviews, MathSciNet, Zentralblatt für Mathematik

INSTRUCTIONS FOR AUTHORS

SCOPE. "KoG" publishes scientific and professional papers from the fields of geometry, applied geometry and computer graphics.

SUBMISSION. Scientific papers submitted to this journal should be written in English or German, professional papers should be written in Croatian, English or German. The papers have not been published or submitted for publication elsewere.
The manuscript with wide margines and double spaced should be sent in PDF format via e-mail to one of the editors:

Sonja Gorjanc	Ema Jurkin
sgorjanc@grad.hr	ema.jurkin@rgn.hr

The first page should contain the article title, author and coauthor names, affiliation, a short abstract in English, a list of keywords and the Mathematical subject classification.

UPON ACCEPTANCE. After the manuscript has been accepted for publication authors are requested to send its LaTex file via e-mail to one of the addresses:

sgorjanc@grad.hr ema.jurkin@rgn.hr

Figures should be included in EPS or PS formats and titled by the figure number that match to the figure number in the text of the paper.

The corresponding author and coauthors will recieve hard copies of the issue free of charge.

UPUTE ZA AUTORE

PODRUČJE. "KoG" objavljuje znanstvene i stručne radove iz područja geometrije, primijenjene geometrije i računalne grafike.

UPUTSTVA ZA PREDAJU RADA. Znanstveni radovi trebaju biti napisani na engleskom ili njemačkom jeziku, a stručni na hrvatskom, engleskom ili njemačkom. Rad ne smije biti objavljen niti predan na recenziju u drugim časopisima.
Rukopis sa širokim marginama i dvostrukim proredom šalje se u PDF formatu elektronskom poštom na adresu jedne od urednica:

Sonja Gorjanc	Ema Jurkin
sgorjanc@grad.hr	ema.jurkin@rgn.hr

Prva stranica treba sadržavati naslov rada, imena autora i koautora, podatke o autoru i koautorima, sažetak na hrvatskom i engleskom, ključne riječi i MSC broj.

PO PRIHVAĆANJU RADA. Tekst prihvaćenog rada autor dostavlja elektronskom poštom kao LaTeX datoteku, a slike u EPS ili PS formatu (s nazivima koji odgovaraju rednom broju slike u tekstu članka) na adresu:

sgorjanc@grad.hr ema.jurkin@rgn.hr

Svaki autor i koautor dobiva po jedan primjerak časopisa.

ISSN 1331–1611

KoG

No. 18
Zagreb, 2014

SCIENTIFIC AND PROFESSIONAL JOURNAL OF
CROATIAN SOCIETY FOR GEOMETRY AND GRAPHICS

CONTENTS

ORIGINAL SCIENTIFIC PAPERS

PROFESSIONAL PAPERS

ISSN 1331–1611

KoG

BROJ 18
Zagreb, 2014

ZNANSTVENO-STRUČNI ČASOPIS
HRVATSKOG DRUŠTVA ZA GEOMETRIJU I GRAFIKU

SADRŽAJ

ORIGINALNI ZNANSTVENI RADOVI

STRUČNI RADOVI

Original scientific paper
Accepted 4. 3. 2014.

NGUYEN LE
NORMAN JOHN WILDBERGER

Incenter Circles, Chromogeometry, and the Omega Triangle

Incenter Circles, Chromogeometry, and the Omega Triangle

ABSTRACT

Chromogeometry brings together planar Euclidean geometry, here called *blue* geometry, and two relativistic geometries, called *red* and *green*. We show that if a triangle has four blue Incenters and four red Incenters, then these eight points lie on a green circle, whose center is the green Orthocenter of the triangle, and similarly for the other colours. Tangents to the incenter circles yield interesting additional standard quadrangles and concurrencies. The proofs use the framework of rational trigonometry together with standard coordinates for triangle geometry, while a dilation argument allows us to extend the results also to Nagel and Speiker points.

Key words: triangle geometry, incenter circles, rational trigonometry, chromogeometry, four-fold symmetry, Nagel points, Spieker points, Omega triangle

MSC 2000: 51M05, 51M10, 51N10

Upisane kružnice, kromogeometrija i Omega trokut

SAŽETAK

Kromogeometrija povezuje ravninsku euklidsku geometriju, ovdje zvanu *plavom* geometrijom, te dvije relativističke geometrije, nazvane *crvenom* i *zelenom* geometrijom. Pokazuje se da ukoliko trokut ima četiri plava i četiri crvena središta upisanih (odnosno pripisanih) kružnica, tada tih osam točaka leži na zelenoj kružnici čije je središte zeleni ortocentar trokuta. Vrijede i druge dvije analogne tvrdnje. Tangente na upisane kružnice stvaraju nove zanimljive četverokute i konkurentnosti. Dokazi se provode u okviru racionalne trigonometrije sa standardnim koordinatama za geometriju trokuta. Transformacija diletacije dozvoljava proširenje rezultata na Nagelove i Speikerove točke.

Ključne riječi: geometrija trokuta, upisane kružnice, racionalna trigonometrija, kromogeometrija, četverostruka simetrija, Nagelove točke, Spiekerove točke, Omega trokut

1 Introduction

This paper investigates a surprising connection between three closely related Incenter hierarchies of a fixed planar triangle. The framework here is that of Rational Trigonometry ([7], [8]) which allows a consistent *universal triangle geometry* valid for any symmetric bilinear form, as described in [5], together with the three-fold symmetry of *chromogeometry* ([9], [10]), which connects the familiar Euclidean (*blue*) geometry based on the symmetric bilinear form $x_1x_2 + y_1y_2$, and two relativistic geometries (*red* and *green*) based respectively on the bilinear forms $x_1x_2 - y_1y_2$ and $x_1y_2 + y_1x_2$. By working with the rational notions of quadrance and spread instead of the transcendental notions of distance and angle, the main laws of Rational Trigonometry allow metrical geometry, and so triangle geometry, to be developed in each of these three geometries in a parallel fashion, with mostly identical formulas and theorems.

The first results of this paper concern the four Incenters of a planar triangle in one of the three geometries, and were announced in [5]. As in that paper, we here refer to all four meets of the vertex bisectors, or bilines, as Incenters, so do not distinguish between the classical incenter and the three excenters. If a triangle $\overline{A_1A_2A_3}$ has four blue Incenters I_0^b, I_1^b, I_2^b and I_3^b, then all four points lie both on a *red incenter circle* $\mathcal{C}_r^b$ with center the red Orthocenter H_r, and on a *green incenter circle* $\mathcal{C}_g^b$ with center the green Orthocenter H_g; this is illustrated in Figure 1. Similarly, if a triangle has red Incenters, then these lie both on a green incenter circle $\mathcal{C}_g^r$ with center H_g, and a blue incenter circle $\mathcal{C}_b^r$ with center the blue Orthocenter H_b. If a triangle has green Incenters, these lie both on a blue incenter circle $\mathcal{C}_b^g$ with center H_b, and on a red incenter circle $\mathcal{C}_r^g$ with center H_r. Furthermore, if *both* red and green Incenters exist, then they lie on the *same* blue incenter circle, and similarly for the other colours. The proofs are algebraic, and rely on non-obvious simplifications found by the help of a

computer. So the Omega triangle formed by the three Orthocenters $\Omega \equiv \overline{H_b H_r H_g}$, introduced in [9], has an intimate connection with the Incenter hierarchies.

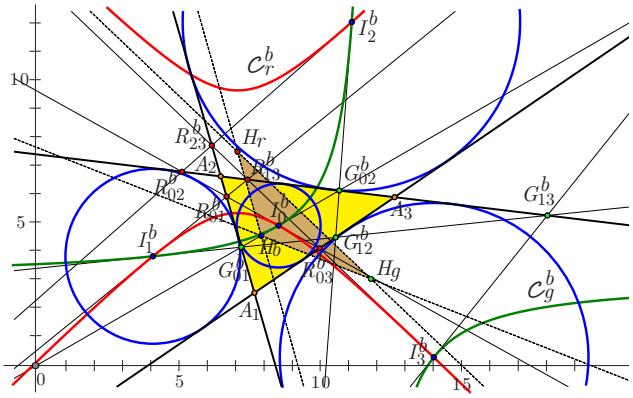


Figure 1: *The four blue Incenters of* $\overline{A_1A_2A_3}$ *and red and green Incenter Circles*

These facts relate also to elegant classical properties of quadrangles. In [1] Haskell showed that if two quadrangles have the same diagonal triangle, then all eight points of these quadrangles lie on a single conic; and in [11] Woods found a synthetic derivation of the same result. Now it is obvious that the four Incenters of a triangle, with respect to any bilinear form, will form a standard quadrangle in this sense, meaning that the diagonal triangle coincides with the original triangle. As a consequence, if blue and red Incenters exist, then they must lie on a conic. Our assertion is that this conic is actually a green circle $\mathcal{C}_g^b = \mathcal{C}_g^r \equiv \mathcal{C}_g$ with center H_g.

In the case of blue Incenters, the four tangent lines to the red incenter circle $\mathcal{C}_r^b$ at the blue Incenters form a standard quadrilateral, implying that they meet in six points R_{ij}^b, which lie two at a time on the three lines of $\overline{A_1A_2A_3}$, where they are harmonic conjugates with respect to A_1, A_2 and A_3; and similarly the four tangent lines to the green incenter circle $\mathcal{C}_g^b$ at the blue Incenters meet in six points G_{ij}^b on the three lines. This is also seen on the above Figure. Similarly there is a corresponding result when we look at red Incenters, and at green Incenters.

The six lines $A_k R_{ij}^b$, for i, j, k distinct, are the lines of a complete quadrangle, so they meet three at a time at four **quad points** Q_{rj}^b. Similarly, the six lines $A_k G_{ij}^b$ meet three at a time at points Q_{gj}^b. Somewhat remarkably, the four **star lines** $s_j^b \equiv Q_{rj}^b Q_{gj}^b$ form a standard quadrilateral $\overline{s_0^b s_1^b s_2^b s_3^b}$.

This paper also illustrates our novel approach to triangle geometry initiated in [5]; using standard coordinates to establish universal aspects of the subject which are *valid over a general bilinear form*. This employs an affine change of coordinates to place an arbitrary triangle into **standard position**, with vertices at $[0,0]$, $[1,0]$ and $[0,1]$. The various triangle centers and constructions are then expressed in terms of the coefficients a, b and c of the matrix

$$C \equiv \begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

of the resulting new bilinear form. This allows a systematic augmentation of Kimberling's *Encyclopedia of Triangle Centers* ([2], [3], [4]) to *arbitrary quadratic forms* and *general fields*.

Standard coordinates also have the advantage of yielding surprisingly simple equations for the three coloured Incenter Circles, which turn out to be, after pleasant simplifications,

$$\mathcal{C}_b : Q_b(X) = b_b(2x + 2y - 1)$$
$$\mathcal{C}_r : Q_r(X) = b_r(2x + 2y - 1)$$
$$\mathcal{C}_g : Q_g(X) = b_g(2x + 2y - 1).$$

However the formulas for the star lines s_j^b become rather formidable, but seem to have interesting algebraic aspects. Some intriguing number theoretical questions arise when we inquire into the existence of triangles, over a given field, which have simultaneously blue, red and green Incenters. Studying concrete examples and using empirical computer investigations of Michael Reynolds [6], we make some tentative conjectures on such triangles, both over the rational numbers and over a finite prime field. Finally we extend the results to Spieker and Nagel points by suitable central dilations.

In the rest of this introduction we recall basic facts from [7] and [5] to formulate triangle geometry over a general bilinear form. We then specialize to the blue, red and green geometries, and use standard coordinates to develop formulas for points and lines (always one of our key aims), and to provide explicit computational proofs of the theorems.

1.1 Quadrilaterals and quadrangles

We begin by reminding the reader of some basic facts from the projective geometry of a quadrangle (four points) or quadrilateral (four lines), using a visual presentation to avoid the need to introduce notation.

In Figure 2 we see four blue lines forming a quadrilateral [in this figure colours are not used in a metrical sense, but only as an aide for explanation]. These four blue lines meet in six points, also in blue. These six blue points determine a further three green **diagonal lines**, forming the **diagonal triangle,** in yellow**,** of the original quadrilateral, whose vertices are three green points. Each green point may be joined via a red line to the two blue points not on either of the two green lines it lies on. This produces six red

lines, which somewhat remarkably meet three at a time at four red points, giving the **opposite quadrangle** from the original blue quadrilateral. Note that there is a natural correspondence between the four original blue lines and the four red points.

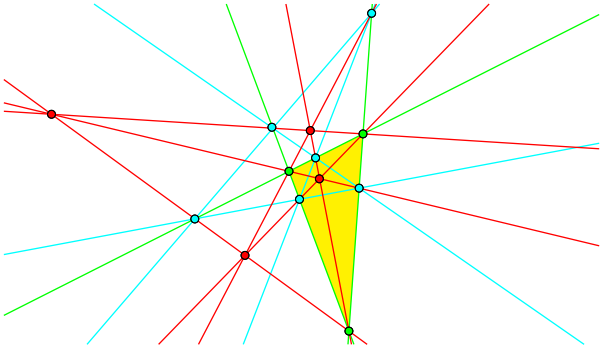


Figure 2: *A quadrilateral and its opposite quadrangle*

The situation is completely symmetric with regard to points and lines. If we had started out with a quadrilateral of four red points, we would join them to form six red lines. These six red lines determine a further three green diagonal points, forming the diagonal triangle of the original quadrilateral, whose sides form three green lines. Each green line meets two of the red lines in two new blue points. These six new blue points lie three at a time on four blue lines, giving the **opposite quadrilateral** from the original red quadrangle.
The diagonal green points on a green line are harmonic conjugates with respect to the two blue points on the same line. The diagonal green lines through a green point are harmonic conjugates with respect to the two red lines through the same point.
There is another more subtle remark to be made here concerning symmetry: each of the three diagonal points is canonically associated to a subdivision of the four original blue lines into two subsets of two: namely those subsets whose joins meet at that diagonal point.
If we start with a triangle, say the yellow triangle in the Figure formed by three green points and three green lines, then any quadrilateral or quadrangle which has that triangle as its diagonal triangle is called **standard**.

1.2 Quadrance, spread and standard coordinates

In this section we briefly summarize the main facts needed from rational trigonometry in the general affine setting (see [7], [8]). We work in the standard two-dimensional vector space V, consisting of row vectors $v = [x, y]$, over a field $\mathbb{F}$.
A **line** l is given by an equation of the form $ax + by + c = 0$, or equivalently the proportion $l \equiv \langle a : b : c \rangle$.
We assume a metrical structure determined by a non-degenerate symmetric 2×2 matrix C: this gives a symmetric bilinear form on vectors:

$$v \cdot u \equiv vCu^T.$$

Non-degenerate means $\det C \neq 0$, and implies that if $v \cdot u = 0$ for all vectors u, then $v = 0$.
Two vectors v and u are then **perpendicular** precisely when $v \cdot u = 0$. Since the matrix C is non-degenerate, for any vector v there is, up to a scalar, exactly one vector u which is perpendicular to v. Two lines l and m are **perpendicular** precisely when they have perpendicular direction vectors.
The bilinear form determines the main metrical quantity: the **quadrance** of a vector v is the number

$$Q_v \equiv v \cdot v.$$

The **quadrance** between the points A and B is $Q(A,B) \equiv Q_{\overrightarrow{AB}}$. A vector v is **null** precisely when $Q_v = v \cdot v = 0$, in other words precisely when v is perpendicular to itself. A line is **null** precisely when it has a null direction vector. The following basic fact appears in [5].

Theorem 1 (Parallel vectors) *Vectors v and u are parallel precisely when*

$$Q_v Q_u = (v \cdot u)^2.$$

This motivates the following measure of the non-parallelism of two vectors; the **spread** between non-null vectors v and u is the number

$$s(v,u) \equiv 1 - \frac{(v \cdot u)^2}{Q_v Q_u} = 1 - \frac{(v \cdot u)^2}{(v \cdot v)(u \cdot u)}.$$

The spread $s(v,u)$ is unchanged if either v or u are multiplied by a non-zero number. We define the **spread** between any non-null lines l and m with direction vectors v and u to be $s(l,m) \equiv s(v,u)$. From Theorem 1, the spread between parallel lines is 0. Two non-null lines l and m are perpendicular precisely when the spread between them is 1.
A **circle** is given by an equation of the form $Q(A,X) = K$ for some fixed point A called the **center**, and a number K called the **quadrance**. Note that it is not required that a circle have any points X lying on it: in this case by enlarging the field to a quadratic extension we can guarantee that it does.
The three particular planar geometries we are most interested in come from the **blue**, **red** and **green** bilinear forms given by the respective matrices

$$C_b \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad C_r \equiv \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad \text{and} \quad C_g \equiv \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The corresponding formulas for the **blue**, **red** and **green** **quadrances** between points $A_1 \equiv [x_1, y_1]$ and $A_2 \equiv [x_2, y_2]$

are

$$Q_b(A_1,A_2) = (x_2-x_1)^2 + (y_2-y_1)^2$$
$$Q_r(A_1,A_2) = (x_2-x_1)^2 - (y_2-y_1)^2$$
$$Q_g(A_1,A_2) = 2(x_2-x_1)(y_2-y_1).$$

It will be useful to discuss triangle geometry then in a general setting: suppose $v_1 \circ v_2 \equiv v_1 B v_2^T$ is a symmetric bilinear form, with B a symmetric 2×2 matrix. Suppose $\phi : V \to V$ is a linear transformation given by an invertible 2×2 matrix M, so that $\phi(v) = vM = w$, with inverse matrix N, so that $wN = v$. The new bilinear form $w_1 \cdot w_2 \equiv (w_1N)\circ(w_2N)$ then has matrix $D = NBN^T$.
Suppose that $\overline{X_1X_2X_3}$ is a triangle in the vector space V which has a distinguished symmetric bilinear form $\circ$. We may move this triangle by a combination of a translation (which does not effect the bilinear form), and a linear transformation ϕ, so that the triangle is in what we call **standard form**, with points

$$A_1 \equiv [0,0], \quad A_2 \equiv [1,0] \quad \text{and} \quad A_3 \equiv [0,1]$$

and lines

$$l_1 \equiv A_2A_3 = \langle 1:1:-1\rangle$$
$$l_2 \equiv A_1A_3 = \langle 1:0:0\rangle$$
$$l_3 \equiv A_2A_1 = \langle 0:1:0\rangle.$$

Whatever the initial matrix B, the new bilinear form $\cdot$ is given by

$$v\cdot u \equiv vDu^T \quad \text{where} \quad D \equiv NBN^T = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \tag{1}$$

for some numbers a,b, and c. We may then replace arguments involving the general triangle $\overline{X_1X_2X_3}$ and the bilinear form $\circ$ with ones involving the simpler triangle $\overline{A_1A_2A_3}$. What we prove for the standard triangle $\overline{A_1A_2A_3}$ with bilinear form given by the matrix D will be true for the original triangle with bilinear form given by the original matrix B.
We will assume that D is invertible, so that

$$\Delta \equiv \det D = ac - b^2$$

is non-zero. Another important quantity is the **mixed trace**

$$d \equiv a + c - 2b$$

that appears in many formulas. With these notations, we have the following result from [5].

Theorem 2 (Standard quadrances and spreads) *The quadrances and spreads of $\overline{A_1A_2A_3}$ are*

$$Q_1 \equiv Q(A_2,A_3) = d$$
$$Q_2 \equiv Q(A_1,A_3) = c$$
$$Q_3 \equiv Q(A_1,A_2) = a$$

and

$$s_1 \equiv s(A_1A_2, A_1A_3) = \frac{\Delta}{ac}$$
$$s_2 \equiv s(A_2A_3, A_2A_1) = \frac{\Delta}{ad}$$
$$s_3 \equiv s(A_3A_1, A_3A_2) = \frac{\Delta}{cd}.$$

Furthermore

$$1 - s_1 = \frac{b^2}{ac}, \quad 1 - s_2 = \frac{(a-b)^2}{ad}, \quad 1 - s_3 = \frac{(c-b)^2}{cd}.$$

Note that the centroid of $\overline{A_1A_2A_3}$ is

$$G = \left[\frac{1}{3}, \frac{1}{3}\right].$$

1.3 Bilines, Incenters and some other triangle centers

A **biline** of the non-null vertex $\overline{l_1l_2}$ is a line b which passes through l_1l_2 and satisfies $s(l_1,b) = s(b,l_2)$. The existence of bilines depends on number theoretical considerations of a particularly simple kind.

Theorem 3 (Existence of Triangle bilines) *The Triangle $\overline{A_1A_2A_3}$ has bilines at each vertex precisely when we can find numbers u,v,w in the field satisfying*

$$ac = u^2, \quad ad = v^2, \quad cd = w^2. \tag{2}$$

In this case we can choose u,v,w so that $acd = uvw$ and

$$du = vw, \quad cv = uw \quad \text{and} \quad aw = uv. \tag{3}$$

We now summarize some basic triangle centers of the standard triangle $\overline{A_1A_2A_3}$, assuming the existence of bilines. These formulas involve the entries a,b,c of D from (1), as well as the secondary quantities u,v and w from (2), satisfying (3). The formulas and proofs are found in [5].
The four Incenters are

$$I_0 = \frac{1}{d+v-w}[-w,v], \qquad I_1 = \frac{1}{d-v+w}[w,-v],$$
$$I_2 = \frac{1}{d+v+w}[w,v], \qquad I_3 = \frac{1}{d-v-w}[-w,-v].$$

Notice that I_1, I_2 and I_3 may be obtained from I_0 by changing signs of: both v and w, just w, and just v respectively. This four-fold symmetry will hold more generally and note that it means that we can generally just record the values of I_0. The Orthocenter H, Circumcenter C and De Longchamps point X_{20} (the orthocenter of the double triangle) are

$$H = \frac{b}{\Delta}[c-b, a-b] \tag{4}$$
$$C = \frac{1}{2\Delta}[c(a-b), a(c-b)]$$
$$X_{20} = \frac{1}{\Delta}\left[b^2 - 2bc + ac, b^2 - 2ab + ac\right].$$

2 The Incenter Circle theorem

Here is the main theorem of the paper, illustrated for green Incenters of the triangle $\overline{A_1A_2A_3}$ in Figure 3. The situation is completely symmetric between the three geometries blue, red and green.

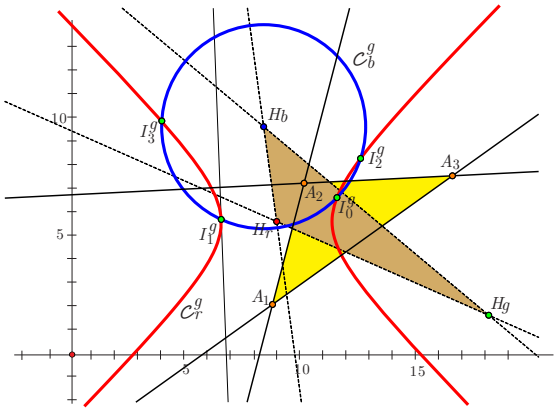


Figure 3: *Green Incenters and the blue and red Incenter Circles*

Theorem 4 (Incenter Circles) *If a triangle $\overline{A_1A_2A_3}$ has four blue Incenters I_0^b, I_1^b, I_2^b and I_3^b, then they all lie both on a red circle $\mathcal{C}_r^b$ with center the red Orthocenter H_r, and on a green circle $\mathcal{C}_g^b$ with center the green Orthocenter H_g, and similarly for the other colours. Furthermore, if both red and green Incenters exist, then they lie on the same blue circle, so that $\mathcal{C}_b^r = \mathcal{C}_b^g = \mathcal{C}_b$, and similarly for the other colours.*

Proof. To prove that the four blue Incenters I_0^b, I_1^b, I_2^b and I_3^b lie on a red circle $\mathcal{C}_r^b$ with center H_r, we need show that

$$Q_r\left(H_r, I_0^b\right) = Q_r\left(H_r, I_1^b\right) = Q_r\left(H_r, I_2^b\right) = Q_r\left(H_r, I_3^b\right).$$

First we find the bilinear forms for the blue, red and green geometries. After translating, and then applying a linear transformation with the matrix M, we send the original triangle to the standard triangle $\overline{A_1A_2A_3}$. If $M^{-1} = N = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$, then the bilinear forms for the blue, red and green geometries become respectively the matrices

$$D_b \equiv \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}^T = \begin{pmatrix} \alpha^2+\beta^2 & \alpha\gamma+\beta\delta \\ \alpha\gamma+\beta\delta & \gamma^2+\delta^2 \end{pmatrix} \equiv \begin{pmatrix} a_b & b_b \\ b_b & c_b \end{pmatrix}$$

$$D_r \equiv \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}^T = \begin{pmatrix} \alpha^2-\beta^2 & \alpha\gamma-\beta\delta \\ \alpha\gamma-\beta\delta & \gamma^2-\delta^2 \end{pmatrix} \equiv \begin{pmatrix} a_r & b_r \\ b_r & c_r \end{pmatrix}$$

$$D_g \equiv \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}^T = \begin{pmatrix} 2\alpha\beta & \alpha\delta+\beta\gamma \\ \alpha\delta+\beta\gamma & 2\gamma\delta \end{pmatrix} \equiv \begin{pmatrix} a_g & b_g \\ b_g & c_g \end{pmatrix}.$$

There are interesting relations between the introduced quantities; for example

$$a_b^2 = a_g^2 + a_r^2, \qquad a_b c_b = b_g^2 + b_r^2,$$
$$a_r c_r = b_b^2 - b_g^2, \qquad a_g c_g = b_b^2 - b_r^2, \qquad c_b^2 = c_g^2 + c_r^2$$

and

$$a_b c_g - 2b_b b_g + c_b a_g = 0, \qquad a_b c_r - 2b_b b_r + c_b a_r = 0,$$

$$a_g c_r - 2b_g b_r + c_g a_r = 0.$$

The determinants of D_b, D_r and D_g are respectively

$$\Delta_b = (\alpha\delta - \beta\gamma)^2, \qquad \Delta_r = \Delta_g = -(\alpha\delta - \beta\gamma)^2 = -\Delta_b$$

and the mixed traces are

$$d_b = (\alpha-\gamma)^2 + (\beta-\delta)^2, \qquad d_r = (\alpha-\gamma)^2 - (\beta-\delta)^2,$$

$$d_g = 2(\alpha-\gamma)(\beta-\delta).$$

Note also the relation $d_b^2 = d_r^2 + d_g^2$.
If the original triangle has four blue Incenters, then the Existence of Triangle bilines theorem shows that we may choose numbers u_b, v_b, w_b satisfying (2) and (3), so that

$$u_b^2 = (\alpha^2+\beta^2)(\gamma^2+\delta^2)$$
$$v_b^2 = (\alpha^2+\beta^2)\left((\alpha-\gamma)^2+(\beta-\delta)^2\right)$$
$$w_b^2 = (\gamma^2+\delta^2)\left((\alpha-\gamma)^2+(\beta-\delta)^2\right).$$

The blue Incenters are then

$$I_0^b = \frac{1}{d_b+v_b-w_b}[-w_b, v_b], \quad I_1^b = \frac{1}{d_b-v_b+w_b}[w_b, -v_b],$$
$$I_2^b = \frac{1}{d_b+v_b+w_b}[w_b, v_b], \quad I_3^b = \frac{1}{d_b-v_b-w_b}[-w_b, -v_b].$$

In exactly the same fashion

$$I_0^r = \frac{1}{d_r+v_r-w_r}[-w_r, v_r] \quad \text{and} \quad I_0^g = \frac{1}{d_g+v_g-w_g}[-w_g, v_g].$$

According to (4), the respective orthocenters are

$$H_b = \frac{b_b}{\Delta_b}[c_b - b_b, a_b - b_b], \qquad H_r = \frac{b_r}{\Delta_r}[c_r - b_r, a_r - b_r],$$

$$H_g = \frac{b_g}{\Delta_g}[c_g - b_g, a_g - b_g].$$

If we set $e_b \equiv d_b + v_b - w_b$ then

$$\overrightarrow{H_r I_0^b} = -\left(\frac{b_r(c_r-b_r)}{\Delta_r} + \frac{w_b}{e_b}, \frac{b_r(a_r-b_r)}{\Delta_r} - \frac{v_b}{e_b}\right)$$
$$= -\frac{1}{\Delta_r e_b}\left(b_r(c_r-b_r)e_b + \Delta_r w_b, b_r(a_r-b_r)e_b - \Delta_r v_b\right)$$

so that

$$Q_r\left(H_r, I_0^b\right) = \left(\overrightarrow{H_r I_0^b}\right) D_r \left(\overrightarrow{H_r I_0^b}\right)^T$$
$$= \begin{pmatrix} \frac{b_r(c_r-b_r)}{\Delta_r} + \frac{w_b}{e_b} & \frac{b_r(a_r-b_r)}{\Delta_r} - \frac{v_b}{e_b} \end{pmatrix} \begin{pmatrix} a_r & b_r \\ b_r & c_r \end{pmatrix} \begin{pmatrix} \frac{b_r(c_r-b_r)}{\Delta_r} + \frac{w_b}{e_b} \\ \frac{b_r(a_r-b_r)}{\Delta_r} - \frac{v_b}{e_b} \end{pmatrix}$$
$$= \frac{1}{\Delta_r^2 e_b^2}\begin{pmatrix} a_r\left(b_r(c_r-b_r)e_b + \Delta_r w_b\right)^2 + c_r\left(b_r(a_r-b_r)e_b - \Delta_r v_b\right)^2 \\ +2b_r\left(b_r(c_r-b_r)e_b + \Delta_r w_b\right)\left(b_r(a_r-b_r)e_b - \Delta_r v_b\right) \end{pmatrix}$$
$$= \frac{1}{\Delta_r^2 e_b^2}\begin{pmatrix} b_r^2(a_r-2b_r+c_r)\left(a_r c_r - b_r^2\right)e_b^2 \\ -2\Delta_r b_r(v_b-w_b)\left(-b_r^2+a_r c_r\right)e_b \\ +\Delta_r^2\left(a_r w_b^2 + c_r v_b^2 - 2b_r v_b w_b\right) \end{pmatrix}.$$

Use the relation $\Delta_r = a_r c_r - b_r^2$ to get

$$Q_r\left(H_r, I_0^b\right) \tag{5}$$
$$= \frac{1}{\Delta_r^2 e_b}\begin{pmatrix} b_r^2(a_r-2b_r+c_r)e_b^2 \\ -2\Delta_r b_r(v_b-w_b)e_b + \Delta_r\left(a_r w_b^2 + c_r v_b^2 - 2b_r v_b w_b\right) \end{pmatrix}$$
$$= \frac{\begin{pmatrix} 2b_r(b_r-c_r)(a_r-b_r)(v_b d_b - v_b w_b - w_b d_b) \\ +a_r(b_r-c_r)^2 v_b^2 + c_r(a_r-b_r)^2 w_b^2 + b_r^2 d_r d_b^2 \end{pmatrix}}{\Delta_r^2 e_b}$$

where we have collected v_b^2, w_b^2 and d_b^2 of the numerator of (5), to rewrite it.
Replace $v_b^2 = a_b d_b$, $w_b^2 = c_b d_b$ and $v_b w_b = u_b d_b$ and the values of $a_b, c_b, d_b, a_r, b_r, c_r$ in terms of $\alpha, \beta, \gamma, \delta$ to get the factorization

$$2b_r(b_r-c_r)(a_r-b_r)(v_b-u_b-w_b)d_b + a_r(b_r-c_r)^2 a_b d_b$$
$$+ c_r(a_r-b_r)^2 c_b d_b + b_r^2(a_r+c_r-2b_r)d_b^2$$
$$= d_b\begin{pmatrix} 2b_r(b_r-c_r)(a_r-b_r)(v_b-u_b-w_b) + a_r(b_r-c_r)^2 a_b \\ +c_r(a_r-b_r)^2 c_b + b_r^2(a_r+c_r-2b_r)d_b \end{pmatrix}$$
$$= 2d_b(\alpha\gamma-\beta\delta)\left(\alpha^2-\alpha\gamma+\gamma^2+\beta^2-\beta\delta+\delta^2-u_b+v_b-w_b\right)$$
$$\times\left(\alpha^2-\beta^2-\alpha\gamma+\beta\delta\right)\left(-\gamma^2+\delta^2+\alpha\gamma-\beta\delta\right) \tag{6}$$

and also note that

$$(d_b+v_b-w_b)^2 = d_b(a_b+c_b+d_b-2u_b+2v_b-2w_b)$$
$$= 2d_b\left(\alpha^2-\alpha\gamma+\gamma^2+\beta^2-\beta\delta+\delta^2-u_b+v_b-w_b\right). \tag{7}$$

Combine (6) and (7), to get the surprisingly simple formula

$$Q_r\left(H_r, I_0^b\right)$$
$$= \frac{(\alpha\gamma-\beta\delta)\left(\alpha^2-\beta^2-\alpha\gamma+\beta\delta\right)\left(-\gamma^2+\delta^2+\alpha\gamma-\beta\delta\right)}{\Delta_r}$$
$$= \frac{b_r(a_r-b_r)(b_r-c_r)}{\Delta_r} \equiv K_r.$$

We may now repeat the calculation to see that $Q_r\left(H_r, I_1^b\right) = Q_r\left(H_r, I_2^b\right) = Q_r\left(H_r, I_3^b\right) = K_r$, showing that indeed the four blue Incenters lie on the red circle $\mathcal{C}_r^b$ with quadrance K_r and center H_r. Note that the expression for K_r depends only on the matrix D_r. Now a similar derivation shows that

$$Q_g\left(H_g, I_i^b\right) = \frac{b_g(a_g-b_g)(b_g-c_g)}{\Delta_g} \equiv K_g, \quad i = 0,1,2,3.$$

Hence the four blue Incenters also lie on a green circle $\mathcal{C}_g^b$ with quadrance K_g and center H_g. Similarly we find that if a triangle has four red Incenters, then they lie on a blue circle $\mathcal{C}_b^r$ with center H_b and quadrance

$$Q_b\left(H_b, I_i^g\right) = Q_b\left(H_b, I_i^r\right) = \frac{b_b(a_b-b_b)(b_b-c_b)}{\Delta_b} \equiv K_b$$

as well as on a green circle $\mathcal{C}_g^r$ with center H_g and quadrance K_g (the same one as above!) Similarly if a triangle has four green Incenters, then they lie on a blue circle $\mathcal{C}_b^g$ with center H_b and quadrance K_b, as well as on a red circle $\mathcal{C}_r^g$ with center H_r and quadrance K_r. The proof is complete. □

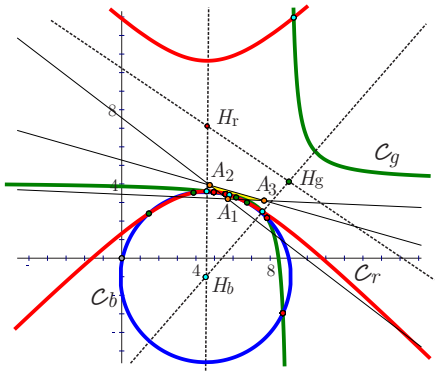


Figure 4: *Three Incenter Circles C_b, C_r and C_g.*

We now call $\mathcal{C}_b = \mathcal{C}_b^r = \mathcal{C}_b^g, \mathcal{C}_r = \mathcal{C}_r^b = \mathcal{C}_r^g$ and $\mathcal{C}_g = \mathcal{C}_g^b = \mathcal{C}_g^r$ the blue, red and green **Incenter Circles** respectively. In Figure 4 we see a (small) triangle $\overline{A_1A_2A_3}$ with its Omega triangle $\overline{H_bH_rH_g}$ and the three Incenter Circles, whose respective meets give the twelve blue, red and green Incenters.

2.1 Equations of Incenter Circles

Theorem 5 (Incenter Circles equations) *In standard coordinates with $X = [x,y]$, the blue, red and green Incenter circles, when they exist, have respective equations*

$$\mathcal{C}_b : Q_b(X) = b_b(2x+2y-1)$$
$$\mathcal{C}_r : Q_r(X) = b_r(2x+2y-1)$$
$$\mathcal{C}_g : Q_g(X) = b_g(2x+2y-1).$$

Proof. The derivation of these equations, using the formulas established above for the orthocenters H_r and coloured Incenters, is somewhat involved algebraically, although the basic idea is simple. We show how to find the equation of the red Incenter Circle $\mathcal{C}_r$, with center H_r, which four blue Incenters and four green Incenters lie on if they exist. From the definition of a red circle, we get the equation $Q_r(H_r, X) = K_r$, and then substitute the values of H_r and K_r to get

$$\begin{pmatrix} \frac{b_r(c_r-b_r)}{\Delta_r} - x & \frac{b_r(a_r-b_r)}{\Delta_r} - y \end{pmatrix} \begin{pmatrix} a_r & b_r \\ b_r & c_r \end{pmatrix} \begin{pmatrix} \frac{b_r(c_r-b_r)}{\Delta_r} - x \\ \frac{b_r(a_r-b_r)}{\Delta_r} - y \end{pmatrix} = \frac{b_r(a_r-b_r)(b_r-c_r)}{\Delta_r}$$

or after expansion

$$\frac{1}{\Delta_r^2} \begin{pmatrix} a_r(b_r(c_r-b_r)-\Delta_r x)^2 + c_r(b_r(a_r-b_r)-\Delta_r y)^2 \\ +2b_r(b_r(c_r-b_r)-\Delta_r x)(b_r(a_r-b_r)-\Delta_r y) \end{pmatrix} = \frac{b_r(a_r-b_r)(b_r-c_r)}{\Delta_r}.$$

This may be rewritten, using $\Delta_r = a_r c_r - b_r^2$, in the form

$$\frac{1}{\Delta_r}(\Delta_r^2 a_r x^2 + 2\Delta_r^2 b_r xy + \Delta_r^2 c_r y^2 + \Delta_r b_r^2(a_r - 2b_r + c_r) - 2\Delta_r^2 b_r x - 2\Delta_r^2 b_r y) = b_r(a_r-b_r)(b_r-c_r).$$

Now cancel Δ_r, and rearrange to get

$$\Delta_r a_r x^2 + 2\Delta_r b_r xy + \Delta_r c_r y^2 - 2\Delta_r b_r x - 2\Delta_r b_r y + b_r(a_r c_r - b_r^2) = 0$$

or more simply

$$a_r x^2 + 2b_r xy + c_r y^2 - 2b_r x - 2b_r y + b_r = 0$$

which has the form stated in the theorem. The same kind of calculation establishes the formulas for $\mathcal{C}_b$ and $\mathcal{C}_g$. □

Note that the equations for the Incenter Circles $\mathcal{C}_b, \mathcal{C}_r$ and $\mathcal{C}_g$ allow them to be defined *whether or not* the corresponding Incenters exist! Incenters then exist precisely as meets of these Incenter Circles: for example the blue Incenters $I_0^b, I_1^b, I_2^b, I_3^b$ are just the meets of $\mathcal{C}_r$ and $\mathcal{C}_g$, if these exist in the field in which we work.

2.2 Tangent lines of Incenter Circles

Now we consider tangent lines to Incenter circles. Figure 5 shows the four blue Incenters of $\overline{A_1A_2A_3}$, together with the red and green Incenter Circles passing through them, namely $\mathcal{C}_r$ and $\mathcal{C}_g$. At each of the four Incenters I_i^b, $i = 1,2,3,4$ we have the tangent lines t_{ri}^b and t_{gi}^b to the red and green Incenter Circles $\mathcal{C}_r$ and $\mathcal{C}_g$ respectively.

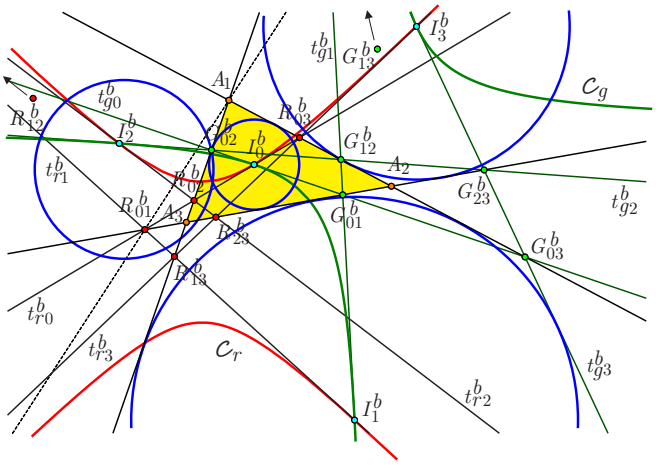


Figure 5: *Incenter tangent meets*

Theorem 6 (Incenter tangent meets) *The tangent lines $t_{r0}^b, t_{r1}^b, t_{r2}^b, t_{r3}^b$ to the red Incenter circle $\mathcal{C}_r$ at the blue Incenters form a standard quadrilateral, as do the tangent lines $t_{g0}^b, t_{g1}^b, t_{g2}^b, t_{g3}^b$ at the green Incenter circle $\mathcal{C}_g$. The same holds for the red and green Incenters, if they exist.*

This implies that the meets $R_{01}^b \equiv t_{r0}^b t_{r1}^b$ and $R_{23}^b \equiv t_{r2}^b t_{r3}^b$ lie on $l_1 = A_2A_3$, and are harmonic conjugates with respect to A_2 and A_3. Similarly $R_{02}^b \equiv t_{r0}^b t_{r2}^b$ and $R_{13}^b \equiv t_{r1}^b t_{r3}^b$ lie on $l_2 = A_1A_3$, and are harmonic conjugates with respect to A_1 and A_3; and $R_{03}^b \equiv t_{r0}^b t_{r3}^b$ and $R_{12}^b \equiv t_{r1}^b t_{r2}^b$ lie on $l_3 = A_1A_2$, and are harmonic conjugates with respect to A_1 and A_2. The points $G_{01}^b \equiv t_{g0}^b t_{g1}^b$ and $G_{23}^b \equiv t_{g2}^b t_{g3}^b$ lie on l_1, and are harmonic conjugates with respect to A_2 and A_3. Similarly $G_{02}^b \equiv t_{g0}^b t_{g2}^b$ and $G_{13}^b \equiv t_{g1}^b t_{g3}^b$ lie on l_2, and are harmonic conjugates with respect to A_1 and A_3, and $G_{03}^b \equiv t_{g0}^b t_{g3}^b$ and $G_{12}^b \equiv t_{g1}^b t_{g2}^b$ lie on l_3, and are harmonic conjugates with respect to A_1 and A_2.

Proof. We prove the result for the meets G_{ij}^b of the green tangent lines t_{gi}^b associated to the blue Incenters; the other cases are similar. We find the joins of a blue Incenter I_i^b and the green Orthocenter H_g to be

$$H_g I_0^b = \left\langle \begin{matrix} (b_g-a_g)b_g d_b + (c_g-b_g)a_g v_b + (a_g-b_g)b_g w_b : \\ (c_g-b_g)b_g d_b + (c_g-b_g)b_g v_b + (a_g-b_g)c_g w_b : \\ (b_g-c_g)b_g v_b + (b_g-a_g)b_g w_b \end{matrix} \right\rangle$$

$$H_g I_1^b = \left\langle \begin{matrix} (b_g-a_g)b_g d_b - (c_g-b_g)a_g v_b - (a_g-b_g)b_g w_b : \\ (c_g-b_g)b_g d_b - (c_g-b_g)b_g v_b - (a_g-b_g)c_g w_b : \\ -(b_g-c_g)b_g v_b - (b_g-a_g)b_g w_b \end{matrix} \right\rangle$$

$$H_g I_2^b = \left\langle \begin{matrix} (b_g-a_g)b_g d_b + (c_g-b_g)a_g v_b - (a_g-b_g)b_g w_b : \\ (c_g-b_g)b_g d_b + (c_g-b_g)b_g v_b - (a_g-b_g)c_g w_b : \\ (b_g-c_g)b_g v_b - (b_g-a_g)b_g w_b \end{matrix} \right\rangle$$

$$H_g I_3^b = \left\langle \begin{matrix} (b_g-a_g)b_g d_b - (c_g-b_g)a_g v_b + (a_g-b_g)b_g w_b : \\ (c_g-b_g)b_g d_b - (c_g-b_g)b_g v_b + (a_g-b_g)c_g w_b : \\ -(b_g-c_g)b_g v_b + (b_g-a_g)b_g w_b \end{matrix} \right\rangle.$$

The tangent line t^b_{gi} is the line green perpendicular to $H_g I^b_i$ passing through I^b_i. These we may calculate to be

$$t^b_{g0} = \left\langle \begin{matrix} (a_g - b_g)\,u_b + b_g v_b + (a_g - 2b_g)\,w_b + d_b b_g - c_b\,(a_g - b_g) : \\ (c_g - b_g)\,u_b + (2b_g - c_g)\,v_b - b_g w_b + d_b b_g + a_b\,(b_g - c_g) : \\ b_g\,(-v_b + w_b - d_b) \end{matrix} \right\rangle$$

$$t^b_{g1} = \left\langle \begin{matrix} (a_g - b_g)\,u_b - b_g v_b - (a_g - 2b_g)\,w_b + b_g d_b - c_b\,(a_g - b_g) : \\ (c_g - b_g)\,u_b - (2b_g - c_g)\,v_b + b_g w_b + d_b b_g + a_b\,(b_g - c_g) : \\ b_g\,(v_b - w_b - d_b) \end{matrix} \right\rangle$$

$$t^b_{g2} = \left\langle \begin{matrix} -(a_g - b_g)\,u_b + b_g v_b - (a_g - 2b_g)\,w_b + b_g d_b - c_b\,(a_g - b_g) : \\ -(c_g - b_g)\,u_b + (2b_g - c_g)\,v_b + b_g w_b + d_b b_g + a_b\,(b_g - c_g) : \\ b_g\,(-v_b - w_b - d_b) \end{matrix} \right\rangle$$

$$t^b_{g3} = \left\langle \begin{matrix} -(a_g - b_g)\,u_b - b_g v_b + (a_g - 2b_g)\,w_b + b_g d_b - c_b\,(a_g - b_g) : \\ -(c_g - b_g)\,u_b - (2b_g - c_g)\,v_b - b_g w_b + d_b b_g + a_b\,(b_g - c_g) : \\ b_g\,(v_b + w_b - d_b) \end{matrix} \right\rangle.$$

We could verify directly that these four lines form a standard quadrilateral. But we prefer to verify that the meets of these four tangent lines agree with the following meets with the side lines of $\overline{A_1A_2A_3}$:

$$G^b_{01} \equiv t^b_{g0} t^b_{g1} = t^b_{g0} l_1$$
$$= \frac{1}{\lambda_{01}} \left[(b_g - c_g)\,(-u_b + v_b + a_b), (a_g - b_g)(-u_b - w_b + c_b) \right]$$
$$G^b_{23} \equiv t^b_{g2} t^b_{g3} = t^b_{g2} l_1$$
$$= \frac{1}{\lambda_{23}} \left[(b_g - c_g)\,(u_b + v_b + a_b)\,, (a_g - b_g)\,(u_b + w_b + c_b) \right]$$
$$G^b_{02} \equiv t^b_{g0} t^b_{g2} = t^b_{g0} l_2 = \frac{1}{\lambda_{02}} \left[0, b_g\,(-v_b + w_b - d_b) \right]$$
$$G^b_{13} \equiv t^b_{g1} t^b_{g3} = t^b_{g1} l_2 = \frac{1}{\lambda_{13}} \left[0, b_g\,(-v_b + w_b + d_b) \right]$$
$$G^b_{03} \equiv t^b_{g0} t^b_{g3} = t^b_{g0} l_3 = \frac{1}{\lambda_{03}} \left[b_g\,(v_b - w_b + d_b)\,, 0 \right]$$
$$G^b_{12} \equiv t^b_{g1} t^b_{g2} = t^b_{g1} l_3 = \frac{1}{\lambda_{12}} \left[b_g\,(v_b - w_b - d_b)\,, 0 \right]$$

where

$$\begin{aligned}
\lambda_{01} &= (c_g - a_g)\,u_b + (b_g - c_g)\,v_b + (b_g - a_g)\,w_b \\
&\quad + (a_b b_g - a_b c_g + c_b a_g - c_b b_g) \\
\lambda_{23} &= (a_g - c_g)\,u_b + (b_g - c_g)\,v_b \\
&\quad + (a_g - b_g)\,w_b + (a_b b_g - a_b c_g + c_b a_g - c_b b_g) \\
\lambda_{02} &= (b_g - c_g)\,u_b + (c_g - 2b_g)\,v_b + b_g w_b \\
&\quad + (a_b c_g - 2a_b b_g + 2b_b b_g - c_b b_g) \\
\lambda_{13} &= (c_g - b_g)\,u_b + (c_g - 2b_g)\,v_b + b_g w_b \\
&\quad + (2a_b b_g - a_b c_g - 2b_b b_g + c_b b_g) \\
\lambda_{03} &= (a_g - b_g)\,u_b + b_g v_b + (a_g - 2b_g)\,w_b \\
&\quad + (a_b b_g - 2b_b b_g - c_b a_g + 2c_b b_g) \\
\lambda_{12} &= (b_g - a_g)\,u_b + b_g v_b + (a_g - 2b_g)\,w_b \\
&\quad + (2b_b b_g - a_b b_g + c_b a_g - 2c_b b_g)\,.
\end{aligned}$$

The fact that $A_2, A_3, G^b_{01,} G^b_{23}$ form a harmonic range etc. is an immediate consequence of a well known fact about standard quadrilaterals in projective geometry, since we have shown that the points A_1, A_2 and A_3 are diagonal points of the quadrilateral formed by the four green tangent lines. □

Following the construction of the red lines in the introductory section on Quadrangles and quadrilaterals, we join a point G^b_{ij} with the triangle point A_k opposite to the triangle line that it lies on; giving six lines $A_k G^b_{ij}$:

$$A_1 G^b_{01} = \left\langle \begin{matrix} (a_g - b_g)\,(u_b + w_b - c_b) : \\ (b_g - c_g)\,(-u_b + v_b + a_b) : \\ 0 \end{matrix} \right\rangle$$

$$A_1 G^b_{23} = \left\langle \begin{matrix} (b_g - a_g)\,(u_b + w_b + c_b) : \\ (b_g - c_g)\,(u_b + v_b + a_b) : \\ 0 \end{matrix} \right\rangle$$

$$A_2 G^b_{02} = \left\langle \begin{matrix} b_g\,(-v_b + w_b - d_b) : \\ (b_g - c_g)\,u_b + (c_g - 2b_g)\,v_b + b_g w_b - b_g d_b + a_b\,(c_g - b_g) : \\ b_g\,(v_b - w_b + d_b) \end{matrix} \right\rangle$$

$$A_2 G^b_{13} = \left\langle \begin{matrix} b_g\,(-v_b + w_b + d_b) : \\ (c_g - b_g)\,u_b + (c_g - 2b_g)\,v_b + b_g w_b + b_g d_b + a_b\,(b_g - c_g) : \\ b_g\,(v_b - w_b - d_b) \end{matrix} \right\rangle$$

$$A_3 G^b_{03} = \left\langle \begin{matrix} (a_g - b_g)\,u_b + b_g v_b + (a_g - 2b_g)\,w_b + b_g d_b + c_b\,(b_g - a_g) : \\ b_g\,(v_b - w_b + d_b) : \\ b_g\,(-v_b + w_b - d_b) \end{matrix} \right\rangle$$

$$A_3 G^b_{12} = \left\langle \begin{matrix} (b_g - a_g)\,u_b + b_g v_b + (a_g - 2b_g)\,w_b - b_g d_b + c_b\,(a_g - b_g) : \\ b_g\,(v_b - w_b - d_b) : \\ b_g\,(-v_b + w_b + d_b) \end{matrix} \right\rangle.$$

Theorem 7 (Quad points) *The triples* $\{A_1 G^b_{23}, A_2 G^b_{13}, A_3 G^b_{12}\}$, $\{A_1 G^b_{23}, A_2 G^b_{02}, A_3 G^b_{03}\}$, $\{A_1 G^b_{01}, A_2 G^b_{13}, A_3 G^b_{03}\}$ *and* $\{A_1 G^b_{01}, A_2 G^b_{02}, A_3 G^b_{12}\}$ *of lines are concurrent in the respective points* $Q^b_{g0}, Q^b_{g1}, Q^b_{g2}$ *and* Q^b_{g3}, *called the* ***blue/green quad points.*** *The triples* $\{A_1 R^b_{23}, A_2 R^b_{13}, A_3 R^b_{12}\}$, $\{A_1 R^b_{23}, A_2 R^b_{02}, A_3 R^b_{03}\}$, $\{A_1 R^b_{01}, A_2 R^b_{13}, A_3 R^b_{03}\}$ *and* $\{A_1 R^b_{01}, A_2 R^b_{02}, A_3 R^b_{12}\}$ *are also concurrent in the respective points* $Q^b_{r0}, Q^b_{r1}, Q^b_{r2}$ *and* Q^b_{r3}, *called the* ***blue/red quad points****. Similar results hold for the red and green Incenters, if they exist.*

Proof. We verify this for the blue/green quad points: this is a consequence of the projective geometry of the complete quadrilateral we mentioned in the first section—if the original four tangent lines are regarded as the blue lines in Figure 6, then the quad points Q^b_{gj} correspond to the red points. However we want to find explicit formulas and check things directly. The quad point Q^b_{gj} is naturally associated to the Incenter I^b_j. After some calculation, we find

that these are

$$Q^b_{g0} = \frac{b_g}{\lambda_0}[(b_g-c_g)(d_bu_b-(b_b-c_b)v_b),\\ (a_g-b_g)((c_b-b_b)w_b+c_bd_b)]$$

$$Q^b_{g1} = \frac{b_g}{\lambda_1}[(b_g-c_g)((a_b-b_b)v_b+a_bd_b),\\ (a_g-b_g)(d_bu_b+(a_b-b_b)w_b)]$$

$$Q^b_{g2} = \frac{b_g}{\lambda_2}[(c_g-b_g)(a_bw_b-b_bv_b),\\ (b_g-a_g)(b_bw_b-c_bv_b)]$$

$$Q^b_{g3} = \frac{b_g}{\lambda_3}[(b_g-c_g)(-d_bu_b+(d_b+b_b)v_b-a_bw_b+a_bd_b),\\ (b_g-a_g)(d_bu_b-c_bv_b+(d_b+b_b)w_b-c_bd_b)]$$

where

$$\begin{aligned}
\lambda_0 &= (b_g-c_g)(b_gd_b+(b_b-c_b)(a_g-b_g))u_b\\
&\quad -(b_g-c_g)(b_bb_g+c_ba_g-2c_bb_g)v_b\\
&\quad -b_g(a_g-b_g)(b_b-c_b)w_b+c_bb_g(a_g-b_g)d_b\\
\lambda_1 &= (a_g-b_g)(b_gd_b+(a_b-b_b)(b_g-c_g))u_b\\
&\quad +b_g(b_g-c_g)(a_b-b_b)v_b\\
&\quad +(a_g-b_g)(2a_bb_g-a_bc_g-b_bb_g)w_b+a_bb_g(b_g-c_g)d_b\\
\lambda_2 &= b_b(b_g-c_g)(a_g-b_g)u_b\\
&\quad +b_g(b_bb_g+c_ba_g-b_bc_g-c_bb_g)v_b\\
&\quad -b_g(a_bb_g+b_ba_g-a_bc_g-b_bb_g)w_b-a_bc_b(b_g-c_g)(a_g-b_g)\\
\lambda_3 &= \left((d_b+b_b)\left(b_g^2+a_gc_g\right)-a_gb_g(2d_b+b_b)-b_bb_gc_g\right)u_b\\
&\quad +\left(b_g((b_g-c_g)(a_b-b_b)+c_b(2a_g-b_g))-c_ba_gc_g\right)v_b\\
&\quad +((d_b+b_b)(b_g-a_g)b_g-a_ba_g(b_g-c_g))w_b\\
&\quad +b_g^2(a_b(d_b-c_b)-c_bd_b)+b_g(c_ba_g(d_b+a_b)-a_bc_g(d_b-c_b))\\
&\quad -a_gc_ga_bc_b.
\end{aligned}$$

We may then check directly that for example Q^b_{g0} is incident with $A_3G^b_{12}$ by computing

$$\begin{aligned}
&\left((b_g-a_g)u_b+b_gv_b+(a_g-2b_g)w_b-b_gd_b+c_b(a_g-b_g)\right)\cdot\\
&\quad\cdot\left(\frac{b_g(b_g-c_g)(d_bu_b-(b_b-c_b)v_b)}{\lambda_0}\right)\\
&+b_g(v_b-w_b-d_b)\left(\frac{b_g(a_g-b_g)((c_b-b_b)w_b+c_bd_b)}{\lambda_0}\right)\\
&+b_g(-v_b+w_b+d_b)\\
&=\frac{b_g(a_g-b_g)(b_g-c_g)\begin{pmatrix}d_bu_b^2+(b_b-a_b)u_bw_b-b_bd_bu_b\\ -c_bv_b^2+b_bv_bw_b+c_b(a_b-b_b)v_b\end{pmatrix}}{\begin{pmatrix}-(b_g-c_g)(b_gd_b+(b_b-c_b)(a_g-b_g))u_b\\ +(b_g-c_g)(b_bb_g+c_ba_g-2c_bb_g)v_b\\ +b_g(a_g-b_g)(b_b-c_b)w_b-c_bb_g(a_g-b_g)d_b\end{pmatrix}}\\
&=0
\end{aligned}$$

since $d_bu_b^2+(b_b-a_b)u_bw_b-b_bd_bu_b-c_bv_b^2+b_bv_bw_b+c_b(a_b-b_b)v_b=0$ by using (2), and similarly for the other indices. In a parallel fashion, we find that the four blue/red quad points Q^b_{rj} have exactly the same formulas as the Q^b_{gj}, except for the replacements $a_g\longrightarrow a_r$, $b_g\longrightarrow b_r$ and $c_g\longrightarrow c_r$, and similarly for the other colours red and green. □

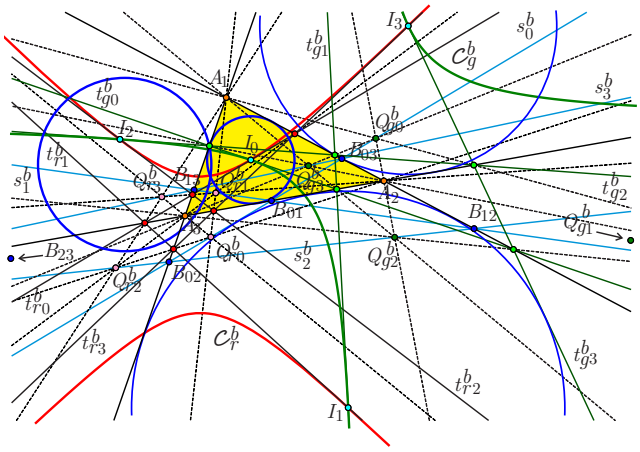


Figure 6: *Quad points and star lines*

Now introduce the **blue star line** s^b_j to be the join of the corresponding blue/red quad point Q^b_{rj} and the blue/green quad point Q^b_{gj}, and similarly for the other colours. There are then four blue star lines s^b_0, s^b_1, s^b_2 and s^b_3.

The **blue star point** B_{ij} is the meet of the two blue star lines s^b_i and s^b_j, that is $B_{ij}\equiv s^b_is^b_j$, and similarly for the other colours.

Note that following the introductory section on Quadrangles and quadrilaterals, we use the correspondence between the Q^b_{gj}; and the tangent lines t^b_{gj}; and the Incenters I^b_j to match up the indices.

Theorem 8 (Star quadrilateral) *The four blue star lines form a standard quadrilateral* $\overline{s^b_0s^b_1s^b_2s^b_3}$. *This holds also for the other colours.*

Proof. The proof we have is surprisingly complicated. The star lines s^b_j have quite involved formulas; for example we find that

$$\begin{aligned}
s^b_0 = Q^{b0}_gQ^{b0}_r = \\
E_0d_bu_b+F_0c_bv_b:\\
b_gb_r\left((b_b-c_b)^2+c_bd_b\right)\cdot\\
\cdot(a_gb_r-b_ga_r-a_gc_r+c_ga_r+b_gc_r-c_gb_r)d_bu_b\\
-2c_bb_gb_r(a_gb_r-b_ga_r-a_gc_r+c_ga_r+b_gc_r-c_gb_r)(b_b-c_b)d_bv_b\\
-a_b(b_r-c_r)(b_g-c_g)(a_gb_r-b_ga_r)\left((b_b-c_b)^2+c_bd_b\right)w_b\\
+2a_bc_b(b_r-c_r)(b_g-c_g)(a_gb_r-b_ga_r)(b_b-c_b)d_b\\
-b_gb_rd_b(a_gb_r-b_ga_r-a_gc_r+c_ga_r+b_gc_r-c_gb_r)\cdot\\
\cdot\left(\left((b_b-c_b)^2+c_bd_b\right)u_b-2c_b(b_b-c_b)v_b\right)
\end{aligned}$$

where E_0 and F_0 are both homogeneous polynomials of degree 6 in the variables a_i, b_i and c_i, with the former having 32 terms and the latter 46 terms. After some trial and error we can present these in the somewhat pleasant, but still mysterious, forms:

$$\begin{aligned}E_0 = &-b_g b_r (a_g c_r - c_g a_r - b_g c_r + c_g b_r)\cdot \\ &\cdot (b_b^2 + 4c_b^2 + a_b c_b - 6b_b c_b) \\ &+ b_g b_r (a_g b_r - b_g a_r)(b_b^2 + 2c_b^2 + a_b c_b - 4b_b c_b) \\ &- 2c_b (a_g c_g b_r^2 - a_r c_r b_g^2 + a_g a_r c_r b_g - a_g c_g a_r b_r)(b_b - c_b)\end{aligned}$$

and

$$\begin{aligned}F_0 = &(a_g c_g b_r^2 - a_r c_r b_g^2 + a_g a_r c_r b_g - a_g c_g a_r b_r)\cdot \\ &\cdot (b_b^2 - 4b_b c_b + 2c_b^2 + a_b c_b) \\ &+ b_g b_r (a_g c_r - c_g a_r - b_g c_r + c_g b_r)\cdot \\ &\cdot (-5b_b^2 - 4c_b^2 + 2a_b b_b - 3a_b c_b + 10b_b c_b) \\ &- 2b_g b_r (a_g b_r - b_g a_r)(b_b - c_b)(a_b - 2b_b + c_b).\end{aligned}$$

We can then calculate the blue star points, for example

$$B_{03} = \left[\frac{\left(b_g b_r d_b (a_g b_r - b_g a_r - a_g c_r + c_g a_r + b_g c_r - c_g b_r)\cdot \right.}{\left.\cdot \left(\left((b_b - c_b)^2 + c_b d_b\right) u_b - 2c_b (b_b - c_b) v_b\right)\right)}{E_0 d_b u_b + F_0 c_b v_b}, 0\right]$$

from which clearly B_{03} lies on l_3. The computations are similar for the other indices, and the other colours. □

3 Explicit examples and some conjectures

3.1 An example over $\mathbb{Q}(\sqrt{30}, \sqrt{217}, \sqrt{741}, \sqrt{2470}, \sqrt{82297})$

We will now explore in detail a particular triangle which has both blue, red and green Incenters; for us this is not only an important tool for checking the consistency of our formulas, but also a way to get a sense of the level of complexity of various constructions. In fact this kind of explicit calculation of examples is much to be encouraged in this subject: especially as working over concrete fields, including finite fields and explicit extension fields of the rationals, allows us to appreciate the number theoretic aspects of our geometrical investigations. For example, finding a triangle with blue, red and green Incenters approximately is easy with a geometry package: finding a concrete example and working out all the points precisely is more challenging.

In particular we were unable, despite a reasonable computer search, to find *any* triangles with purely rational points that have blue, red and green Incenters! We would like to thank Michael Reynolds for his contributions to this search. So we tentatively conjecture that *there are no such triangles*.

In any case, to get an explicit example we use an algebraic extension field of the rationals; so by $\sqrt{30}$ we mean an appropriate symbol in the extension field $\mathbb{Q}(\sqrt{30})$ etc.. Note that although our use of square roots is entirely algebraic, our representation of these square roots as approximate rational numbers (we prefer to avoid discussion of "real numbers"), necessarily brings an *approximate aspect into our diagrams*.

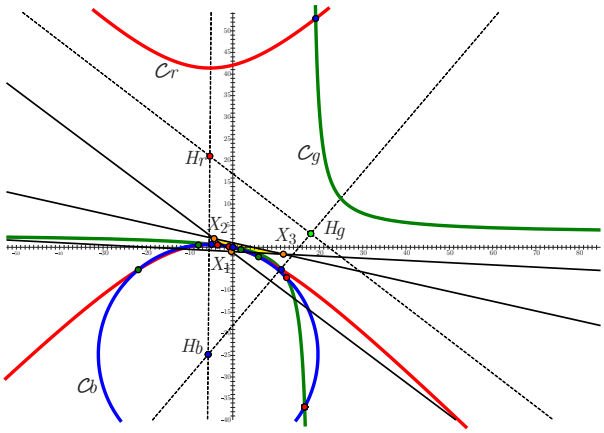


Figure 7: *An example triangle* $\overline{X_1X_2X_3}$

Example 1 *One may check that the basic Triangle with points*

$$X_1 \equiv [-21/59, -58/59], \quad X_2 \equiv [-13/3, 2] \quad \textit{and}$$

$$X_3 \equiv [35/3, -8/5]$$

in $\mathbb{Q}(\sqrt{30}, \sqrt{217}, \sqrt{741}, \sqrt{2470}, \sqrt{82297})$ *has both blue, red and green Incenters. After translation by* $(21/59, 58/59)$ *we obtain* $\widetilde{X}_1 = [0,0]$, $\widetilde{X}_2 = [-704/177, 176/59]$ *and* $\widetilde{X}_3 = [2128/177, -182/295]$. *The matrix N and its inverse M, where*

$$N = \begin{pmatrix} -\frac{704}{177} & \frac{176}{59} \\ \frac{2128}{177} & -\frac{182}{295} \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \qquad \textit{and}$$

$$M = N^{-1} = \begin{pmatrix} \frac{13}{704} & \frac{5}{56} \\ \frac{95}{264} & \frac{5}{42} \end{pmatrix}$$

respectively send $[1,0]$ *and* $[0,1]$ *to* $\widetilde{X}_2$ *and* $\widetilde{X}_3$, *and* $\widetilde{X}_2$ *and* $\widetilde{X}_3$ *to* $[1,0]$ *and* $[0,1]$. *From now on we discuss only the standard triangle* $\overline{A_1A_2A_3}$ *associated to* $\overline{X_1X_2X_3}$; *to convert back into the original coordinates, we would multiply by N and translate by* $(-21/59, -58/59)$. *The bilinear forms in these new standard coordinates, for the blue, red and green*

geometries respectively, are given by matrices

$$D_b = N\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}N^T = \begin{pmatrix}\frac{774400}{31329} & -\frac{7778848}{156645}\\ -\frac{7778848}{156645} & \frac{113507716}{783225}\end{pmatrix} = \begin{pmatrix}a_b & b_b\\ b_b & c_b\end{pmatrix}$$
$$D_r = N\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}N^T = \begin{pmatrix}\frac{216832}{31329} & -\frac{7202272}{156645}\\ -\frac{7202272}{156645} & \frac{112911484}{783225}\end{pmatrix} = \begin{pmatrix}a_r & b_r\\ b_r & c_r\end{pmatrix}$$
$$D_g = N\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}N^T = \begin{pmatrix}-\frac{247808}{10443} & \frac{2000768}{52215}\\ \frac{2000768}{52215} & -\frac{774592}{52215}\end{pmatrix} = \begin{pmatrix}a_g & b_g\\ b_g & c_g\end{pmatrix}.$$

The determinants of D_b, D_r and D_g are $\Delta_b = \frac{97140736}{87025}$ and $\Delta_r = \Delta_g = -\frac{97140736}{87025}$, while the mixed traces are $d_b = \frac{6724}{25}$, $d_r = \frac{6076}{25}$ and $d_g = -\frac{576}{5}$. The orthocenters of $\overline{A_1A_2A_3}$ are

$$H_b = \left[-\frac{8825537}{1019520}, -\frac{84337}{25488}\right], H_r = \left[\frac{87833227}{11214720}, \frac{55537}{25488}\right],$$
$$H_g = \left[\frac{7105}{3894}, \frac{377}{177}\right].$$

Blue, red and green Incenters exist over $\mathbb{F} = \mathbb{Q}\left(\sqrt{30}, \sqrt{217}, \sqrt{741}, \sqrt{2470}, \sqrt{82297}\right)$ and we may choose

$$u_b = \frac{1875104}{31329}, \quad v_b = \frac{14432}{177}, \quad w_b = \frac{873628}{4425}$$
$$u_r = \frac{17248}{156645}\sqrt{82297}, \quad v_r = \frac{2464}{885}\sqrt{217}$$
$$w_r = \frac{196}{4425}\sqrt{217}\sqrt{82297}$$
$$u_g = \frac{19712}{52215}\sqrt{2470}, \quad v_g = \frac{2816}{295}\sqrt{30}, \quad w_g = \frac{448}{295}\sqrt{741}.$$

Then the four blue Incenters, the four red Incenters and the four green Incenters of $\overline{A_1A_2A_3}$ respectively are

$$I_0^b = \left[-\frac{761}{590}, \frac{220}{413}\right] \qquad I_1^b = \left[\frac{5327}{10384}, -\frac{25}{118}\right]$$
$$I_2^b = \left[\frac{761}{2112}, \frac{25}{168}\right] \qquad I_3^b = \left[\frac{5327}{270}, \frac{220}{27}\right]$$

$$I_0^r = \begin{bmatrix}\frac{1}{22429440}(4032553\sqrt{217} - 20461\sqrt{217}\sqrt{82297}\\ +210343\sqrt{82297} - 76618507),\\ \frac{1}{50976}(2923\sqrt{217} - 7\sqrt{217}\sqrt{82297}\\ +133\sqrt{82297} - 30049)\end{bmatrix}$$

$$I_1^r = \begin{bmatrix}\frac{1}{22429440}(20461\sqrt{217}\sqrt{82297} - 4032553\sqrt{217}\\ +210343\sqrt{82297} - 76618507),\\ \frac{1}{50976}(7\sqrt{217}\sqrt{82297} - 2923\sqrt{217}\\ +133\sqrt{82297} - 30049)\end{bmatrix}$$

$$I_2^r = \begin{bmatrix}\frac{1}{22429440}(4032553\sqrt{217} + 20461\sqrt{217}\sqrt{82297}\\ -210343\sqrt{82297} - 76618507),\\ \frac{1}{50976}(7\sqrt{217}\sqrt{82297} + 2923\sqrt{217}\\ -133\sqrt{82297} - 30049)\end{bmatrix}$$

$$I_3^r = \begin{bmatrix}\frac{-1}{22429440}(4032553\sqrt{217} + 20461\sqrt{217}\sqrt{82297}\\ +210343\sqrt{82297} + 76618507),\\ \frac{-1}{50976}(7\sqrt{217}\sqrt{82297} - 2923\sqrt{217}\\ -133\sqrt{82297} - 30049)\end{bmatrix}$$

$$I_0^g = \begin{bmatrix}\frac{203}{7788}\sqrt{741} - \frac{247}{3894}\sqrt{30} + \frac{35}{3894}\sqrt{2470} - \frac{3211}{7788},\\ \frac{13}{1239}\sqrt{2470} - \frac{29}{177}\sqrt{30} + \frac{20}{1239}\sqrt{741} - \frac{100}{177}\end{bmatrix}$$
$$I_1^g = \begin{bmatrix}\frac{247}{3894}\sqrt{30} - \frac{203}{7788}\sqrt{741} + \frac{35}{3894}\sqrt{2470} - \frac{3211}{7788},\\ \frac{29}{177}\sqrt{30} + \frac{13}{1239}\sqrt{2470} - \frac{20}{1239}\sqrt{741} - \frac{100}{177}\end{bmatrix}$$
$$I_2^g = \begin{bmatrix}-\frac{247}{3894}\sqrt{30} - \frac{203}{7788}\sqrt{741} - \frac{35}{3894}\sqrt{2470} - \frac{3211}{7788},\\ -\frac{29}{177}\sqrt{30} - \frac{13}{1239}\sqrt{2470} - \frac{20}{1239}\sqrt{741} - \frac{100}{177}\end{bmatrix}$$
$$I_3^g = \begin{bmatrix}\frac{247}{3894}\sqrt{30} + \frac{203}{7788}\sqrt{741} - \frac{35}{3894}\sqrt{2470} - \frac{3211}{7788},\\ \frac{29}{177}\sqrt{30} - \frac{13}{1239}\sqrt{2470} + \frac{20}{1239}\sqrt{741} - \frac{100}{177}\end{bmatrix}.$$

The Incenter circle quadrances are

$$K_b = \frac{18154129609}{28196100}, \qquad K_r = -\frac{11681819191}{28196100}$$
$$K_g = \frac{1182272}{10443}.$$

The blue, red and green Incenter Circles themselves have respective equations

$$4840000x^2 - 19447120xy + 28376929y^2$$
$$+19447120x + 19447120y - 9723560 = 0$$
$$19360x^2 - 62524xy + 12103y^2$$
$$+62524x + 62524y - 31262 = 0$$
$$193600x^2 - 2572240xy + 4032553y^2$$
$$+2572240x + 2572240y - 1286120 = 0.$$

The four tangent lines t_{gj}^b are

$$t_{g0}^b = \langle 1570 : -11823 : 8323\rangle$$
$$t_{g1}^b = \langle -127512 : -33761 : 58261\rangle$$
$$t_{g2}^b = \langle -18216 : -11823 : 8323\rangle$$
$$t_{g3}^b = \langle -1570 : 4823 : 8323\rangle.$$

The meets of these four tangent lines agree with the following meets with the side lines of $\overline{A_1A_2A_3}$:

$$G^b_{01} \equiv t^b_{g0}t^b_{g1} = t^b_{g0}l_1 = \left[\frac{3500}{13393}, \frac{9893}{13393}\right]$$
$$G^b_{23} \equiv t^b_{g2}t^b_{g3} = t^b_{g2}l_1 = \left[-\frac{3500}{6393}, \frac{9893}{6393}\right]$$
$$G^b_{02} \equiv t^b_{g0}t^b_{g2} = t^b_{g0}l_2 = \left[0, \frac{1189}{1689}\right]$$
$$G^b_{13} \equiv t^b_{g1}t^b_{g3} = t^b_{g1}l_2 = \left[0, \frac{1189}{689}\right]$$
$$G^b_{03} \equiv t^b_{g0}t^b_{g3} = t^b_{g0}l_3 = \left[-\frac{8323}{1570}, 0\right]$$
$$G^b_{12} \equiv t^b_{g1}t^b_{g2} = t^b_{g1}l_3 = \left[\frac{8323}{18216}, 0\right].$$

The blue/red quad points Q^b_{rj} *associated to* $I^b_0, I^b_1, I^b_2, I^b_3$ *respectively are*

$$Q^b_{r0} = \left[\frac{18005811535}{21082889161}, -\frac{12129669559}{21082889161}\right]$$
$$Q^b_{r1} = \left[-\frac{18005811535}{9330605209}, \frac{12129669559}{9330605209}\right]$$
$$Q^b_{r2} = \left[\frac{18005811535}{14928733909}, \frac{12129669559}{14928733909}\right]$$
$$Q^b_{r3} = \left[\frac{18005811535}{45342228279}, \frac{12129669559}{45342228279}\right].$$

The respective blue/green quad points Q^b_{gj} *are*

$$Q^b_{g0} = \left[-\frac{4161500}{2654777}, \frac{11762777}{2654777}\right]$$
$$Q^b_{g1} = \left[-\frac{4161500}{12547777}, \frac{11762777}{12547777}\right]$$
$$Q^b_{g2} = \left[\frac{4161500}{10977777}, \frac{11762777}{10977777}\right]$$
$$Q^b_{g3} = \left[\frac{4161500}{20870777}, \frac{11762777}{20870777}\right].$$

The blue star lines are then

$$s^b_0 = Q^b_{r0}Q^b_{g0} =$$
$$\langle 1796063533088 : 868804574039 : -1034074074039\rangle$$
$$s^b_1 =$$
$$\langle 272084614990 : 1199343574039 : -1034074074039\rangle$$
$$s^b_2 =$$
$$\langle 272084614990 : 868804574039 : -1034074074039\rangle$$
$$s^b_3 =$$
$$\langle 1796063533088 : 1199343574039 : -1034074074039\rangle$$

and they meet at the blue star points

$$B_{01} = \left[\frac{165269500000}{927258959049}, \frac{761989459049}{927258959049}\right]$$
$$B_{23} = \left[-\frac{165269500000}{596719959049}, \frac{761989459049}{596719959049}\right]$$
$$B_{02} = \left[0, \frac{1034074074039}{868804574039}\right], B_{13} = \left[0, \frac{1034074074039}{1199343574039}\right]$$
$$B_{03} = \left[\frac{1034074074039}{1796063533088}, 0\right], B_{12} = \left[\frac{1034074074039}{272084614990}, 0\right].$$

Note the pleasant rationality of the previous objects.

3.2 An example over $\mathbb{F}_{13}$

Now we look at an example over a finite field.

Theorem 9 (Null quadrances incenters) *Suppose that the field* $\mathbb{F}$ *contains an element* i*, where* $i^2 = -1$*, and the characteristic of* $\mathbb{F}$ *is not* 2*. If*

$$K_b \equiv \frac{b_b(a_b - b_b)(b_b - c_b)}{\Delta_b} = K_r \equiv \frac{b_r(a_r - b_r)(b_r - c_r)}{\Delta_r}$$
$$= K_g \equiv \frac{b_g(a_g - b_g)(b_g - c_g)}{\Delta_g} = 0$$

then the standard Triangle $\overline{A_1A_2A_3}$ *has four distinct blue, red and green Incenters.*

Proof. If $K_b = 0$ then from the definition of the blue incenter circle $\mathcal{C}_b$, which is $Q_b(H_b, X) = K_b$, $\mathcal{C}_b$ is a null circle, so it is a product of lines. Similarly, if $K_r = 0$ then $\mathcal{C}_r$ is a null circle, and if $K_g = 0$ then $\mathcal{C}_g$ is a null circle. These null lines have distinct direction vectors $(1, \pm i), (1, \pm 1)$ and $(1,0), (0,1)$ respectively, and they are never parallel since $\text{char}(\mathbb{F}) \neq 2$, so $i \neq \pm 1$. Therefore, any two null circles meet in exactly four points. □

Here is an example found by Michael Reynolds [6] which illustrates explicitly the above theorem.

Example 2 *The triangle* $\overline{X_1X_2X_3}$ *with points* $X_1 \equiv [3,4], X_2 \equiv [1,9]$ *and* $X_3 \equiv [12,3]$ *in* $\mathbb{F}_{13}$ *has four blue, red and green Incenters. In* $\mathbb{F}_{13}$ *the squares are* $0, 1, 3, 4, 9, 10$ *and* 12*, and in particular* $-1 = 12 = 5^2$ *is a square. After translation by* $(3,4)$ *we obtain* $\widetilde{X}_1 = [0,0], \widetilde{X}_2 = [11,5]$ *and* $\widetilde{X}_3 = [9,12]$*. The matrix N and its inverse M*

$$N = \begin{pmatrix} 11 & 5 \\ 9 & 12 \end{pmatrix}, \qquad M = N^{-1} = \begin{pmatrix} 10 & 11 \\ 12 & 7 \end{pmatrix}$$

send $[1,0]$ *and* $[0,1]$ *to* $\widetilde{X}_2$ *and* $\widetilde{X}_3$*, and* $\widetilde{X}_2$ *and* $\widetilde{X}_3$ *to* $[1,0]$ *and* $[0,1]$ *respectively. The bilinear form in these new standard coordinates for the blue, red and green geometries*

respectively are

$$D_b = N\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}N^T = \begin{pmatrix}3 & 3\\ 3 & 4\end{pmatrix}$$
$$D_r = N\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}N^T = \begin{pmatrix}5 & 0\\ 0 & 2\end{pmatrix}$$
$$D_g = N\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}N^T = \begin{pmatrix}6 & 8\\ 8 & 8\end{pmatrix}.$$

We can see immediately that $K_b = K_r = K_g = 0$ from the definitions

$$Q_b(H_b, I_{ir}) = \frac{b_b(a_b - b_b)(b_b - c_b)}{\Delta_b} \equiv K_b$$
$$Q_r(H_r, I_{ib}) = \frac{b_r(a_r - b_r)(b_r - c_r)}{\Delta_r} \equiv K_r$$
$$Q_g(H_g, I_{ir}) = \frac{b_g(a_g - b_g)(b_g - c_g)}{\Delta_g} \equiv K_g, \quad i = 0,1,2,3.$$

The four blue, red and green Incenters respectively are

$$I_0^b = [4,8], \quad I_1^b = [3,6], \quad I_2^b = [8,10], \quad I_3^b = [11,4]$$
$$I_0^r = [10,9], \quad I_1^r = [8,2], \quad I_2^r = [6,5], \quad I_3^r = [4,12]$$
$$I_0^g = [9,8], \quad I_1^g = [5,3], \quad I_2^g = [12,11], \quad I_3^g = [2,4]$$

and the blue, red and green Incenter Circles respectively have equations

$$\mathcal{C}_b : (y - x + 1)(x + 3y - 1) = 0$$
$$\mathcal{C}_r : (x - 6y)(x + 6y) = 0$$
$$\mathcal{C}_g : (x + 2y - 2)(x + 5y - 5) = 0.$$

From Michael Reynolds' computer investigations, we tentatively conjecture that for finite fields $\mathbb{F}_p$ where $p \equiv 3 \bmod 4$, there are *no* triangles which have both blue, red and green Incenters, and for finite fields $\mathbb{F}_p$ where $p \equiv 1 \bmod 4$, blue, red and green Incenters exists precisely when $K_b = K_r = K_g = 0$, as in the above example.

4 Spieker circles and Nagel circles

Now we recall from [5] that the central dilation $\delta_{-1/2}$ about the centroid takes the Orthocenter to the Circumcenter, and the Incenters to the *Spieker centers*. In standard coordinates

$$\delta_{-1/2}([x,y]) = (1/2)[1 - x, 1 - y].$$

The inverse central dilation δ_{-2} takes the Orthocenter to the *De Longchamps point* X_{20}, and takes the Incenters to the *Nagel points*. In standard coordinates

$$\delta_{-2}([x,y]) = [1 - 2x, 1 - 2y].$$

Theorem 10 (Spieker circles) *If a triangle has four blue Incenters I_0^b, I_1^b, I_2^b and I_3^b, then the four blue Spieker centers all lie both on a red Spieker circle with center the red Circumcenter C_r, and on a green Spieker circle with center the green Circumcenter C_g. If both say blue and red Incenters exist, then all 8 blue and red Spieker points lie on the same green circle. The same holds for the other colours.*

Proof. We see that if we use the central dilation formula to transform Incenter circles centred at the Orthocenters, we get the Spieker circles centred at Circumcenters, so this theorem is a direct consequence of the Incenter circles theorem and the fact that a dilation preserves circles of any colour. □

Here are the formulas for the coloured Circumcenters in standard coordinates:

$$C_b = \frac{1}{2\Delta_b}[c_b(a_b - b_b), a_b(c_b - b_b)]$$
$$C_r = \frac{1}{2\Delta_r}[c_r(a_r - b_r), a_r(c_r - b_r)]$$
$$C_g = \frac{1}{2\Delta_g}[c_g(a_g - b_g), a_g(c_g - b_g)].$$

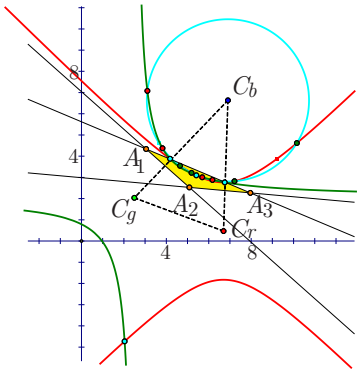


Figure 8: *Blue, red and green Speiker circles*

Theorem 11 (Nagel circles) *If a triangle has four blue Incenters I_0^b, I_1^b, I_2^b and I_3^b, then the four blue Nagel centers all lie both on a red Nagel circle with center the red De Longchamps point X_{20r}, and on a green Nagel circle with center the green De Longchamps point X_{20g}. If both say blue and red Incenters exist, then all 8 blue and red Nagel points lie on the same green circle. The same holds for the other colours.*

Proof. In the same fashion as in the previous theorem, if we use the inverse central dilation δ_{-2} to transform Incenter circles centred at the Orthocenters, we get the Nagel circles centred at De Longchamps points. □

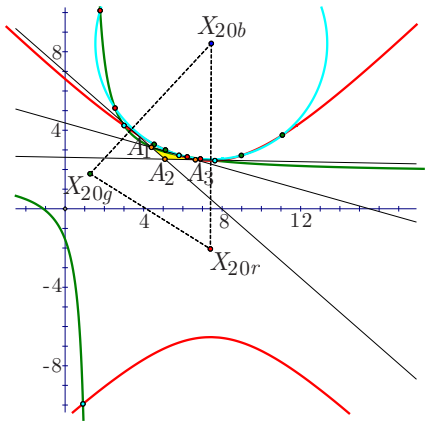


Figure 9: *Blue, red and green Nagel circles*

Here are the formulas for the blue, red and green De Longchamps points:

$$X_{20b} = \frac{1}{\Delta_b}\left[b_b^2 - 2b_bc_b + a_bc_b, b_b^2 - 2a_bb_b + a_bc_b\right]$$

$$X_{20r} = \frac{1}{\Delta_r}\left[b_r^2 - 2b_rc_r + a_rc_r, b_r^2 - 2a_rb_r + a_rc_r\right]$$

$$X_{20g} = \frac{1}{\Delta_g}\left[b_g^2 - 2b_gc_g + a_gc_g, b_g^2 - 2a_gb_g + a_gc_g\right].$$

In Figure 10 we see the relations between the three coloured Orthocenters, Circumcenters and De Longchamps points. The lines joining these are the three coloured Euler lines. Note that the centroids of the triangles of Orthocenters, Circumcenters and De Longchamps points all agree with the centroid G of the original triangle $\overline{A_1A_2A_3}$. We conclude with a simple observation about De Longchamps points.

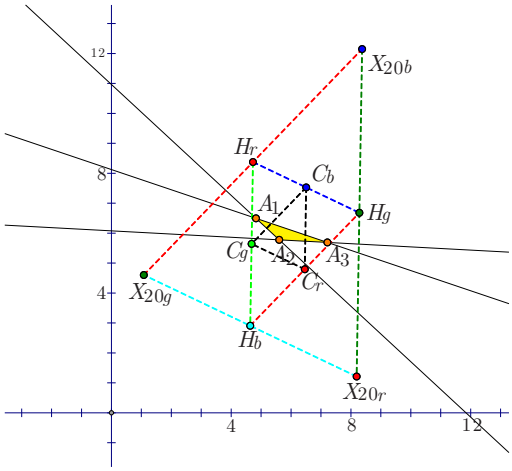


Figure 10: *Blue, red, green Orthocenters, Circumcenters and De Longchamps points*

Theorem 12 (Orthocenters as midpoints) *For any triangle, a coloured orthocenter H is the midpoint of the two De Longchamps points X_{20} of the other two colours.*

Proof. This follows by considering the action of the central dilation δ_{-2} which takes the circumcenter C_i to the orthocenter H_i, and the orthocenter H_i to the De Longchamps point X_{20i}. □

References

[1] M. W. HASKELL, The Construction of Conics under given conditions, *Bulletin of the Amer. Math. Soc.* **11 (5)** (1905), 268–273.

[2] C. KIMBERLING, *Triangle Centers and Central Triangles*, vol. 129 Congressus Numerantium, Utilitas Mathematica Publishing, Winnepeg, MA, 1998.

[3] C. KIMBERLING, *Encyclopedia of Triangle Centers*, `http://faculty.evansville.edu/ck6/encyclopedia/ETC.html`.

[4] C. KIMBERLING, Major Centers of Triangles, *Amer. Math. Monthly* **104** (1997), 431–438.

[5] N. LE, N. J. WILDBERGER, Universal Affine Triangle Geometry and Four-fold Incenter Symmetry, *KoG* **16** (2012), 63–80.

[6] M. REYNOLDS, *Personal Communication*, 2013.

[7] N. J. WILDBERGER, *Divine Proportions: Rational Trigonometry to Universal Geometry*, Wild Egg Books, Sydney, 2005.

[8] N. J. WILDBERGER, Affine and projective metrical geometry, *arXiv: math/0612499v1*.

[9] N. J. WILDBERGER, Chromogeometry, *Mathematical Intelligencer* **32 (1)** (2010), 26–32.

[10] N. J. WILDBERGER, Chromogeometry and Relativistic Conics, *KoG* **13** (2009), 43–50.

[11] B. M. WOODS, The Construction of Conics under given conditions, *Amer. Math. Monthly* **21 (6)** (1914), 173–180.

Nguyen Le
e-mail: n.h.le@unsw.edu.au
Norman John Wildberger
e-mail: n.wildberger@unsw.edu.au
School of Mathematics and Statistics UNSW
Sydney 2052 Australia

Original scientific paper
Accepted 20. 10. 2014.

HIROSHI OKUMURA

The Arbelos with Overhang

The Arbelos with Overhang

ABSTRACT

We consider a generalized arbelos consisting of three semicircles with collinear centers, in which only two of the three semicircles touch. Many Archimedean circles of the ordinary arbelos are generalized to our generalized arbelos.

Key words: arbelos, arbelos with overhang, Archimedean circles

MSC2010: 51M04, 51N20

Arbelos s privjeskom

SAŽETAK

Promatra se poopćeni arbelos koji se sastoji od tri polukružnice s kolinearnim središtima, pri čemu se dvije on njih dodiruju. Mnoge Arhimedove kružnice obićnog arbelosa su poopćene za poopćeni arbelos.

Ključne riječi: arbelos, arbelos s privjeskom, Arhimedove kružnice

1 Introduction

The *arbelos* is a plane figure consisting of three mutually touching semicircles with collinear centers. It has three points of tangency. In [5], [7] and [9], we have considered a generalized arbelos called a *collinear arbelos* consisting of three circles with collinear centers, in which one of the circles touches the remaining two circles, but the two circles do not touch in general. Thereby the collinear arbelos has two points of tangency.

In this paper, we consider the remaining case. We consider a configuration consisting of three semicircles with collinear centers, in which only two semicircles touch, i.e., it has only one point of tangency. Many Archimedean circles of the ordinary arbelos are generalized to our generalized arbelos, but also several new Archimedean circles of the ordinary arbelos are induced by this.

2 An arbelos with overhang

Let O be a point on the segment AB with $|AO| = 2a$ and $|BO| = 2b$. We use a rectangular coordinate system with origin O such that the coordinates of the points A and B are $(2a,0)$ and $(-2b,0)$, respectively. For two points P and Q, (PQ) and $P(Q)$ denote the circle with diameter PQ and the circle with center P passing through Q, respectively. However if their centers lie on the line AB, we consider them as semicircles lying in the region $y \geq 0$ unless otherwise stated. Let A' (resp. B') be a point on the half line with endpoint O passing through A (resp. B), and let $|A'O| = 2a'$ (resp. $|B'O| = 2b'$) (see Figure 1). Let $\gamma = (AB)$, and let δ'_α be the circle touching the semicircle $(A'O)$ externally γ internally and the perpendicular to AB passing through O from the side opposite to the point B. The circle δ'_β is defined similarly.

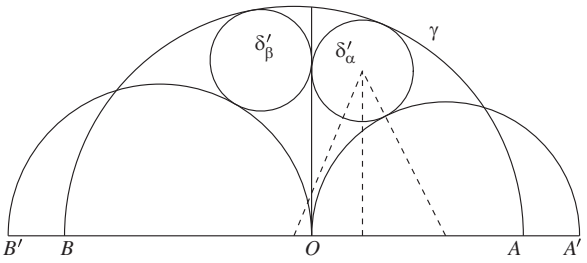


Figure 1

Proposition 1 *The two circles δ'_α and δ'_β are congruent if and only if $a' - a = b' - b$.*

Proof. Let r be the radius of δ'_α. The center of the circle with a diameter $A'O$ or AB, the center of $\delta_{\alpha'}$ and the foot of perpendicular from this point to AB form a right triangle. Hence by the Pythagorean theorem, we get

$$(r+a')^2 - (r-a')^2 = ((a+b)-r)^2 - (r-(a-b))^2.$$

Notice that if the foot of perpendicular coincides with the center of the circle with a diameter $A'O$ or AB, then one of the triangles degenerates to a segment. But the equation still holds. Solving the equation we get $r = ab/(a'+b)$. Similarly, δ'_β has radius $ab/(a+b')$. Therefore the two circles are congruent if and only if $a'+b = a+b'$. □

Let $\alpha = (AO)$, $\beta = (BO)$, and let $a' = a+h$, $b' = b+h$ with $-\min(a,b) < h$. We relabel a', b', A', B', δ'_α and δ'_β as a_h, b_h, A_h, B_h, δ_h^α and δ_h^β, respectively and let $\alpha_h = (A_hO)$ and $\beta_h = (B_hO)$. The configuration consisting of the three semicircles α_h, β_h and γ is denoted by $(\alpha_h,\beta_h,\gamma)$. We call $(\alpha_h,\beta_h,\gamma)$ an *arbelos with overhang h*, and $(\alpha_h,\beta_h,\gamma)$ is said to have overhang h. The ordinary arbelos (α,β,γ) has overhang 0. The perpendicular to AB passing through O is called the axis, which overlaps with the y-axis.

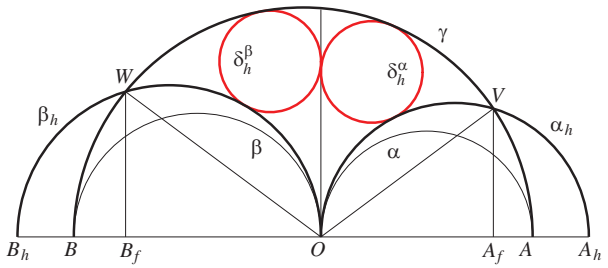


Figure 2

Now the circles δ_h^α and δ_h^β have the same radius $r_{\mathrm{A}}^{\mathrm{h}} = ab/(a+b+h)$ by Proposition 1. The two circles are a generalization of the twin circles of Archimedes of the ordinary arbelos (α,β,γ). Circles of radius $r_{\mathrm{A}}^{\mathrm{h}}$ are said to be Archimedean circles of $(\alpha_h,\beta_h,\gamma)$ or Archimedean with respect to $(\alpha_h,\beta_h,\gamma)$. Also we say that $(\alpha_h,\beta_h,\gamma)$ has Archimedean circles of radius $r_{\mathrm{A}}^{\mathrm{h}}$. The common radius of Archimedean circles of (α,β,γ) is denoted by r_{A}, i.e., $r_{\mathrm{A}} = ab/(a+b)$.

We define A_f and B_f as the points with coordinates $(2ab/b_h,0)$ and $(-2ab/a_h,0)$, respectively. Let γ have points V and W in common with the semicircles α_h and β_h respectively in the case $h \geq 0$ (see Figure 2). The points V and W have coordinates

$$(2ab/b_h, f(a,b)/b_h) \text{ and } (-2ab/a_h, f(a,b)/a_h),$$

respectively, where $f(a,b) = 2\sqrt{abh(a+b+h)}$. Therefore the points A_f and B_f are the feet of perpendiculars from V and W to the line AB, respectively. By the coordinates of V and W, we get $\tan\angle WOB = \tan\angle VOA$. Therefore $\angle WOB = \angle VOA$ holds.

The circle touching γ internally and the segment AB at the point O has radius $2r_{\mathrm{A}}$ [11]. The fact is generalized as follows.

Proposition 2 *If $h > 0$, the radius of the circle touching γ internally and the segments OV and OW is $2r_{\mathrm{A}}^{\mathrm{h}}$.*

Proof. Let r and $(0,c)$ be the radius and the coordinates of the center of the touching circle. Then we get

$$(a-b)^2 + c^2 = (a+b-r)^2. \tag{1}$$

Also by similar triangles, we get

$$\frac{r}{c} = \cos\angle VOA = \sqrt{\frac{ab}{(a+h)(b+h)}}. \tag{2}$$

Eliminating c from (1) and (2), and solving the resulting equation for r with $h > 0$, we get $r = 2r_{\mathrm{A}}^{\mathrm{h}}$. □

Let $\alpha_f = (A_fO)$ and $\beta_f = (B_fO)$. Archimedean circles of the ordinary arbelos $(\alpha_f,\beta,(A_fB))$ have radius

$$\frac{(ab/b_h)b}{ab/b_h+b} = \frac{ab}{a+b_h} = r_{\mathrm{A}}^{\mathrm{h}}.$$

Similarly Archimedean circles of the ordinary arbelos $(\alpha,\beta_f,(AB_f))$ have the same radius. Hence we get:

Proposition 3 *The ordinary arbeloi $(\alpha_f,\beta,(A_fB))$ and $(\alpha,\beta_f,(AB_f))$ share Archimedean circles with $(\alpha_h,\beta_h,\gamma)$.*

The circle touching the axis at the point O from the side opposite to the point B and also touching the tangent of β from the point A is an Archimedean circle of the ordinary arbelos (α,β,γ), which is denoted by W_6 in [4]. Hence by Proposition 3, we get the following proposition. By this proposition we can construct the point A_f (also B_f) even in the case $h < 0$ (see Figures 7 and 14).

Proposition 4 *The point A_f coincides with the point of intersection of the line AB and the external common tangent of β and the Archimedean circle of $(\alpha_h,\beta_h,\gamma)$ touching the axis at the point O from the side opposite to the point B.*

Since $|AB_f| : |AB_h| = a : a_h$ holds, we get the following proposition, which also enable us to construct the points A_f and B_f in the case $h < 0$.

Proposition 5 *The point B_f divides the segment AB_h in the ratio $a : |h|$ internally or externally, according as $h > 0$ or $h < 0$.*

3 Several twin circles

In this section we show that several twin circles exist for $(\alpha_h,\beta_h,\gamma)$, if $h > 0$. Let us assume $h > 0$, and let ε_1^α be the circle touching the semicircles α externally α_h internally and the segment A_fV from the side opposite to the point A (see Figure 3). Let ε_2^α be the circle touching the semicircles α externally α_h and γ internally. Also let ε_3^α be the circle touching α_h and γ externally and the axis from the side opposite to the point B. The circles ε_1^β, ε_2^β and ε_3^β are defined similarly. The following proposition has a straightforward proof that is omitted.

Proposition 6 *If $h > 0$, the following statements hold.*
(i) *The circles ε_1^α and ε_1^β have the same radius*

$$\left(\frac{1}{a}+\frac{1}{b}+\frac{h}{ab}+\frac{1}{h}\right)^{-1} = \left(\frac{1}{r_{\mathrm{A}}^{\mathrm{h}}}+\frac{1}{h}\right)^{-1}.$$

(ii) *The circles ε_2^α and ε_2^β have the same radius*

$$\left(\frac{1}{a}+\frac{1}{b}+\frac{1}{h}\right)^{-1} = \left(\frac{1}{r_{\mathrm{A}}}+\frac{1}{h}\right)^{-1}.$$

(iii) *The circles ε_3^α and ε_3^β have the same radius ab/h.*

The proposition also shows that the sum of the curvatures of the circles ε_2^α and ε_3^α equals the curvature of the circle ε_1^α.

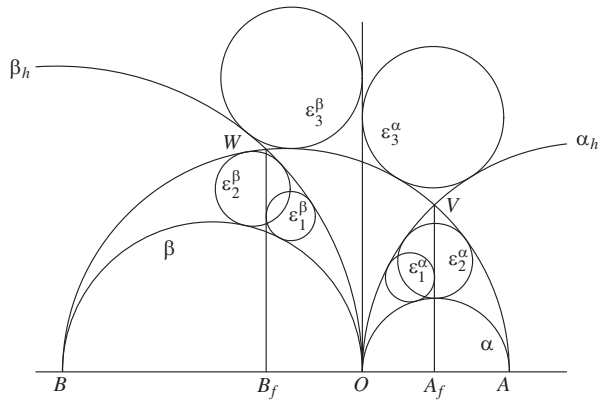


Figure 3

4 Bankoff circles

The circle orthogonal to the semicircles α, β and to the circle touching α and β externally and γ internally is an Archimedean circle of (α,β,γ) called the Bankoff triplet circle, which is denoted by W_3 in [4]. The maximal circle touching the external common tangent of α and β and the arc of γ cut by the tangent internally is an Archimedean circle of (α,β,γ) called the Bankoff quadruplet circle, which is denoted by W_4 in [4]. In this section we generalize the two circles (see Figures 4 and 6). Let $\gamma_f = (A_f B_f)$.

Theorem 1 *The following two circles are Archimedean with respect to $(\alpha_h,\beta_h,\gamma)$, and coincide.*
(i) *The circle orthogonal to the semicircles α, β and to the circle touching α and β externally and γ_f internally.*
(ii) *The circle orthogonal to the semicircles α_f, β_f and to the circle touching α_f and β_f externally and γ internally.*

Proof. Let δ be the circle touching α and β externally and γ_f internally, and let ε be the circle denoted by (i). We invert the figure in the circle with center O and radius $2\sqrt{ab}$, and label the images with an overline (see Figure 5). The x-coordinates of the points $\overline{A}$, $\overline{B}$, $\overline{A_f}$ and $\overline{B_f}$ are $2b$, $-2a$, $2b_h$ and $-2a_h$, respectively, and the circle $\overline{\gamma_f} = (\overline{A_f}\,\overline{B_f})$ has center $(b-a,0)$ and radius a_h+b_h. Let $(x_{\overline{\delta}}, y_{\overline{\delta}})$ and $r_{\overline{\delta}}$ be the coordinates of the center of the circle $\overline{\delta}$ and its radius. The circle $\overline{\delta}$ touches $\overline{\alpha}$ and $\overline{\beta}$, which are the perpendiculars to AB passing through the points $\overline{A}$ and $\overline{B}$, respectively. Therefore we get $x_{\overline{\delta}} = b-a$ and $r_{\overline{\delta}} = a+b$. Since $\overline{\delta}$ touches $\overline{\gamma_f}$ externally and the x-coordinates of their centers are the same, we get

$$y_{\overline{\delta}} = a_h + b_h + r_{\overline{\delta}} = 2(a+b+h).$$

Since $\overline{\varepsilon}$ is the line perpendicular to the line $\overline{\alpha}$ and passes through the point of tangency of $\overline{\alpha}$ and $\overline{\delta}$, it is parallel to AB and passes through the center of $\overline{\delta}$. Hence the distance between AB and the farthest point on ε equals $4ab/y_{\overline{\delta}} = 2r_{\mathrm{A}}^{\mathrm{h}}$. Therefore ε is Archimedean with respect to $(\alpha_h,\beta_h,\gamma)$. The part (ii) is proved similarly. □

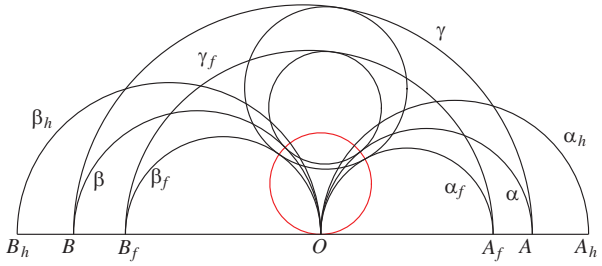


Figure 4

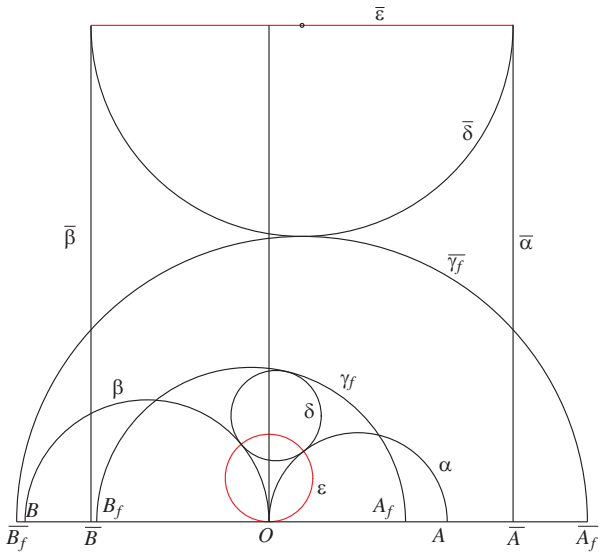


Figure 5

We call the circle in Theorem 1 the Bankoff triplet circle of $(\alpha_h,\beta_h,\gamma)$.

Theorem 2 *If $\mathcal{E}$ is the external common tangent of the semicircles α and β_h, or α_h and β, then the maximal circle touching $\mathcal{E}$ and the arc of γ cut by $\mathcal{E}$ (the part of γ between the two points of intersection of γ and $\mathcal{E}$) internally is Archimedean with respect to $(\alpha_h, \beta_h, \gamma)$.*

Proof. We prove the case $\mathcal{E}$ being the common tangent of α and β_h. The other case is proved similarly. Let d be the distance between $\mathcal{E}$ and the centers of γ, and let T be the point of intersection of the lines $\mathcal{E}$ and AB. If T lies in the region $x > 0$, let $|AT| = t$. By similar triangles, we get

$$a/(t+a) = d/(t+a+b) = b_h/(t+2a+b_h).$$

Eliminating t and solving the resulting equations for d, we get $d = a+b-2r_{\mathrm{A}}^{\mathrm{h}}$. Therefore δ is an Archimedean circle of $(\alpha_h, \beta_h, \gamma)$. The case T lying in the region $x < 0$ is proved similarly. If $\mathcal{E}$ and AB are parallel, then $a = b+h = d$ and $r_{\mathrm{A}}^{\mathrm{h}} = ab/(2a) = b/2$. Therefore we also get $d = a+b-2r_{\mathrm{A}}^{\mathrm{h}}$. □

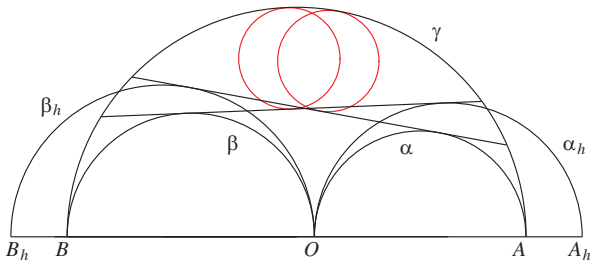


Figure 6

5 Miscellaneous Archimedean circles

In this section we consider miscellaneous Archimedean circles of $(\alpha_h, \beta_h, \gamma)$ obtained from points dividing given segments in the ratio $a : b_h$ or $a_h : b$ internally, some of which seem to be new even for the ordinary arbelos. Let I be the point of intersection of the axis and the semicircle γ. The minimal circle touching the axis and passing through the point of intersection of the semicircle α and the segment AI is an Archimedean circle of (α, β, γ), which is denoted by W_9 in [4]. Also the minimal circle touching the axis and passing through the point of intersection of the semicircles γ and $A(O)$ is an Archimedean circle of (α, β, γ), which is denoted by W_{13} in [4]. The two facts are generalized. Let A_m and B_m be the midpoints of the segments AA_h and BB_h, respectively (see Figure 7).

Theorem 3 (i) *If I_α is the point of intersection of the axis and the semicircle (A_hB), then the point dividing the segment A_fI_α in the ratio $a : b_h$ internally lies on the semicircle α and its distance from the axis is $2r_{\mathrm{A}}^{\mathrm{h}}$.*
(ii) *The distance between the axis and the point of intersection of the semicircles $A_m(O)$ and γ is $2r_{\mathrm{A}}^{\mathrm{h}}$.*

Proof. Since I_α has coordinates $\left(0, 2\sqrt{a_hb}\right)$, the point dividing the segment A_fI_α in (i) has coordinates

$$\left(\frac{b_h \cdot 2ab/b_h}{a+b_h}, \frac{a \cdot 2\sqrt{a_hb}}{a+b_h}\right) = \left(2r_{\mathrm{A}}^{\mathrm{h}}, 2r_{\mathrm{A}}^{\mathrm{h}}\sqrt{\frac{a_h}{b}}\right).$$

This proves (i). The point of intersection of γ and $A_m(O)$ has coordinates $\left(2r_{\mathrm{A}}^{\mathrm{h}}, 2\sqrt{(a-r_{\mathrm{A}}^{\mathrm{h}})(b+r_{\mathrm{A}}^{\mathrm{h}})}\right)$. This proves (ii). □

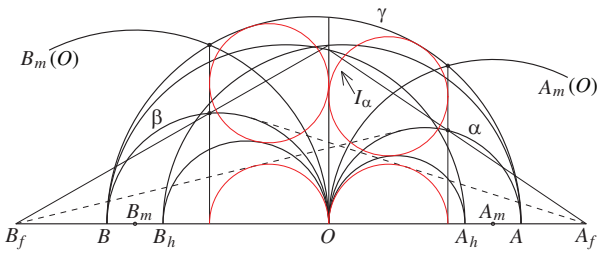


Figure 7

For a circle or a semicircle δ, its center is denoted by O_δ. The farthest point on δ from AB lying in the region $y \geq 0$ is denoted by T_δ. If the segments $T_\alpha T_\beta$ and $T_\gamma O_\gamma$ intersect at a point P, the circle (PT_γ) is an Archimedean circle of (α, β, γ), which is denoted by W_{20} in [4]. The fact is generalized (see Figure 8).

Theorem 4 *The segments $T_{\alpha_h}T_\beta$, $T_\alpha T_{\beta_h}$ and $T_\gamma O_\gamma$ intersect at a point P, which divides $T_{\alpha_h}T_\beta$ and $T_\alpha T_{\beta_h}$ in the ratios $b_h : a$ and $b : a_h$ internally, respectively. The circle (PT_γ) is Archimedean with respect to $(\alpha_h, \beta_h, \gamma)$.*

Proof. The points dividing $T_{\alpha_h}T_\beta$ in the ratio $b_h : a$ internally and $T_\alpha T_{\beta_h}$ in the ratio $b : a_h$ internally have the same coordinates $(a-b, a+b-2r_{\mathrm{A}}^{\mathrm{h}})$. □

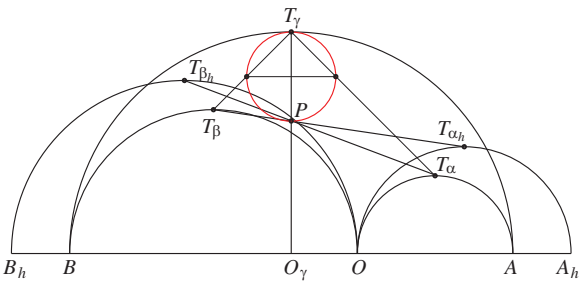


Figure 8

In the theorem, the endpoints of the diameter of (PT_γ) parallel to AB divide the segments $T_\gamma T_\alpha$ and $T_\gamma T_\beta$ in the ratios $a : b_h$ and $b : a_h$ internally, respectively.

Theorem 5 *Let T'_{α_h} and T'_β be the reflected images of the points T_{α_h} and T_β in the line AB, respectively. The following statements hold.*

(i) *Let C be the internal center of similitude of the semicircles γ and α_h. If D is the point of intersection of the lines CT'_β and AT_α, then D divides AT_α in the ratio $b : a_h$ internally and the circle touching α or AB at the point A and passing through D is Archimedean with respect to $(\alpha_h, \beta_h, \gamma)$.*
(ii) *The segments $A_f T_\beta$, AT_{β_f} and $T_\alpha O$ intersect at a point H, which divides $A_f T_\beta$ in the ratio $a : b_h$ internally AT_{β_f} and $T_\alpha O$ in the ratio $a_h : b$ internally, respectively. The circle touching α or AB at the point O and passing through H is Archimedean with respect to $(\alpha_h, \beta_h, \gamma)$.*
(iii) *If E is the point of intersection of the segments $T'_{\alpha_h} T_\gamma$ and OT_α, then E divides $T'_{\alpha_h} T_\gamma$ and OT_α in the ratio $a_h : b$ internally and the circle touching α or the line $O_\alpha T_\alpha$ at the point T_α and passing through E is Archimedean with respect to $(\alpha_h, \beta_h, \gamma)$.*

Proof. The point C has coordinates $(2aa_h/(2a+b_h), 0)$ (see Figure 9). If we regard β as a circle, then C coincides with the internal center of similitude of β and the Archimedean circle of $(\alpha_h, \beta_h, \gamma)$ touching α at the point A internally. Hence D has coordinates $(2a - r_A^h, r_A^h)$. This proves (i). The points dividing $A_f T_\beta$ in the ratio $a : b_h$ internally, AT_{β_f} and $T_\alpha O$ in the ratio $a_h : b$ internally have the same coordinates (r_A^h, r_A^h). This proves (ii). The points dividing OT_α and $T'_{\alpha_h} T_\gamma$ in the ratio $a_h : b$ internally have the same coordinates $(a - r_A^h, a - r_A^h)$. This proves (iii). □

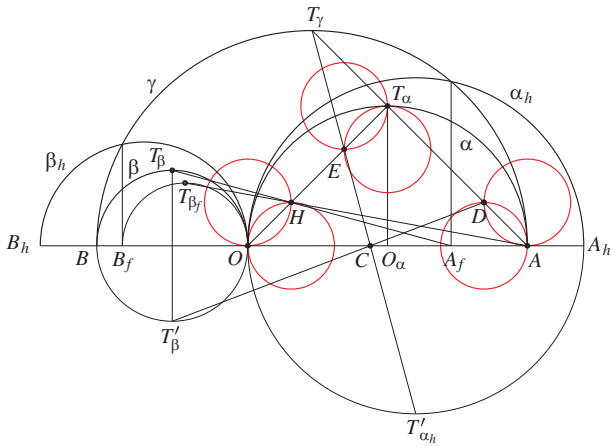


Figure 9

The circle touching AB at O and passing through H in (ii) is the Bankoff triplet circle of $(\alpha_h, \beta_h, \gamma)$.

6 Archimedean circles touching γ

In [12], we gave necessary and sufficient conditions that a circle touching the semicircle γ internally is Archimedean with respect to (α, β, γ). In this section we generalize this.

[1]The notations are slightly changed from [12]

Let $\alpha(z)$ and $\beta(z)$ be the semicircles constructed in the region $y \geq 0$ touching the axis at the point O and having the centers with coordinates $(za, 0)$ and $(-zb, 0)$, respectively for a real number z[1]. Let $\mathcal{C}(m,n)$ be the circle touching the semicircles γ internally and $\alpha(m)$ and $\beta(n)$ at points different from O such that the points of tangency on $\alpha(m)$, γ and $\beta(n)$ lie counterclockwise in this order for real numbers m and n. The radius of $\mathcal{C}(m,n)$ is expressed as follows [12, Theorem 1]:

$$\frac{ab(ma+nb)}{ma^2+nb^2+mnab}. \tag{3}$$

Let $\alpha_h(z)$ and $\beta_h(z)$ be the semicircles constructed in the region $y \geq 0$ touching the axis at the point O and having the centers with coordinates $(za_h, 0)$ and $(-zb_h, 0)$, respectively for a real number z. Let $\mathcal{C}_h(m,n)$ be the circle touching the semicircles γ internally and $\alpha_h(m)$ and $\beta_h(n)$ at points different from O such that the points of tangency on $\alpha_h(m)$, γ and $\beta_h(n)$ lie counterclockwise in this order for real numbers m and n.

Theorem 6 *The circle $\mathcal{C}_h(m,n)$ has radius*

$$\frac{ab(ma_h+nb_h)}{maa_h+nbb_h+mna_hb_h}. \tag{4}$$

Proof. Notice that $\alpha_h(m) = \alpha(ma_h/a)$ and $\beta_h(n) = \beta(nb_h/b)$. Replacing m and n by ma_h/a and nb_h/b respectively in (3), we get (4). □

Theorem 7 *The circle $\mathcal{C}_h(m,n)$ is Archimedean with respect to $(\alpha_h, \beta_h, \gamma)$ if and only if*

$$\frac{1}{m} + \frac{1}{n} = 1. \tag{5}$$

Proof. The theorem follows from Theorem 6, because

$$\begin{aligned} &\frac{ab(ma_h+nb_h)}{maa_h+nbb_h+mna_hb_h} - r_A^h \\ = \ &\frac{(m+n-mn)a_hb_hr_A^h}{maa_h+nbb_h+mna_hb_h}. \end{aligned}$$

□

Corollary 1 *The following circles are Archimedean with respect to $(\alpha_h, \beta_h, \gamma)$.*
(i) *The circle touching the semicircles $A_h(O)$ and $B_h(O)$ externally and γ internally.*
(ii) *The circle touching γ internally and the two distinct circles of radius $a_h + b_h$ touching the axis at the point O externally.*

Proof. The part (i) follows from the fact $A_h(O) = \alpha_h(2)$ and $B_h(O) = \beta_h(2)$ (see Figure 10). The part (ii) follows from the fact that $m = (a_h + b_h)/a_h$ and $n = (a_h + b_h)/b_h$ satisfy (5). □

The circle described in (i) is a generalization of Schoch circle of the ordinary arbelos which is denoted by W_{15} in [4]. The circle described in (ii) is a generalization of the Archimedean circle of the ordinary arbelos in [10].
Let $\gamma_h = (A_hB_h)$. If $m > 0$ (resp. $m < 0$), let $P_{\alpha_h}(m)$ be the external (resp. internal) center of similitude of the semicircles $\alpha_h(m)$ and γ_h. Similarly the point $P_{\beta_h}(m)$ is defined.

Theorem 8 *The points $P_{\alpha_h}(m)$ and $P_{\beta_h}(n)$ coincide if and only if* (5) *holds.*

Proof. The semicircles $\alpha_h(m)$ and γ_h have radii ma_h and a_h+b_h and centers with x-coordinates ma_h and a_h-b_h, respectively. Hence $P_{\alpha_h}(m)$ has x-coordinate

$$\frac{(a_h+b_h)ma_h - ma_h(a_h-b_h)}{-ma_h+(a_h+b_h)} = \frac{2ma_hb_h}{a_h+b_h-ma_h}.$$

Similarly $P_{\beta_h}(n)$ has x-coordinate

$$\frac{(a_h+b_h)(-nb_h) - nb_h(a_h-b_h)}{-nb_h+(a_h+b_h)} = \frac{-2na_hb_h}{a_h+b_h-nb_h}.$$

While

$$\frac{2ma_hb_h}{a_h+b_h-ma_h} - \frac{-2na_hb_h}{a_h+b_h-nb_h}$$

$$= \frac{2(m+n-mn)(a_h+b_h)a_hb_h}{(a_h+b_h-ma_h)(a_h+b_h-nb_h)}.$$

Therefore the proof is complete. □

Corollary 2 *The circle $\mathcal{C}_h(m,n)$ is Archimedean with respect to $(\alpha_h,\beta_h,\gamma)$ if and only if the points $P_{\alpha_h}(m)$ and $P_{\beta_h}(n)$ coincide.*

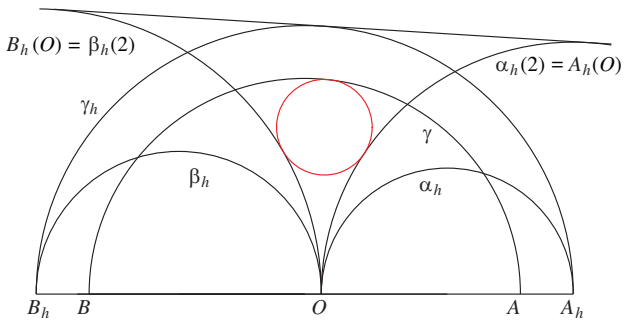


Figure 10: $m = n = 2$

If the external common tangents from γ_h to both $\alpha_h(m)$ and $\beta_h(n)$ exist, then the circle $\mathcal{C}_h(m,n)$ is Archimedean with respect to $(\alpha_h,\beta_h,\gamma)$ if and only if the two tangents coincide (see Figure 10).

7 Woo's Archimedean circles

We assume $a \neq b$ in this section. Peter Woo found that the circle touching $\alpha(z)$ and $\beta(z)$ externally with center on the line $x = (b-a)r_A/(b+a)$ is Archimedean with respect to (α,β,γ) for a non-negative real number z [4]. The line is called the Schoch line of (α,β,γ). The fact was generalized for a real number $z \geq -r_A/(a+b)$ in [12]. We generalize this.
A circle is said to touch $\alpha_h(z)$ appropriately if it touches $\alpha_h(z)$ externally in the case $z > 0$ and it touches the reflected image of $\alpha_h(z)$ in the line AB internally in the case $z < 0$. The same notion of appropriate tangency applies to $\beta_h(z)$. Let $s_h = (b_h-a_h)r_A^h/(b_h+a_h)$. We call the line $x = s_h$ the Schoch line of $(\alpha_h,\beta_h,\gamma)$.

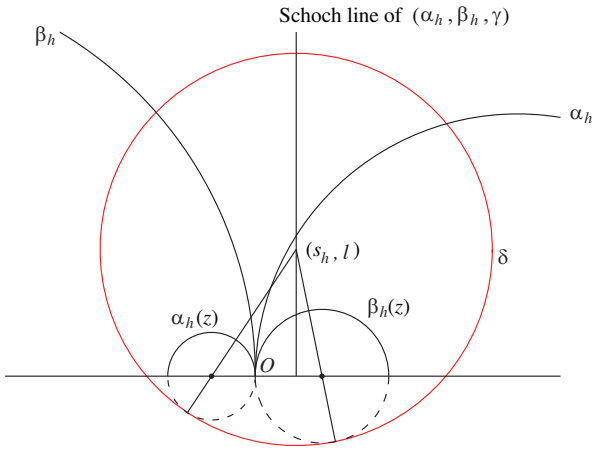


Figure 11: $z < 0$

Theorem 9 *Let δ be the circle touching $\alpha_h(z)$ and $\beta_h(z)$ appropriately and having its center on the Schoch line of $(\alpha_h,\beta_h,\gamma)$ for a real number $z \neq 0$. The following statements hold.*
(i) *The circle δ is Archimedean with respect to $(\alpha_h,\beta_h,\gamma)$.*
(ii) *The circle δ exists if and only if $-a_hb_h/(a_h+b_h)^2 \leq z < 0$ or $0 < z$.*

Proof. If r is the radius of δ and l is the y-coordinate of its center (see Figure 11), then we get

$$(za_h+r)^2 - (s_h-za_h)^2 = (zb_h+r)^2 - (s_h+zb_h)^2 = l^2. \quad (6)$$

Solving the equation for r, we get $r = s_h(b_h+a_h)/(b_h-a_h) = r_A^h$. This proves (i). From (6) we also get

$$l^2 = \frac{4a_hb_hs_h(s_h+(b_h-a_h)z)}{(a_h-b_h)^2}$$

$$= \frac{4a_h^2b_h^2(a_hb_h+(a_h+b_h)^2z)}{(a_h+b_h)^4}.$$

Therefore l satisfying (6) is real if and only if $-a_hb_h/(a_h+b_h)^2 \leq z < 0$ or $0 < z$. This proves (ii). □

Notice that the circle δ in Theorem 9 is not uniquely determined if $a = b$. We get an infinite set of Archimedean circles of $(\alpha_h, \beta_h, \gamma)$, whose centers lie on the line $x = s_h$ by the theorem. However the Archimedean circle of $(\alpha_h, \beta_h, \gamma)$ with center $(s_h, 2a_h b_h\sqrt{a_h b_h}/(a_h + b_h)^2)$ is not a member of this set. In fact, there are infinitely many circles passing through O with center on this line. But it seems to be natural to consider this circle as a member of this set.

8 Dilation

In [8], we have shown that if σ is a dilation with center O, the circle touching the semicircles $(A^\sigma O)$ externally (AB^σ) internally and the axis from the side opposite to the point B is an Archimedean circle of (α, β, γ), where A^σ and B^σ are the images of A and B by σ, respectively. In this section we generalize this fact.

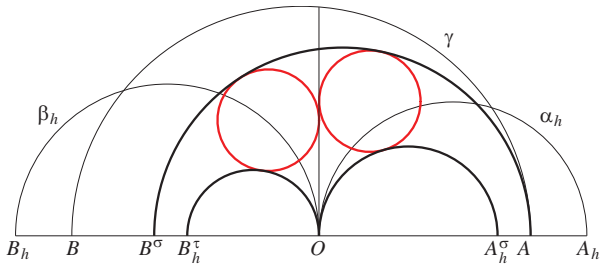


Figure 12: $t = 2/3$

Theorem 10 *Let σ and τ be the dilations with center O and A with the same ratio of magnification t, respectively. Then the following statements are true.*
(i) *The circle touching the semicircles $(A_h^\sigma O)$ externally (AB^σ) internally and the axis from the side opposite to the point B is Archimedean with respect to $(\alpha_h, \beta_h, \gamma)$.*
(ii) *If $t > a/(a+b_h)$, then $((A_h^\sigma O), (B_h^\tau O), (AB^\sigma))$ is an arbelos with overhang $ta_h - a$, and has Archimedean circles of radius $r_{\mathrm{A}}^{\mathrm{h}}$.*
(iii) *If $t = a/a_h$, then $((A_h^\sigma O), (B_h^\tau O), (AB^\sigma))$ coincides with $(\alpha, \beta_f, (AB_f))$, and the points B_h^τ and B^σ also coincide with the point B_f.*

Proof. Let r be the radius of the touching circle in (i) (see Figure 12). Then we get

$$(ta_h + r)^2 - (ta_h - r)^2 = (bt + a - r)^2 - ((-bt + a) - r)^2.$$

Solving the equation, we get $r = r_{\mathrm{A}}^{\mathrm{h}}$. This proves (i). If $t > a/(a+b_h)$, then $|AB_h^\tau| = 2t(a+b_h) > 2a = |AO|$. Hence the point B_h^τ lies on the half line with endpoint O passing through B. While $|A_h^\sigma O| - |AO| = 2(ta_h - a)$ and $|B_h^\tau O| - |B^\sigma O| = 2t(a+b_h) - 2(tb - a) = 2(ta_h - a)$. Hence the configuration $((A_h^\sigma O), (B_h^\tau O), (AB^\sigma))$ is an arbelos with overhang $ta_h - a$. The rest of (ii) follows from (i). If $t = a/a_h$, the points A and A_h^σ coincide, i.e., $(A_h^\sigma O) = \alpha$. While $a/a_h > a/(a_h + b) = a/(a + b_h)$ holds. Therefore we get an ordinary arbelos $(\alpha, (B_h^\tau O), (AB^\sigma))$, whose Archimedean circles have radius $r_{\mathrm{A}}^{\mathrm{h}}$ by (ii). While $(\alpha, \beta_f, (AB_f))$ is also an ordinary arbelos having Archimedean circles of the same radius by Proposition 3. Therefore the two ordinary arbeloi coincide. The rest of (iii) is obvious. □

9 New type of Archimedean circles

Quang Tuan Bui has found a pair of new type of Archimedean circles such that the endpoints of the diameter parallel to the line AB lie on a given circle [1], which has been rediscovered by us [6]. One of the circles is obtained as follows: If the line $T_\alpha O_\alpha$ intersects the semicircle γ at a point S and the lines SA and SO intersect the semicircle α at points T and U respectively, the circle (TU) is Archimedean with respect to (α, β, γ). The fact is generalized (see Figures 13 and 14). Notice that $h + r_{\mathrm{A}}^{\mathrm{h}} > 0$ and $a - r_{\mathrm{A}}^{\mathrm{h}} > 0$.

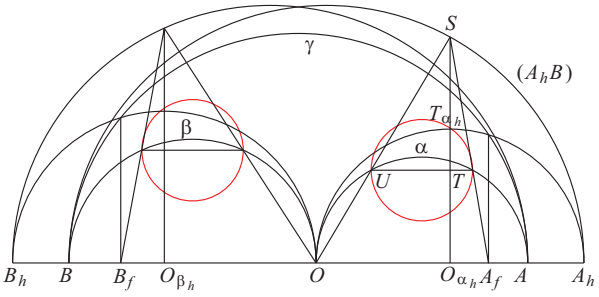


Figure 13

Theorem 11 (i) *Let S be the point of intersection of the semicircle (A_hB) and the line $T_{\alpha_h}O_{\alpha_h}$. If T is the point dividing the segment SA_f in the ratio $(h + r_{\mathrm{A}}^{\mathrm{h}}) : (a - r_{\mathrm{A}}^{\mathrm{h}})$ internally and U is the point of intersection of the line SO and the semicircle α, then T lies on α and the line TU is parallel to AB and the circle (TU) is Archimedean with respect to $(\alpha_h, \beta_h, \gamma)$.*
(ii) *Let S be the point of intersection of the semicircle (AB_h) and the line $T_\alpha O_\alpha$. If the line SO intersects the semicircle α_f at a point U and the line parallel to AB passing through U intersects α_f at a point T again, then the circle (TU) is Archimedean with respect to $(\alpha_h, \beta_h, \gamma)$.*

Proof. The point S in (i) has coordinates $(a_h, g(a,b))$, where $g(a,b) = \sqrt{a_h(a_h + 2b)}$. Hence the points T and U have coordinates $(a + r_{\mathrm{A}}^{\mathrm{h}}, g(a,b)r_{\mathrm{A}}^{\mathrm{h}}/b)$ and $(a - r_{\mathrm{A}}^{\mathrm{h}}, g(a,b)r_{\mathrm{A}}^{\mathrm{h}}/b)$, respectively. This proves (i). The point U in (ii) has x-coordinate $ab/b_h - r_{\mathrm{A}}^{\mathrm{h}}$. This proves (ii). □

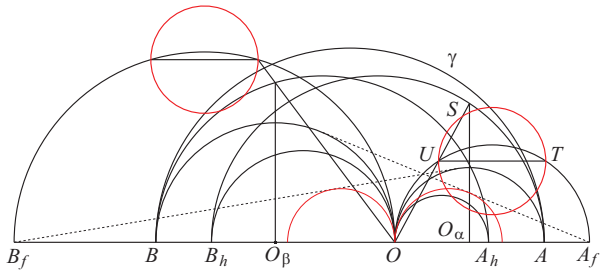


Figure 14

Notice that $(h+r_{\mathrm{A}}^{\mathrm{h}}) : (a-r_{\mathrm{A}}^{\mathrm{h}}) = b : a$ if $h = 0$.

10 Power type Archimedean circles

From now on we consider all the semicircles with centers on the line AB as circles. If two congruent circles of radius r touching at a point D also touch a given circle δ at points different from D, we say that D generates circles of radius r with δ, and the two circles are said to be generated by D with δ. If the two generated circles are Archimedean with respect to $(\alpha_h, \beta_h, \gamma)$, we say that D generates Archimedean circles of $(\alpha_h, \beta_h, \gamma)$ with δ.
Frank Power has found that the point T_α generates Archimedean circles of (α, β, γ) with the circle γ [13]. Quang Tuan Bui has found that the circles (AO_β), (BO_α) and the axis belong to the same intersecting pencil of circles and the points of intersection generate Archimedean circles of (α, β, γ) with γ [3]. We generalize the two facts. The following lemma is needed [7].

Lemma 1 *For a circle δ of radius r, a point D generates circles of radius $||DO_\delta|^2 - r^2|/(2r)$ with δ.*

The parts (i) and (ii) of the next theorem are generalizations of Power's result and Bui's result, respectively (see Figure 15).

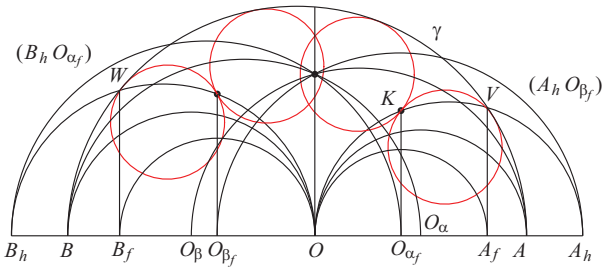


Figure 15

Theorem 12 (i) *If the circle α_h and the line $T_{\alpha_f} O_{\alpha_f}$ have a point in common, the point generates Archimedean circles of (α, β, γ) with γ.*
(ii) *The circles (AO_β), (BO_α), $(A_h O_{\beta_f})$, $(B_h O_{\alpha_f})$ and the axis belong to the same intersecting pencil of circles, and the points of intersection generate Archimedean circles of (α, β, γ) with γ.*

Proof. Let K be the point in (i) lying on α_h and $T_{\alpha_f} O_{\alpha_f}$. Then $|KO_{\alpha_f}|^2 = a_h^2 - (a_h - ab/b_h)^2$. Therefore $|KO_\gamma|^2 - (a+b)^2 = |O_\gamma O_{\alpha_f}|^2 + |KO_{\alpha_f}|^2 - (a+b)^2 = (ab/b_h - (a-b))^2 + a_h^2 - (a_h - ab/b_h)^2 - (a+b)^2 = -2ab$. Therefore K generates Archimedean circles of (α, β, γ) with γ by Lemma 1. The part (ii) follows from the fact that the powers of the point O with respect to (AO_β), $(A_h O_{\beta_f})$ and $(B_h O_{\alpha_f})$ are the same. □

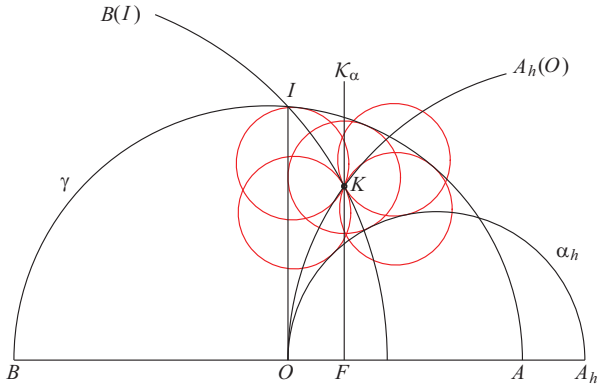


Figure 16

Recall that I is the point of intersection of the axis and the circle γ lying in the region $y > 0$. Quang Tuan Bui has also found that the points of intersection of the circles (AO) and $B(I)$ generate Archimedean circles of (α, β, γ) with the circle γ [2]. Let J be the point of intersection of the circle $B(I)$ and the line AB lying in the region $x > 0$. If $a < b$, we can choose h so that $4a_h < |OJ|$ holds. Then the circles $A_h(O)$ and $B(I)$ have no points in common. Let $\mathcal{K}_\alpha$ be the perpendicular to the line AB from the center of the circle δ_h^α. Quang Tuan Bui's result is generalized as follows.

Theorem 13 (i) *The circles $A_h(O)$, $B(I)$ and the line $\mathcal{K}_\alpha$ belong to the same pencil of circles. If the pencil is intersecting, the points of intersection generate Archimedean circles of $(\alpha_h, \beta_h, \gamma)$ with each of the circles γ and α_h.*
(ii) *The circles $A_m(O)$, $B_m(I)$ and the line $\mathcal{K}_\alpha$ belong to the same intersecting pencil of circles, and the points of intersection generate Archimedean circles of (α, β, γ) with the circle γ.*

Proof. The circles $A_h(O)$ and $B(I)$ are expressed by the equations

$$(x - 2a_h)^2 + y^2 = 4a_h^2 \tag{7}$$

and

$$(x + 2b)^2 + y^2 = 4b^2 + 4ab, \tag{8}$$

respectively (see Figure 16). Subtracting (8) from (7) and rearranging, we get $x = r_{\mathrm{A}}^{\mathrm{h}}$. Therefore $A_h(O)$, $B(I)$ and

$\mathcal{K}_\alpha$ belong to the same pencil of circles. Let us assume that the pencil is intersecting and K is one of the points of intersection. Let F be the foot of the perpendicular from K to AB. Then $|KF|^2 = 4a_h^2 - (r_A^h - 2a_h)^2$. Then $|KO_\gamma|^2 - (a+b)^2 = (r_A^h - (a-b))^2 + |KF|^2 - (a+b)^2 = -2(a+b)r_A^h$. Therefore K generates Archimedean circles of $(\alpha_h, \beta_h, \gamma)$ with γ by Lemma 1. The rest of (i) follows from $|KO_{\alpha_h}|^2 - a_h^2 = (r_A^h - a_h)^2 + |KF|^2 - a_h^2 = 2a_h r_A^h$. We prove (ii). The circles $A_m(O)$ and $B_m(I)$ are expressed by the equations

$$(x - (2a+h))^2 + y^2 = (2a+h)^2 \tag{9}$$

and

$$(x + 2b + h)^2 + y^2 = (2b+h)^2 + 4ab, \tag{10}$$

respectively. Subtracting (10) from (9), we get $x = r_A^h$. Therefore $A_m(O)$, $B_m(I)$ and $\mathcal{K}_\alpha$ belong to the same pencil. Substituting $x = r_A^h$ in (9), and using $a + h > 0$, we get $y^2 = (2a+h)^2 - (r_A^h - (2a+h))^2 = r_A^h(2(2a+h) - r_A^h) > r_A^h(2a - r_A^h) > 0$. Therefore $A_m(O)$ and $\mathcal{K}_\alpha$ intersect. The rest of (ii) can be proved similarly as the proof of (i). Let K be one of the points of intersection in (ii), and let F be the foot of the perpendicular from K to AB. Then $|KF|^2 = r_A^h(2(2a+h) - r_A^h)$. Therefore $|KO_\gamma|^2 - (a+b)^2 = (r_A^h - (a-b))^2 + |KF|^2 - (a+b)^2 = -2ab$. This proves (ii). □

References

[1] Q. T. BUI, A newly born pair of siblings to Archimedes' twins, `http://www.cut-the-knot.org/Curriculum/Geometry/ArbelosBui.shtml`.

[2] Q. T. BUI, Two more Powerian pairs in the arbelos, *Forum Geom.* **8** (2008), 149–150.

[3] Q. T. BUI, The arbelos and nine points circles, *Forum Geom.* **7** (2007), 115–120.

[4] C. W. DODGE, T. SCHOCH, P. Y. WOO, P. YIU, Those ubiquitous Archimedean circles, *Math. Mag.* **72** (1999), 202–213.

[5] H. OKUMURA, Lamoenian circles of the collinear arbelos, *KoG* **17** (2013), 9–13.

[6] H. OKUMURA, Archimedean twin circles in the arbelos, *Math. Gaz.* **97** (2013), 512.

[7] H. OKUMURA, Archimedean circles of the collinear arbelos and the skewed arbelos, *J. Geom. Graph.* **17** (2013), 31–52.

[8] H. OKUMURA, Dilations and the arbelos, *Normat* **60** (2012), 4–8.

[9] H. OKUMURA, Ubiquitous Archimedean circles of the collinear arbelos, *KoG* **16** (2012), 17–20.

[10] H. OKUMURA, M. WATANABE, Remarks on Woo's Archimedean circles, *Forum Geom.* **7** (2007), 125–128.

[11] H. OKUMURA, M. WATANABE, The twin circles of Archimedes in a skewed arbelos, *Forum Geom.* **4** (2004), 229–251.

[12] H. OKUMURA, M. WATANABE, The Archimedean circles of Schoch and Woo, *Forum Geom.* **4** (2004), 27–34.

[13] F. POWER, Some more Archimedean circles in the arbelos, *Forum Geom.* **5** (2005), 133–134.

Hiroshi Okumura
e-mail: okumura.hiroshi@yamato-u.ac.jp

Department of Mathematics, Yamato University
2-5-1 Katayama Suita Osaka 564-0082, Japan

Original scientific paper
Accepted 12. 11. 2014.

BORIS ODEHNAL

Distances and Central Projections

Distances and Central Projections

ABSTRACT

Given a point P in Euclidean space $\mathbb{R}^3$ we look for all points Q such that the length $\overline{PQ}$ of the line segments PQ from P to Q equals the length of the central image of the segment. It turns out that for any fixed point P the set of all points Q is a quartic surface Φ. The quartic Φ carries a one-parameter family of circles, has two conical nodes, and intersects the image plane π along a proper line and the three-fold ideal line p_2 of π if we perform the projective closure of the Euclidean three-space. In the following we shall describe and analyze the surface Φ.

Key words: central projection, distance, principal line, distortion, circular section, quartic surface, conical node

MSC 2010: 51N20, 14H99, 70B99

Udaljenosti i centralna projekcija

SAŽETAK

Za danu točku P u euklidskom prostoru $\mathbb{R}^3$ traže se sve točke Q takve da je duljina $\overline{PQ}$ dužine PQ jednaka duljini njezine centralne projekcije. Pokazuje se da je za čvrstu točku P skup svih točaka Q kvartika Φ. Kvartika Φ sadrži jednoparametarsku familiju kružnica, ima dvije dvostruke točke, te siječe ravninu slike π po jednom pravom pravcu i tri puta brojanom idealnom pravcu p_2 ravnine π (promatra se projektivno proširenje trodimenzionalnog euklidskog prostora). U radu se opisuje i istražuje ploha Φ.

Ključne riječi: centralna projekcija, udaljenost, glavni pravac, distorzija, kružni presjek, kvartika, dvostruka točka

1 Introduction

It is well-known that segments on lines which are parallel to the image plane π or, equivalently, orthogonal to the fibres of an *orthogonal projection* have images of the same length, *i.e.*, they appear undistorted, see [1, 4, 5, 7]. The lines orthogonal to the fibres of an orthogonal projection are usually called *principal lines* and they are the only lines with undistorted images under this kind of projection.

In case of an *oblique parallel projection*, *i.e.*, the fibres of the projection are not orthogonal (and, of course, not parallel) to the image plane, the principal lines are still parallel to the image plane π. Nevertheless, there is a further class of principal lines in the case of a parallel projection $\iota: \mathbb{R}^3 \to \mathbb{R}^2$. As illustrated in Figure 1, we can see that in between the parallel fibres f_P and f_Q of two arbitrary points P and Q on a principal line $l \parallel \pi$ we can find a second segment emanating from P and ending at $\widetilde{Q}$ with $\overline{PQ} = \overline{P\widetilde{Q}} = \overline{P'Q'}$. (Here and in the following we write P' for the image point of P instead of $\iota(P)$.) In case of an orthogonal projection, we have $Q = \widetilde{Q}$, cf. Figure 1.

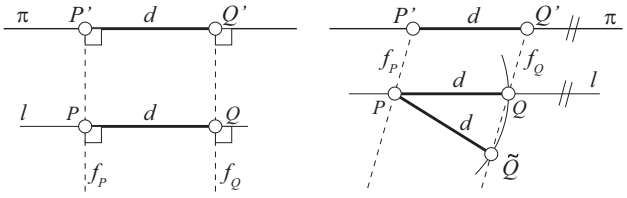


Figure 1: *Principal lines: orthogonal projection (left), oblique parallel projection (right).*

In both cases, the orthogonal projection and the oblique parallel projection, the principal lines are mapped *congruent* onto their images.

What about the central projection? Let $\kappa: \mathbb{R}^3 \setminus \{O\} \to \pi$ be the a central projection with center (eyepoint) O and image plane π. For the sake of simplicity, we shall write P' instead of $\kappa(P)$. Again the lines parallel to π serve as principal lines. Of course, the restriction $\kappa|_l$ of κ to a line $l \parallel \pi$ is a similarity mapping. The mapping $\kappa|_l$ is a congruent transformation if, and only if, $l \subset \pi$ because it is the identity in this case.

From Figure 2 we can easily guess that even in the case of central projections there are more line segments than

those in the image plane π having central images of the same length. Once we have chosen a point P on the fibre f_P through P' we can find up to two points Q, $\widetilde{Q}$ on the fibre f_Q through Q' such that $\overline{P'Q'} = \overline{PQ} = \overline{P\widetilde{Q}}$ holds as long as $\overline{Pf_Q} < \overline{P'Q'}$. The points Q and $\widetilde{Q}$ coincide exactly if $\overline{Pf_Q} = \overline{P'Q'}$. Finally, there are no points Q and $\widetilde{Q}$ if $\overline{Pf_Q} > \overline{P'Q'}$.

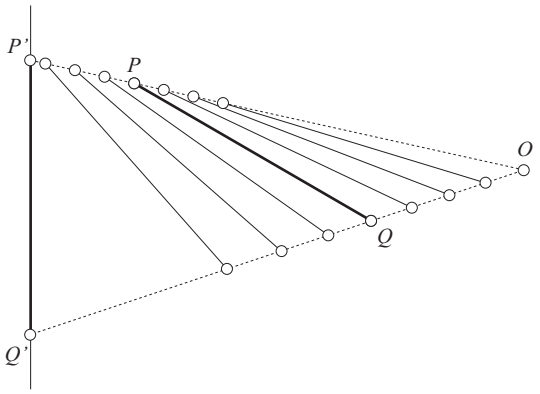


Figure 2: *Some of infinitely many segments of length s with the same image $P'Q'$ and, therefore, also of length s.*

In the case of a central projection κ, only the lines in the image plane are mapped *congruent* onto their images. All the other lines which carry segments whose images are of the same length are *not mapped congruent* onto their images. Just one segment on all these lines has a κ-image of the same length.

Note that if either Q or P equals O the line $[P,Q]$ is mapped to a point. Thus $s = \overline{PQ} \neq \overline{P'Q'}$ since the latter quantity is undefiend for either Q' or P' does not exist.

Assume further that $P \neq O$ is an arbitrary point in Euclidean three-space. Now we can ask for the set of all points Q at fixed distance, say $s \in \mathbb{R} \setminus \{0\}$, such that

$$s = \overline{PQ} = \overline{P'Q'} \tag{1}$$

where $P' := \kappa(P)$ and $Q' = \kappa(Q)$ and $s \in \mathbb{R} \setminus \{0\}$. The left-hand equation of (1) can also be skipped. Then, we are looking for all points Q being the endpoints of line segments emanating from P whose central image has the same length. It is clear that the set of all Q is an algebraic surface. In Section 2 we shall describe and analyze this surface in more detail. Section 3 is devoted to the study of algebraic properties of this surface. Surprisingly, this type of quartic surface does appear among the huge number of quartic surfaces in [3].

In the following $\mathbf{x} = (x,y,z)^{\mathrm{T}} \in \mathbb{R}^3$ are Cartesian coordinates. For any two vectors $\mathbf{u}$ and $\mathbf{v}$ from $\mathbb{R}^3$ we denote the canonical scalarproduct by

$$\langle \mathbf{u}, \mathbf{v} \rangle = u_x v_x + u_y v_y + u_z v_z.$$

Based on the canonical scalarproduct, we can compute the length $\|\mathbf{v}\|$ of a vector $\mathbf{v}$ by $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$.

2 The set of all endpoints

In the following we assume that there is the central projection $\kappa : \mathbb{R}^{3*} \to \pi \cong \mathbb{R}^2$ with the *image plane* π where $\mathbb{R}^{3*} := \mathbb{R}^3 \setminus \{O\}$ and $O \notin \pi$ shall be the center of the projection, *i.e.*, the *eyepoint*. The *principal (vanishing) point* $H \in \pi$ is π's closest point to the eyepoint O and $d := \overline{OH} = \overline{O\pi}$ is called the *distance* of κ. Therefore, H is the pedalpoint of the normal from the eyepoint O to the image plane π.

Let us assume that $P \in \mathbb{R}^{3*} \setminus \{\pi\}$ is a point in Euclidean three-space (neither coincident with O nor in π). With $P' = [O,P] \cap \pi$ we denote the κ-image of P. The set of all points $Q' \in \pi$ with a certain fixed distance $s \in \mathbb{R} \setminus \{0\}$ from P' is a circle $c_{P',s}$ in the image plane π centered at P' with radius s, see Figure 3.

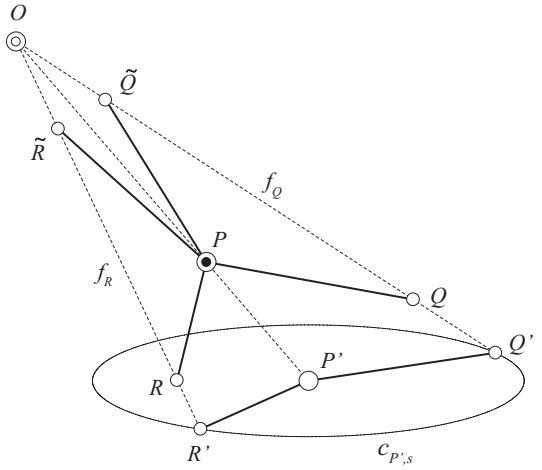


Figure 3: *Line segments in π and their equally long preimages.*

We find all possible preimages of Q' on the quadratic cone $\Gamma_{P',s} = c_{P',s} \vee O$ of κ-fibres through all points on $c_{P',s}$. The preimages shall satisfy

$$s = \overline{P'Q'} = \overline{PQ}$$

and, therefore, they are located on a Euclidean sphere $\Sigma_{P,s}$ centered at P with radius s. Consequently, we can say:

Theorem 1 *The set of all points $Q \in \mathbb{R}^3$ with $\overline{PQ} = \overline{P'Q'} = s \in \mathbb{R} \setminus \{0\}$ for some point $P \in \mathbb{R}^{3*} \setminus \{\pi\}$ is a quartic space curve q being the intersection of a sphere $\Sigma_{P,s}$ (centered at P with radius s) with a quadratic cone $\Gamma_{P',s}$ whose vertex is the eyepoint O and the circle $c_{P',s}$ (lying in π, centered at P's κ-image P', and with radius s) is a directrix.*

The quartic curve q mentioned in Theorem 1 has always two branches, since the two points on each generator f_Q of $\Gamma_{P',S}$ are the points of intersection of the generator f_Q with the sphere $\Sigma_{P,S}$. Therefore, q is in general not rational. An example of such a quartic is displayed in Figure 4 where the sphere $\Sigma_{P,S}$ and the cone $\Gamma_{P',S}$ are also shown.

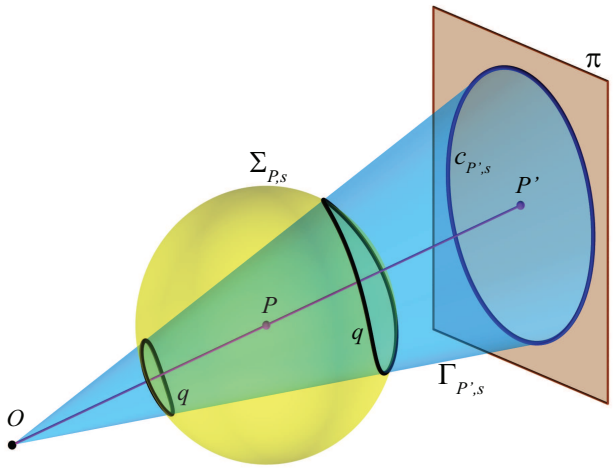


Figure 4: *The quartic curve q of possible endpoints of line segments starting at P with length s and equally long image segments. The curve q is the intersection of the quadratic cone $\Gamma_{P',S}$ and the sphere $\Sigma_{P,S}$.*

Not even in the cases $[O,P]\perp\pi$ and $P\in\pi$ an exeption occurs: q happens to be the union of two circles (rational curves). However, the union of rational curves is (in general) not rational. In the first case $\Gamma_{P',S}$ is a cone of revolution and $\Sigma_{P,S}$ is centered on the cone's axis. Consequently, q degenerates and becomes a pair of parallel circles on both surfaces. In the second case the quartic q is also the union of two circles, namley $c_{P',S}$ and a further circle on $\Sigma_{P,S}$ and $\Gamma_{P',S}$.
Figure 4 shows an example of such a quartic curve (in the non-rational or generic case) carrying the preimages of possible endpoints Q.
As the length s of PQ as well as of $P'Q'$ can vary freely, there is a linear family of quartic curves depending on s. Thus, from Theorem 1 we can deduce the following:

Theorem 2 *The set of all points Q being the endpoints of line segments PQ starting at an arbitrary point $P\in\mathbb{R}^{3*}\setminus\{\pi\}$ with $\overline{PQ}=\overline{P'Q'}$ is a quartic surface Φ.*

Proof. There exists a $(1,1)$-correspondence between the pencil of quadratic cones $\Gamma_{P',S}$ and the pencil of spheres $\Sigma_{P,S}$. Consequently, the manifold of common points, *i.e.*, the set of points common to any pair of assigned surfaces is a quartic variety, cf. [6]. □

Figure 5 shows the one-parameter family of quartic curves mentioned in Theorem 1.
Figures 5 and 6 show the quartic surface Φ mentioned in Theorem 2.

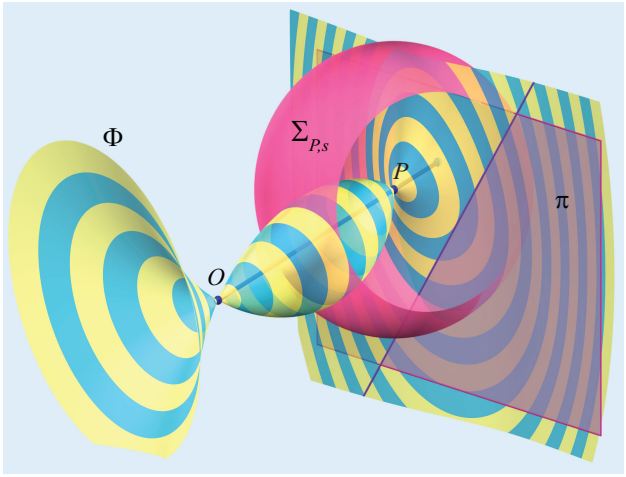


Figure 5: *The linear one-parameter family of spherical quartic curves covers a quartic surface.*

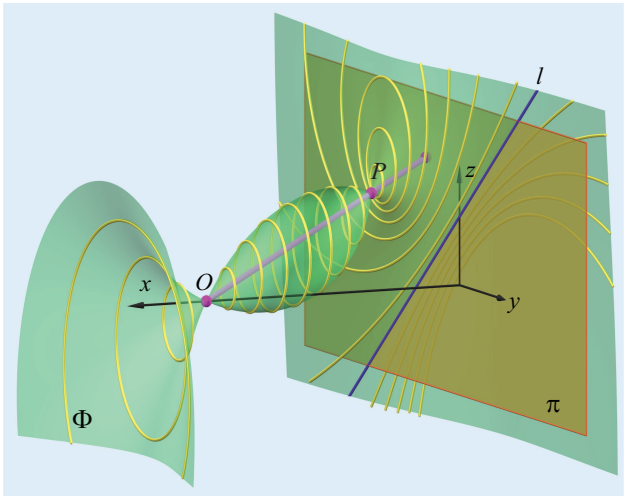


Figure 6: *The quartic surface Φ with its circles in planes parallel to π has a singularity at O and P. Φ intersects π in the line l and the ideal line p_2 of π, the latter with multiplicity three.*

3 The quartic surface

In order to describe and investigate the quartic surface Φ, we introduce a Cartesian coordinate system: It shall be centered at H, the x-axis points towards O, and π shall serve as the $[yz]$-plane. Thus, $O=(d,0,0)^{\mathrm{T}}$ and the image plane π is given by the equation $x=0$.

For any point $P \in \mathbb{R}^{3*}$ with coordinate vector $\mathbf{p} = (\xi, \eta, \zeta)^{\mathrm{T}}$ with $\xi \neq d$ the central image $P' := \kappa(P) = [O,P] \cap \pi$ is given by

$$\mathbf{p}' = \left(0, \frac{d\eta}{d-\xi}, \frac{d\zeta}{d-\xi}\right)^{\mathrm{T}}. \tag{2}$$

Obviously, $P' = P$ if $P \in \pi$, *i.e.*, $\xi = 0$. The points in the plane

$$\pi_v : \; x = d \tag{3}$$

have no image in the affine part of the plane π. Therefore, the plane π_v is called *vanishing plane*. The plane π_v contains the center O and is parallel to π at distance d. Performing the projective closure of $\mathbb{R}^3$ the images of all points of $\pi_v \setminus \{O\}$ are the ideal points of π gathering on π's ideal line p_2 .
Let now Q be the variable endpoint of a segment starting at P. The point Q shall be given by its coordinate vector $\mathbf{x} = (x,y,z)^{\mathrm{T}}$. Then, an implicit equation of Φ is given by

$$\Phi : \; \overline{PQ}^2 - \overline{P'Q'}^2 = 0. \tag{4}$$

Using Eq. (2) we can write Eq. (4) in terms of coordinates as

$$\begin{aligned} \Phi : \; & d^2((\eta(d-x) - y\delta)^2 + \\ & + (\zeta(d-x) - z\delta)^2) = \\ = \big((x-\xi)^2 + (y-\eta)^2 + (z-\zeta)^2\big) & \cdot \\ \cdot \delta^2 (d-x)^2 & \end{aligned} \tag{5}$$

where $\delta := d - \xi$.

4 Properties of Φ

A closer look at the equation of Φ as given by Eq. (5) allows us to formulate the following theorem which holds in projectively extended Euclidean space $\mathbb{R}^3$:

Theorem 3 *Let $\kappa : \mathbb{R}^{3*} \to \pi$ be a central projection from a point $O \in \mathbb{R}^3$ to a plane $\pi \not\ni O$ and let further $P \in \mathbb{R}^{3\star}$ be a point in Euclidean three-space. The set of all points Q satisfying*

$$\overline{PQ} = \overline{P'Q'}$$

(where $P' = \kappa(P)$ and $Q' = \kappa(Q)$) is a uni-circular algebraic surface Φ of degree four. The ideal line p_2 of π is a double line of Φ.

Proof. The algebraic degree Φ can be easily read off from Eq. (5).
In order to show the circularity of Φ, we perform the projective closure of $\mathbb{R}^3$ and write Φ's equation (5) in terms of homogeneous coordinates: We substitute

$$x = X_1 X_0^{-1}, \;\; y = X_2 X_0^{-1}, \;\; z = X_3 X_0^{-1}$$

and multiply by X_0^4. The intersection of the (projectively) extended surface Φ with the ideal plane $\omega : \; X_0 = 0$ is given by inserting $X_0 = 0$ into the homogeneous equation of Φ which yields the equations of a quartic cycle

$$\phi : \; X_1^2(X_1^2 + X_2^2 + X_3^2) = X_0 = 0. \tag{6}$$

The first factor of the latter equation tells us that the ideal line p_2 of the image plane $\pi : \; X_1 = 0$ is a part of $\phi = \omega \cap \Phi$ and has multiplicity two. In order to be sure that p_2 is a double line on Φ, we compute the Hessian $\mathrm{H}(\Phi)$ of the homogeneous equation of Φ and evaluate at

$$p_2 = (0 : 0 : X_2 : X_3)$$

(with $X_2 : X_3 \neq 0 : 0$ or equivalently $X_2^2 + X_3^2 \neq 0$). This yields

$$\mathrm{H}(\Phi) = 2\delta^2(X_2^2 + X_3^2) \begin{pmatrix} 0 & -d & 0 & 0 \\ -d & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{7}$$

which shows that all but two partial derivatives of Φ's homogeneous equation do not vanish along p_2. Therefore, p_2 is a double line on Φ.

The second factor of the left-hand side of (6) defines the equation of the *absolute conic* of Euclidean geometry with multiplicity one. Thus, Φ is uni-circular. □

A part of the double line p_2 is shown in Figure 7 which shows a perspective image of the surface Φ and the circles and lines on Φ.

Corollary 1 *In the case $P \in \pi$,* i.e., *$\xi = 0$, the surface Φ is the union of the image plane π (a surface of degree one) and a cubic surface.*

Proof. If $P \in \pi$, we have $\xi = 0$. Inserting $\xi = 0$ into Eq. (5) we find

$$x(\|\mathbf{x}\|^2(x-2d) - 2(x-d)(\eta y + \zeta z) + d^2 x) = 0.$$

Obviously, Φ is the union of the plane π (with the equation $x = 0$) and a cubic surface. □

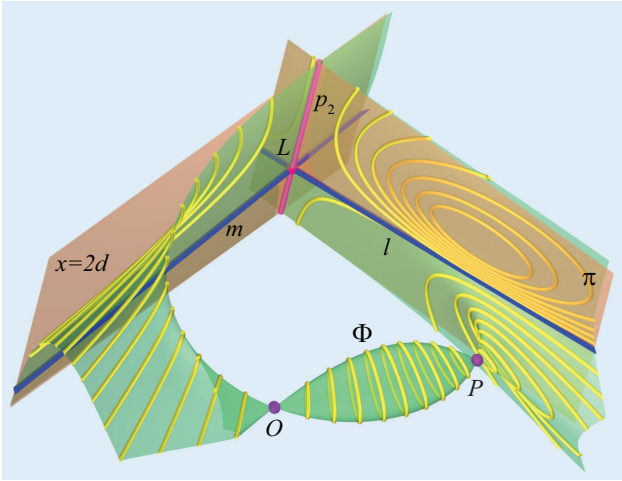


Figure 7: *A perspective image of the situation in space: The ideal line p_2 of the image plane π of κ is a part of the double curve of Φ. The two parallel lines l and m meet in the common ideal point $L \in p_2$. The two planes π and $x = 2d$ serve as tangent planes of Φ along p_2 and meet Φ along p_2 with multiplicity three and l and m appear as the remaining linear part.*

The spheres of the one-parameter family of concentric spheres centered at P carrying the one-parameter family of quartic curves $q \subset \Phi$ intersect Φ along the quartics q and the absolute circle of Euclidean geometry. At the latter the spheres are in concact with each other and with the quartic surface Φ. This can easily be shown by computing the resultants of Φ's and the spheres' homogeneous equations with respect to X_0. From this resultant the factor $X_1^2+X_2^2+X_3^2$ splits off with multiplicity 2. In other words: Φ and all spheres about P share an isotropic tangent cone with vertex at P.
The shape of the curve $\omega \cap \Phi$ together with $\eta^2+\zeta^2 \neq 0$, *i.e.*, $P \notin [O,H]$, tells us:

Theorem 4 *A plane $x = k$ ($k \in \mathbb{R}$) parallel to the image plane π intersects Φ along*

1. *the union of a circle whose center lies on a rational planar cubic curve γ and the two-fold ideal line p_2 if $k \neq 0, d, 2d, \xi$,*
2. *the union of a line l and the three-fold line p_2 if $k = 0$,*
3. *the union of a line $m \parallel l$ and the three-fold line p_2 if $k = 2d$, and*
4. *the union of a pair of isotropic lines and the two-fold line p_2 if $k = d, \xi$.*

Proof. Each planar section of the affine part of Φ is an algebraic curve whose degree is at most 4. As we have seen in the proof of Theorem 3, the ideal line p_2 of the image plane π is a two-fold line in Φ. Thus, the intersection of (the projectively extended) surface Φ with any plane parallel to π also contains this repeated line. The remaining part r of these planar intersetions is at most of degree 2.
The planes parallel to π meet the absolute conic of Euclidean geometry at their *absolute points* which induce Euclidean geometry in these planes. Since the absolute conic is known to be a part of ϕ, the curves r are Euclidean circles (including pairs of isotropic lines and the join p_2 of the two absolute points as limiting cases). The equations of the intersections of Φ with planes parallel to π can be found by rearranging Φ's equation (5) considering y and z as variables in these planes. The coefficients are univariate functions in x and we find

$$\begin{aligned} &x(x-2d)\delta^2(\underline{y^2}+\underline{z^2})+\\ &+2\delta(d-x)(\delta x+d\xi)(\eta\underline{y}+\zeta\underline{z})+\\ &+(d-x)^2\delta^2(\langle \mathbf{p},\mathbf{p}\rangle + x(x-2\xi))\\ &-d^2(\eta^2+\zeta^2)=0. \end{aligned} \tag{8}$$

The essential monomials y^2, z^2, y, and z are underlined in order to emphasize them. Note that the monomial yz does not show up. Since $\text{coeff}(x^2) = \text{coeff}(y^2)$ the curves in Eq. (8) are Euclidean circles.

1. We only have to show that the centers of the circles given in Eq. (8) on Φ in planes $x = k$ (with $k \neq 0, d, 2d, \xi$) are located on a rational planar cubic curve. For that purpose we consider Φ's inhomogeneous equation (5) as an equation of conics in the $[y,z]$ plane. By completing the squares in Eq. (8), we find the center of these conics. Keeping in mind that x varies freely in $\mathbb{R} \setminus \{0, d, 2d, \xi\}$ we can parametrize the centers by

$$\gamma(x) = \begin{pmatrix} x \\ \dfrac{\eta(d-x)(d\xi+dx-x\xi)}{\delta x(2d-x)} \\ \dfrac{\zeta(d-x)(d\xi+dx-x\xi)}{\delta x(2d-x)} \end{pmatrix} \tag{9}$$

which is the parametrization of a rational cubic curve. The cubic passes through O and P which can be verified by inserting either $x = d$ or $x = \xi$. In order to show that m is planar, we show that any four points on γ are coplanar. We insert $t_i \neq 0, d, 2d, \xi$ with $i \in \{1,2,3,4\}$ into (9) and show that the inhomogeneous coordinate vectors of the four points $\gamma(t_i)$ are linearly dependent for any choice of mutually distinct t_i.

From

$$\det\begin{pmatrix} 1 & \gamma(t_1)^{\mathrm{T}} \\ 1 & \gamma(t_2)^{\mathrm{T}} \\ 1 & \gamma(t_3)^{\mathrm{T}} \\ 1 & x\,y\,z \end{pmatrix} = 0$$

we obtain the equation

$$\eta y - \eta z = 0$$

of the plane that carries γ.

Figure 8 shows the cubic curve γ with its three asymptotes.

2. The image plane $\pi: x = 0$ of the underlying central projection κ touches (the projective extended surface) Φ along the ideal line p_2 of π. This can be concluded from the following: We write down the quadratic form

$$\mathbf{X}^{\mathrm{T}}\mathrm{H}(\Phi)\mathbf{X} = X_1(X_1 - 2dX_0) = 0$$

with $\mathrm{H}(\Phi)$ being the Hessian from (7) and $\mathbf{X} = (X_0, X_1, X_2, X_3)^{\mathrm{T}}$ being homogeneous coordinates. (Non-vanishing factors are cancelled out.) This form gives the equations of the two planes through p_2 that intersect Φ along p_2 with higher multiplicity than two, *i.e.*, in this case with multiplicity three. Thus, the multiplicity of the line p_2 considered as the intersection of π and Φ is of multiplicity three and a single line l of multiplicity one remains. This line is given by

$$l: (2d - \xi)\langle \mathbf{p}, \mathbf{p}\rangle - d^2\xi = 2\delta(\eta y + \zeta z)$$

where y and z are used as Cartesian coordinates in the image plane π.

3. In a similar manner we find the line m which is the only proper intersection of Φ with the plane $x = 2d$:

$$m: d(2d^2 - 5d\xi + 4\xi^2) - \xi\langle \mathbf{p}, \mathbf{p}\rangle = \\ = 2\delta(\eta y + \zeta z)$$

The plane of the cubic curve γ is orthogonal to the lines l and m.

4. In case of $x = \xi$, the plane runs through P. Again, the ideal line p_2 splits off with multiplicity two. The remaining part r is the pair of isotropic lines through P with the equation

$$x = \xi, \quad (y - \eta)^2 + (z - \zeta)^2 = 0.$$

The same situation occurs at O, *i.e.*, $x = d$ where the isotropic lines have the equation

$$x = d, \quad y^2 + z^2 = 0. \qquad \square$$

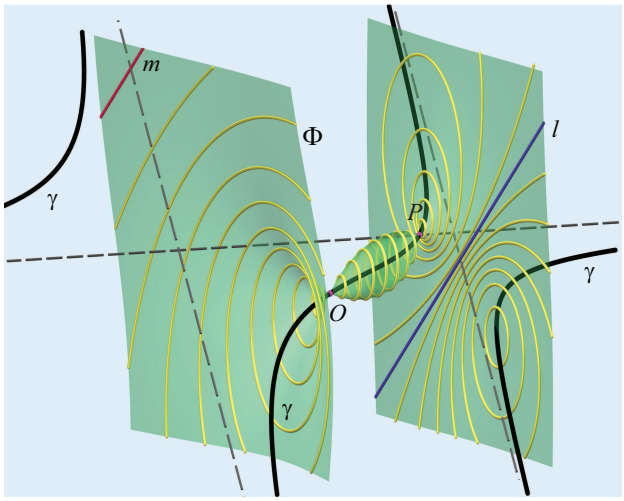


Figure 8: *The cubic curve γ carries the centers of all circles on Φ. Its ideal doublepoint $(0:0:\eta:\zeta)$ is the ideal point of the lines orthogonal to $l \parallel m$. The tangent of c at the third ideal point $(0:1:0:0)$ passes through P. The three dashed lines are γ's asymptotes.*

The circles as well as the line l on the quartic surface Φ can be seen in Figures 6, 9 and 8. In Figure 8, a small piece of the line m shows up.

Remark 1 *In the case of $P \in [O,H]$, or equivalently, $\eta^2 + \zeta^2 = 0$ the lines l and m coincide with the ideal line of π and, thus, $\pi \cap \Phi$ is the ideal line of π with multiplicity four. The same holds true for the plane $x = 2d$ if $P \in [O,H]$.*

Remark 2 *The planes π and $x = 2d$ behave like the* tangents *of a planar algebraic curve c at an ordinary double point D because these tangents intersect c at D with multiplicity three. This cannot just be seen from Figure 7.*

The lines l and m from the proof of Theorem 4 are parallel to each other but skew and orthogonal to the line $[O,P]$ as long as $\xi(\xi - 2d) \neq 0$. If $\xi = 0$ or $\xi = 2d$, we have the case mentioned in Remark 1 and l and m are ideal lines. They are still skew to $[O,P]$ but orthogonality is not defined in that case.

The set of singular surface points on Φ contains only points of multiplicity two. A more detailed description of the set of singular surface points is given by:

Theorem 5 *The set of singular surface points on Φ is the union of eyepoint O, the object point P, and the ideal line p_2 of the image plane π. The eyepoint O and the object point P are conical nodes on Φ.*

Proof. The ideal line of π is a line with multiplicity two on Φ. The planes $\pi:\ x=0$ and $x=2d$ intersect Φ along this ideal line with multiplicity three as shown in the proof of Theorem 4. Therefore, the points on π's ideal line are singular points considered as points on Φ.
The points O and P are singular surface points on Φ since the gradients of Φ vanish at both points:

$$\operatorname{grad}(\Phi)(\mathrm{d},0,0)=(0,0,0)^{\mathrm{T}}$$

and

$$\operatorname{grad}(\Phi)(\xi,\eta,\zeta)=(0,0,0)^{\mathrm{T}}$$

Now we apply the translation $\tau_1:\ O\mapsto(0,0,0)^{\mathrm{T}}$ to Φ, *i.e.*, the singular point O moves to the origin of the new coordinate system. The equation of Φ does not alter its degree. However, the monomials in the equation of Φ are at least of degree two in the variables x, y, z. If we remove the monomials of degree three and four, we obtain the equation of a quadratic cone Γ_O centered at O. Its equation (in the new coordinate system, but still labelled x, y, z) reads

$$\Gamma_O:\ d^2\delta^2\langle\mathbf{x},\mathbf{x}\rangle+2d^2\delta x(\eta y+\zeta z)=$$
$$=(\delta^4+\xi(2d+\xi)\langle\mathbf{p},\mathbf{p}\rangle+\xi^3(d+\delta))x^2.$$

Γ_O is the second order approximation of Φ at O. Since Γ_O is a quadratic cone the singular point O is a conical node, see [2].
In order to show that P is also a conical node of Φ we apply the translation $\tau_2:\ P\mapsto(0,0,0)^{\mathrm{T}}$. Again we use x, y, z as the new coordinates and the quadratic term of the transformed equation of Φ given by

$$\Gamma_P:\ \xi(\delta+d)\delta^2\langle\mathbf{x},\mathbf{x}\rangle+2d^2\delta x(\eta y+\zeta z)+$$
$$+d^2(\langle\mathbf{p},\mathbf{p}\rangle-\delta^2-2\xi^2)x^2=0.$$

is the equation of a quadratic cone Γ_P centered at P. Consequently, P is also a conical node (cf. [2]). □

Remark 3 *The homogeneous equations of the quadratic cones Γ_O and Γ_P are the quadratic forms whose coefficient matrices are (non-zero) scalar multiples of the Hessian matrix of Φ's homogeneous equation evaluated at O and P.*

Figure 9 illustrates the two quadratic cones Γ_O and Γ_P. The planes parallel to π (except $x=k$ with $k\in\{d,\xi\}$) intersect both quadratic cones Γ_O and Γ_P along circles.

If $P=P'$ but $[0,P]\not\perp\pi$, *i.e.*, $P\in\pi$ and $P\neq H$, then Φ is the union of the image plane π and a cubic surface $\overline{\Phi}$ with the equation

$$(x-2d)\langle\mathbf{x},\mathbf{x}\rangle=2(x-d)(\eta y+\zeta z)-d^2x. \tag{10}$$

The cubic surface $\overline{\Phi}$ has only one singularity at O which is a conical node.

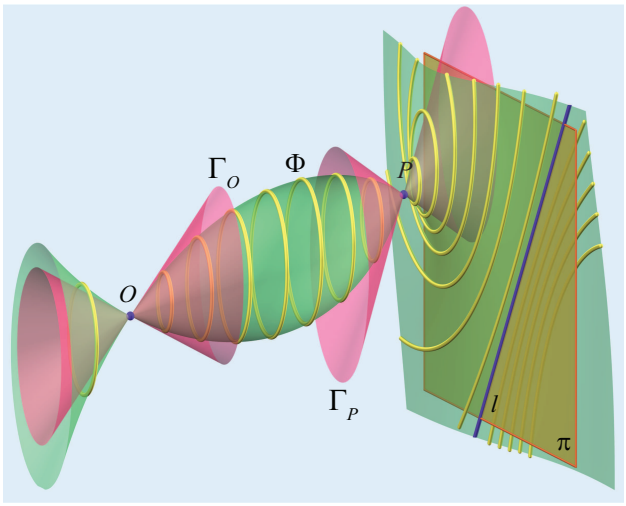


Figure 9: *The two singular points O and P are conical nodes,* i.e., *the terms of degree two of Φ's equation when translated to O or P are the equations of quadratic cones.The circular sections of Φ lie in planes that meet the quadratic cones Γ_O and Γ_P along circles.*

If $P\in[O,H]$ (but $P\neq O,H$), then $\overline{\Phi}$ is a surface of revolution with the equation

$$x(x-2d)\langle\mathbf{x},\mathbf{x}\rangle+\xi(\xi-2x)(x-d)^2-d^2x^2=0 \tag{11}$$

where $\eta^2+\zeta^2\neq0$ in contrast to earlier assumptions.
The set of singular surface points on Φ contains only points of multiplicity two. A more detailed description of the set of singular surfaces points is given by:

Theorem 6 *The set of singular surface points on Φ is the union of eyepoint O, the object point P, and the ideal line p_2 of the image plane π. The eyepoint O and the object point P are conical nodes on Φ.*

Proof. The ideal line of π is a line with multiplicity two on Φ. The planes $\pi:\ x=0$ and $x=2d$ intersect Φ along this ideal line with multiplicity three as shown in the proof of Theorem 4. Therefore, the points on π's ideal line are singular points considered as points on Φ.
The points O and P are singular surface points on Φ since the gradients of Φ vanish at both points:

$$\operatorname{grad}(\Phi)(\mathrm{d},0,0)=(0,0,0)^{\mathrm{T}}$$

and

$$\operatorname{grad}(\Phi)(\xi,\eta,\zeta)=(0,0,0)^{\mathrm{T}}$$

Now we apply the translation $\tau_1:\ O\mapsto(0,0,0)^{\mathrm{T}}$ to Φ, *i.e.*, the singular point O moves to the origin of the new coordinate system. The equation of Φ does not alter its degree.

However, the monomials in the equation of Φ are at least of degree two in the variables x, y, z. If we remove the monomials of degree three and four, we obtain the equation of a quadratic cone Γ_O centered at O. Its equation (in the new coordinate system, but still labelled x, y, z) reads

$$\Gamma_O: \; d^2\delta^2\langle \mathbf{x},\mathbf{x}\rangle + 2d^2\delta x(\eta y+\zeta z) = \\ = (\delta^4+\xi(2d+\xi)\langle \mathbf{p},\mathbf{p}\rangle + \xi^3(d+\delta))x^2.$$

Γ_O is the second order approximation of Φ at O. Since Γ_O is a quadratic cone the singular point O is a conical node, see [2].

In order to show that P is also a conical node of Φ we apply the translation $\tau_2: \; P \mapsto (0,0,0)^{\mathrm{T}}$. Again we use x, y, z as the new coordinates and the quadratic term of the transformed equation of Φ given by

$$\Gamma_P: \; \xi(\delta+d)\delta^2\langle \mathbf{x},\mathbf{x}\rangle + 2d^2\delta x(\eta y+\zeta z) + \\ + d^2(\langle \mathbf{p},\mathbf{p}\rangle - \delta^2 - 2\xi^2)x^2 = 0.$$

is the equation of a quadratic cone Γ_P centered at P. Consequently, P is also a conical node (cf. [2]). □

Figures 10 and 11 show the two distinct cases where Φ is a surface of revolution.

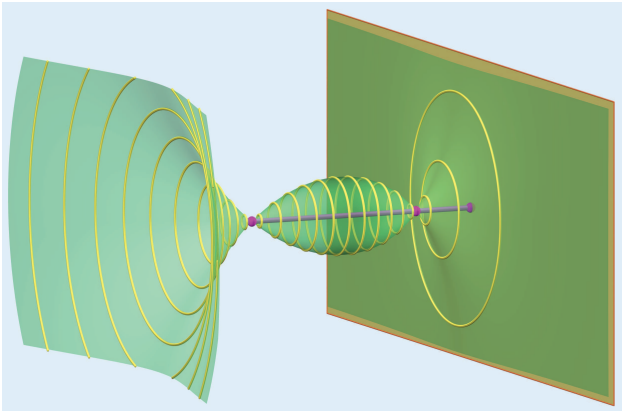


Figure 10: *The set Φ of all points Q is a quartic surface of revolution if $P \in [O,H]$ and $P \neq O,H$.*

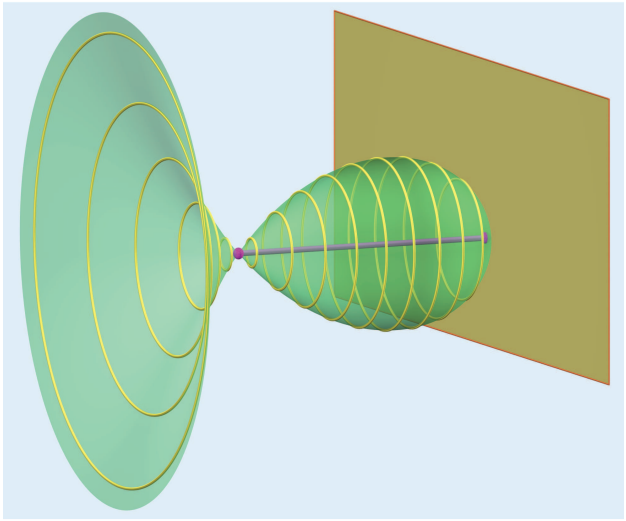


Figure 11: *Φ is the union of π and a cubic surface of revolution touching π at H if $P = H$.*

References

[1] H. BRAUNER, *Lehrbuch der konstruktiven Geometrie*, Springer-Verlag, Wien, 1986.

[2] W. BURAU, *Algebraische Kurven und Flächen*, De Gruyter, 1962.

[3] K. FLADT, A. BAUR, *Analytische Geometrie spezieller Flächen und Raumkurven*, Vieweg, Braunschweig, 1975.

[4] F. HOHENBERG, *Konstruktive Geometrie in der Technik*, 3rd Edition, Springer-Verlag, Wien, 1966.

[5] E. MÜLLER, *Lehrbuch der Darstellenden Geometrie*, Vol. 1, B. G. Teubner, Leipzig-Berlin, 1918.

[6] B. L. VAN DER WAERDEN, *Einführung in die Algebraische Geometrie*, Springer-Verlag, Berlin, 1939.

[7] W. WUNDERLICH, *Darstellende Geometrie*, 2 Volumes, BI Wissenschaftsverlag, Zürich, 1966 & 1967.

Boris Odehnal
e-mail: boris.odehnal@uni-ak.ac.at
University of Applied Arts Vienna
Oskar-Kokoschka-Platz 2, A-1100 Vienna, Austria

Original scientific paper
Accepted 29. 12. 2014.

GUNTER WEISS
SYBILLE MICK

Non-standard Visualizations of Fibonacci Numbers and the Golden Mean

Dedicated to Prof. Dr. Otto Röschel on the occasion of his 60^{th} birthday.

Non-standard Visualizations of Fibonacci Numbers and the Golden Mean

ABSTRACT

Fibonacci numbers and the Golden Mean are numbers and thus 0-dimensional objects. Usually, they are visualized in the Euclidean plane using squares and rectangles in a spiral arrangement. The Golden Mean, as a ratio, is an affine geometric concept and therefore Euclidean visualizations are not mandatory. There are attempts to visualize the Fibonacci number sequence and Golden Spirals in higher dimensions [11], in Minkowski planes [12], [4] and in hyperbolic planes (again [4]). The latter has to replace the not existing squares by sequences of touching circles. This article aims at visualizations in all Cayley-Klein planes and makes use of three different visualization ideas: nested sets of squares, sets of touching circles and sets of triangles that are related to Euclidean right angled triangles.

Key words: Cayley-Klein geometries, Fibonacci numbers, Golden Mean

MSC 2010: 51M04, 51M10, 51F20, 11B39

Nestandardne vizualizacije Fibonaccijevih brojeva i zlatni rez

SAŽETAK

Fibonaccijevi brojevi i zlatni rez su brojevi, stoga su to 0-dimenzionalni objekti. Najčešće se vizulaiziraju u euklidskoj ravnini, pomoću kvadrata i pravokutnika u spiralnom poretku. Zlatni rez, kao omjer, je pojam afine geometrije pa euklidske vizualizacije nisu nužne. Postoje pokušaji vizualizacije Fibonaccijevog niza i zlatne spirale u višim dimenzijama [11], u ravninama Minkowskog [12], [4], i u hiperboličkim ravninama, također [4], gdje se nepostojeći kvadrati zamjenjuju kružnicama koje se dodiruju. Cilj ovog rada je vizualizacija u svim Cayley-Kleinovim ravninama uz korištenje triju različitih ideja: grupiranih skupova kvadrata, skupova kružnica koje se dodiruju i skupova trokuta koji su analogni euklidskim pravokutnim trokutima.

Ključne riječi: Cayley-Kleinove geometrije, Fibonaccijevi brojevi, zlatni rez

1 Euclidean Visualizations

In this paper we continue a study of visualizing the classical sequence of Fibonacci numbers and Golden Spirals [4] and aim at visualizations in general Cayley-Klein planes.

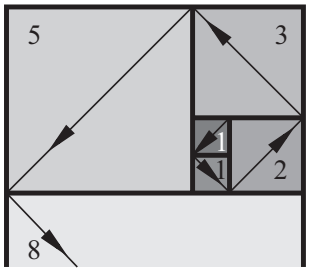


Figure 1: *Set of Fibonacci squares*

In the Euclidean plane there are mainly three cases:

(α) The standard visualization by nested sets of Golden Rectangles and gnomon squares, see Figures 1, 2 and e.g. [1], [4], [14]. Inscribing quarter circles into the gnomon squares results in discrete spirals of C^1-continuity.

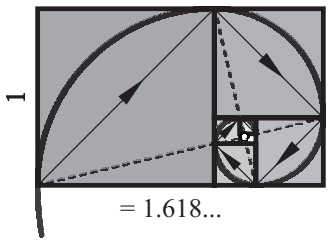


Figure 2: *Nested set of Golden rectangles and quater circle biarc spiral*

(β) In [4] the authors propose to use a chain of circles where each circle touches the former two circles, see Figures 3, 4. This type of visualization even allows generally normed planes (Minkowski planes) and also a hyperbolic plane as places of action, i.e. planes without a (proper) concept of squares.

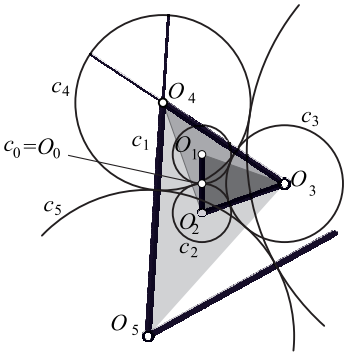


Figure 3: *Set of touching Fibonacci circles and Fibonacci spiral polygon of circle centres*

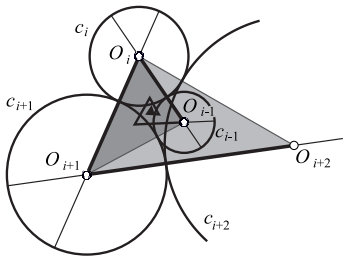


Figure 4: *Set of "Golden Circles" and spiral polygon of circle centres*

From the construction of touching circles in Figure 3 follows that we get a limit triangle of circle centres with side ratio

$$a:b:c = F_{i+1} : (F_{i-1}+F_{i+1}) : F_{i+2} = 1 : (1+\frac{1}{\phi^2}) : \phi$$
$$= (1+\phi) : (2+\phi) : (1+2\phi). \qquad (1)$$

Figure 4 contains a nested set of such triangles with side ratio (1).

(γ) In this paper we propose an additional way of visualizing Fibonacci numbers and Golden spirals using sets of Pythagoras triplets and right angled triangles akin to the classical root spiral. Similar to Fibonaccis rule Pythagoras formula adds two numbers and gives a new one. This in mind, one can generate the set of natural numbers as well as the Fibonacci sequence via iterative processes applied to the classical formula of Pythagoras, see Figure 5.

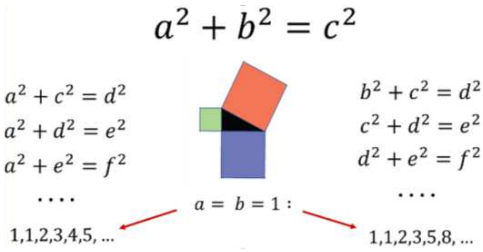


Figure 5: *Natural numbers and Fibonacci numbers derived from Pythagoras' formula*

While the left column leads to the so-called root-spiral and triangles with cathetes $\sqrt{n}, 1$ and the hypotenuse $\sqrt{n+1}$, see Figure 6, the right column leads to cathetes $\sqrt{F_{i-1}}$, $\sqrt{F_i}$ and the hypotenuse $\sqrt{F_{i+1}}$, see Figure 7. Again, we get a 'limit' triangle with side ratio

$$a:b:c = \sqrt{F_{i-1}} : \sqrt{F_i} : \sqrt{F_{i+1}} \qquad (2)$$

Such a triangle might as well be called Golden (right-angled) Triangle. Figure 8 shows the spiral polygon derived from such Golden Triangles.

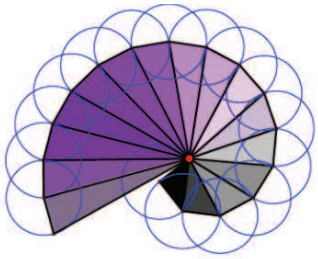


Figure 6: *The classical "root spiral"*

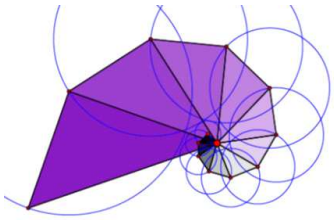


Figure 7: *Fibonacci number root spiral*

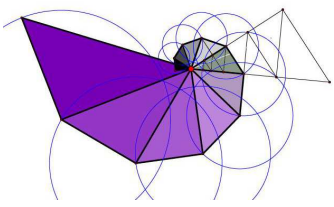


Figure 8: *Golden Root Spiral Polygon*

While in the Euclidean plane triangles with side ratio (2) are right angled by themselves, this is not the case in the hyperbolic plane. But as long the triangle in-equation for the side ratio (2) remains valid this third type of visualizations is also possible in a wide range of settings as it is shown in the following chapters. It should be mentioned that the ratio (2) is supposed to be connected with some of the ancient Egyptian pyramids, c.f. the concepts "Kepler triangle" and "Golden Pyramid", see [14].

2 The Golden Mean and the Fibonacci-sequence

There is such a huge number of publications dealing with the Golden Mean and Fibonacci numbers and the topic has become common knowledge among mathematicians that one can refrain from citing more than a few basic books on that theme, e.g. [1], [9] and the Wikipedia article [14] which contains a long list of references.

Fibonacci numbers and the Golden Mean value ϕ are numbers, thereby ϕ is a root of $x^2 - x - 1 = 0$ and the result of the "most irrational continued fraction" $\phi = 1 + \frac{1}{1+\frac{1}{\ddots}}$. Being *numbers*, these objects are 0-dimensional. As a "ratio of 3 collinear points" ϕ has a 1-dimensional visualization and it is an affine geometric concept independent from any Euclidean structure. Obviously, 2-dimensional visualizations in the Euclidean plane using squares or circles cannot be mandatory! Visualizations in other settings are at least possible and might even enjoy some aesthetic value. Such non-standard visualizations also give some deeper insight into the interplay of visualization assumptions and the structure of the places of action. This might justify the following considerations.

3 Other places of action

At first we collect those visualizations in places of actions already treated in references and show some figures:

3.1 Higher dimensional Euclidean visualizations

See [11] and Figure 9.

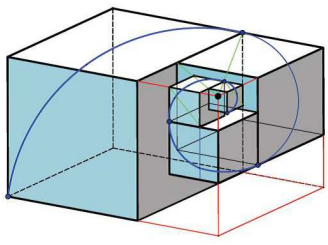


Figure 9: *Nested set of Golden Prisms with cubes as gnomon figures and a Golden helix biarc spiral*

3.2 Visualizations in (affine) normed planes (so-called Minkowski planes)

See [1], [12] and Figures 10, 11.

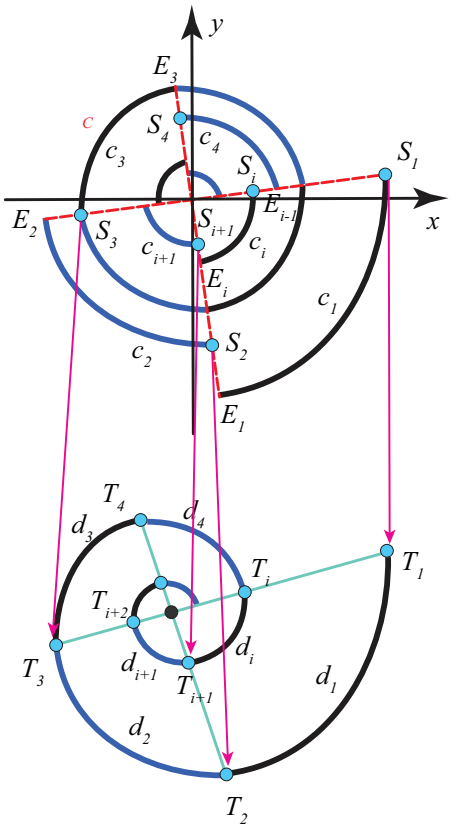


Figure 10: *Minkowski geometric analogue to Figure 2. (Applying translations τ_i to the partial arcs c_i of concentric Minkowski circles under the end-point is start-point condition results in a Minkowski circular bi-arc spiral)*

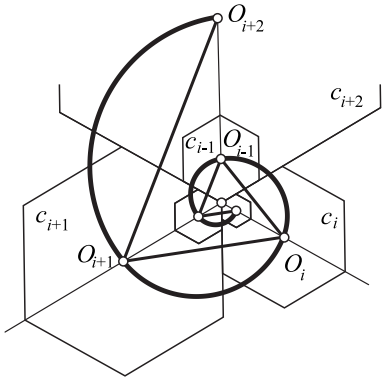


Figure 11: *Minkowski geometric analogue to Figure 4. (Sequence of touching Golden Minkowski-circles and Golden Triangles in a Minkowski plane with hexagons as circles.*

3.3 Visualizations in a hyperbolic plane

See [4] and Figure 12.

As there are no similarities and no squares in a hyperbolic plane one cannot use the visualization method (α). In [4] the authors propose method (β) and handle the different radii of the Golden circles via a hyperbolic scaled line. For visualizing the circle chain they use the F. Klein model of a hyperbolic plane and base the construction on an arbitrarily given scale on a hyperbolic line, see Figure 12.

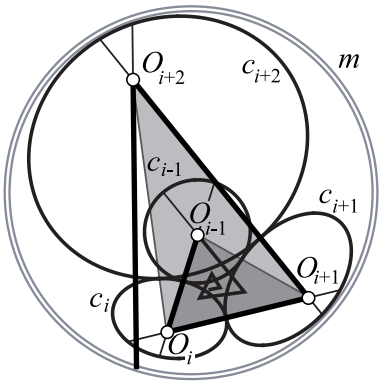


Figure 12: *A sequence of tangent Golden circles and Golden triangles in the hyperbolic plane.*

The hyperbolic case encourages us to look for visualizations also in other Cayley-Klein planes. Of course, one faces the problem of finite length of lines in e.g. the elliptic plane, while a line in affine planes and the hyperbolic plane has infinite length. It turns out that also some affine Cayley-Klein planes need greater modifications of the three visualisation schemes, as shown in the next chapters. For an overview of all Cayley-Klein planes see e.g. [3].

4 Visualisations in Cayley-Klein planes

4.1 Affine cases

4.1.1 The Euclidean case

See Chapter 1.

4.1.2 The pseudo-Euclidean case (*pe*-case)

Here, the visualization method (α) by squares does work. The constructions are based on affine parallelograms, which, by a suitably chosen affine coordinate frame can be called affine squares, affine Golden rectangles, see [4] and Figure 13.

Note that, from the pseudo-Euclidean point of view, one side of Golden Rectangles is space-like, while the other is time-like (with a negative length). Therefore one needs to modify the side ratio concept of *pe*-Golden Rectangles by using absolute values as

$$|a| : |b| = 1 : \phi. \tag{3}$$

Furthermore, the biarc spiral curve in Figure 13 consists of general conic section arcs and not of *pe*-quarter circles.

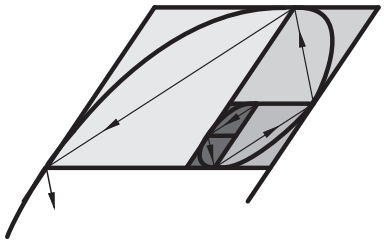


Figure 13: *"Affine Golden biarc spiral", "affine Golden Rectangles" and "squares" in an affine plane, which is endowed with a suitable affine coordinate frame.*

Method (β) does not work: It is not possible to construct a real space-like *pe*-circle (i.e. a Euclidean equilateral hyperbola with predifined directions of asymptotes), which touches two mutually touching space-like *pe*-circles. Their centres would have to form triangles with side length ratio (2). For $a < b < c$ we would have $a + b < c$, expressing that in the *pe*-plane the triangle inequality would not held.

This is why method (γ) is problematic, too. It would have to be modified according to the norm-function of the *pe*−plane by absolute values similar to (3). But as there exist the (continuous) group of *pe*-rotations, the group of translations and the group of dilatations, one can at least construct *pe*−spirals as orbits of a one-parameter group of *pe*−similarities. Generating such a discrete spiral polygon with the *pe*-similarity factor ϕ (or $\sqrt{\phi}$) then leads to a visualization of a Golden *pe*-root spiral, see Figure 14.

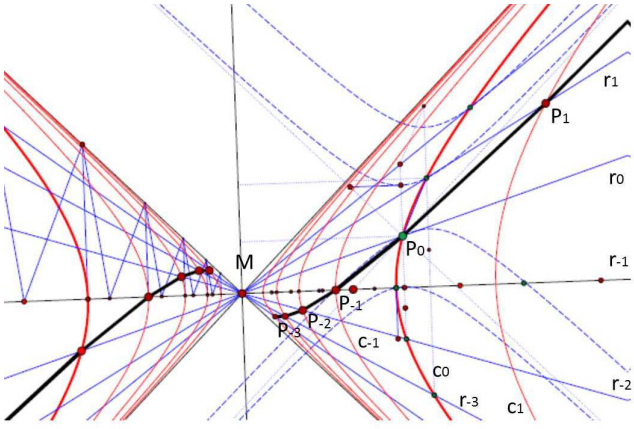


Figure 14: *Golden pseudo-Euclidean spiral polygon*

4.1.3 The isotropic case (*i*-case)

In the projective extension of an isotropic plane the ideal line u and ideal point U act as absolute figure. Choosing an affine coordinate frame $(O;E,F)$ with O,F,U collinear and E,F as unit points of the axes $x = OE, y = OU$ allows to measure the *i*-distance of two points $P(x_P,y_P)$ and $Q(x_Q,y_Q)$ as $d(P,Q) := |x_Q - x_P|$. If P,Q are collinear with U, one uses $|y_Q - y_P|$ as a substitute for their vanishing *i*-distance and calls it "spacing (Sperrung)". Obviously it is possible to construct a nested set of "Golden Rectangles" in a seemingly spiral arrangement, see Figure 15, but this affine spiral arrangement does not suit to an isotropic spiral (c.f. [7]). This way, visualization method (α) works well.

β The set of points with fixed distance to a given point is a pair of *y*-parallel lines, so-called isotropic lines. This circle concept is not useful for our purpose. Let us consider the following circle concept: An *i*-circle is a conic section touching the absolute line u in the absolute point U. In our model of the *i*-plane *i*-circles are parabolas with *y*-parallel diameters. The Euclidean parameter is a proper replacement of the concept "radius", because two *i*-circles are either similar or one is the translated of the other. It is not possible to construct an *i*-circle that touches two mutually touching *i*-circles in proper points. At least one of the tangent points must be the ideal point U. Therefore visualization method (β)is not viable.

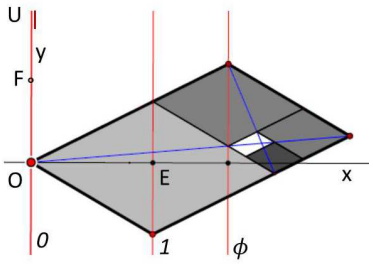


Figure 15: *Nested set of Golden Rectangles and squares in an isotropic plane.*

(γ) Similar to the *pe*-case one can generate an *i*-spiral polygon (see [7]) and can use it for the construction of a Golden *i*-root spiral, see Figure 16. For such a spiral polygon even C^1-smooth *i*-circular bi-arc spirals are possible, see Figure 17.

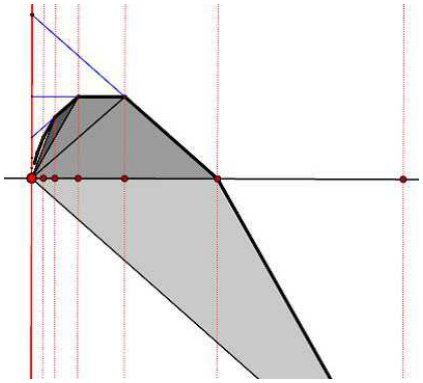


Figure 16: *Golden isotropic spiral polygon*

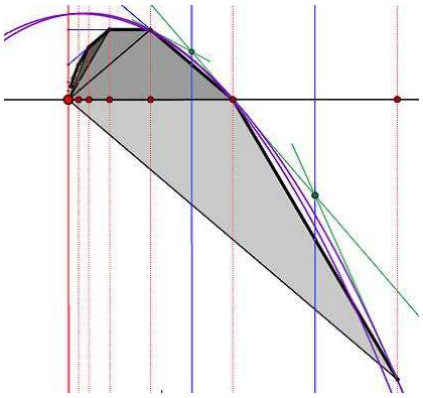


Figure 17: *i-circular biarc spiral to an i-spiral polygon*

4.1.4 The dual *pe*-plane

The dual *pe*-plane is also called quasi-hyperbolic plane (*qh*-plane). Its absolute is a pair of real lines e, f, whereby one (say $f =: u$) can act as "line at infinity" u of a projective embedding of the affine plane. Therefore, we can discuss this case in this sub-chapter. In this model of a *qh*-plane the absolute involution with the fixed lines e, f simply becomes the (Euclidean) reflection at the proper line e. Strictly speaking, the *qh*-plane is just one half-plane with respect to e, but as we did in the *pe*-plane case we also consider the projective embedding of the *qh*-structure. Applying a suitable regular polarity Ω to the pseudo-Euclidean plane one receives this model of the projectively embedded *qh*-plane, where the two half-planes represent the space-like world and the time-like world. The fixed line e represents one set of light-like lines. It is easy to translate concepts concerning ratios (of collinear points) and angles (of lines) to *qh*-ratios of lines and *qh*-angles of points. Parallel lines occur as points on a line parallel to e and orthogonal lines map to points on lines symmetric to e, such that rectangles $(ABCD)$ occur as trapecia $(abcd)$ with two sides parallel and symmetric to e. In particular, squares map to parallelograms.

For the *pe*-plane visualization method (α) can be based on Figure 13, if we modify the side ratios of the Golden Parallelograms according to (3). So we receive a visualization (α) in the *qh*-plane by dualizing Figure 13, see Figure 18.

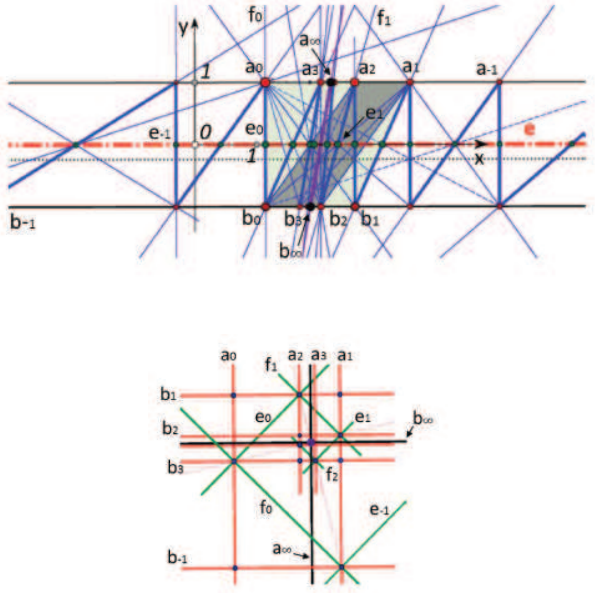


Figure 18: *Golden qh-polygon according to visualization method (α) by squares and rectangles (by dualizing the figure at top).*

In Figure 18 the points a, b represent the sides of the limit-quadrangle and their connection represents the pole of the spiral polygon (at right). Thereby the ratio of a with respect to a_1, a_0 reads as

$$\begin{aligned} r(a,a_1,a_0 &= r(a,a_3,a_2) = \ldots = 1:(3-\phi) \\ &= 1/5(2+\phi) = 0.723\ldots \end{aligned} \tag{4}$$

Visualization method (β) does not work here: In this model *qh*-circles occur as parabolas with common tangent *e*. It is impossible to construct non-trivial triples of mutually (outward) touching parabolas having a common tangent. Similarly, also method (γ) is not suited to *qh*-planes.

4.2 Projective cases

4.2.1 The hyperbolic plane (see 3.3)

Here we just refer to [4] and Figure 12 in 3.3.

4.2.2 The elliptic plane (*el*-plane)

Place of action is the full projective plane endowed with an elliptic absolute polarity. There are no similarities and no squares in the el-plane, such that visualisation method (α) cannot be performed. Because of the finite length of elliptic lines method (β) works well up to a certain number n of circles, see Figure 19. By choosing a suitably small unit for the scaling of the elliptic line this number can be any finite number n. For visualization purposes, this might be sufficient. If the *el*-circle radius would exceed the length l of a line one had to replace it by a circle with a radius modulo 1, what makes visualization very confusing. For e.g. a chain of golden *el*-circles on the sphere model of the *el*-plane it could be an idea to start with the largest possible *el*-circle, i.e. a great circle, and work from large to small, see Figure 20.

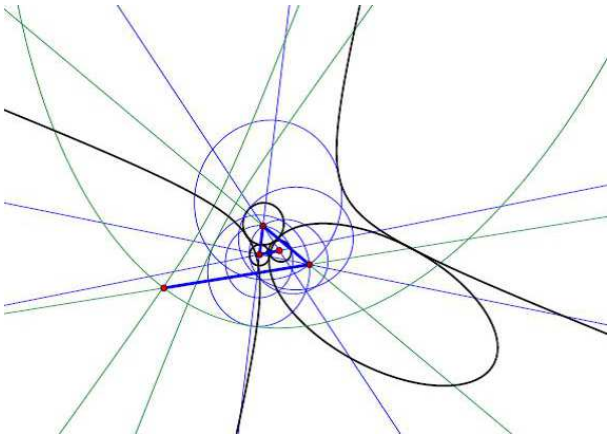


Figure 19: *Chain of Fibonacci circles in the projective model of the elliptic plane.*

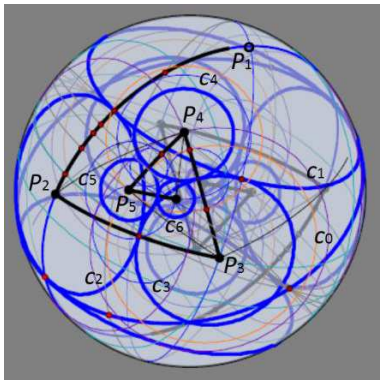


Figure 20: *Chain of Golden Circles on the Euclidean sphere representing the elliptic plane*

4.2.3 The dual Euclidean plane

is also called quasi-elliptic plane (*qe*-plane). Its absolute figure is a pair of conjugate imaginary lines i, j with a real intersection point U. A very convenient model of a qe-plane in the (projectively extendend) plane of visual perception (represented by a sheet of paper, the PC-screen or

the blackboard) takes U as a proper point and i and j as the fixed lines of the (Euclidean) "right angle involution" in the pencil of lines with support U (see Figure 21 left). We use homogeneous Euclidean coordinates in both, the projectively enlarged Euclidean plane (*e*-plane) and the *qe*-plane. The transfer from the *e*-plane to the *qe*-plane can be carried out by the (regular) polarity $\Omega : \mathrm{P} \to \mathrm{L}$ defined by the regular imaginary conic section $x_0^2 + x_1^2 + x_2^2 = 0$ and the transformation matrix

$$T_\Omega = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{5}$$

In the *e*-plane the parabolic distance measure in the point set P and the elliptic angle measure in the line set L are described by the usual Pythagoras formula (6) and Brauner's formula (7), see [2]:

$$d(P,Q) = \sqrt{(x_Q - x_P)^2 + (y_Q - y_P)^2} \tag{6}$$

with $x = \frac{x_1}{x_0}, y = \frac{x_2}{x_0}$ and

$$\tan\angle(p,q) = \sqrt{(-cr(p,p^\perp,q,q^\perp)}, \tag{7}$$

with $p^\perp$ *e*-orthogonal p, etc. These measures become a parabolic *qe*-angle measure in the line set L and an elliptic *qe*-distance measure in the point set P of a *qe*-plane. In our visualization of the *qe*-plane the *qe*-distance of two points P,Q which are non-collinear with U appears as the Euclidean angle between the lines UP, UQ, see Figure 21 (left). Thus the formulae (6), (7) just exchange their roles. Note that, similar to the isotropic case, *qe*-circles are not defined as (planar) point sets having constant distance from a centre point, but as conic sections touching the complex absolute lines i and j. This means that, in our visualization of a *qe*-plane, the absolute point U is a common (Euclidean) focus of the conic sections representing of *qe*-circles. One can choose one of the qe-circles as unit circle c and, similar to the isotropic case, extend the distance measure also to two parallel points $R,S(U)$, that are collinear with U : That additional distance can again be named *spacing* (*Sperrung*). It is defined by the difference of ratios as

$$d(R,S) := |r(S,E,U) - r(R,E,U)|, \tag{8}$$

with the point $E \in c$ acting as *unit point* on UR , see Figure 21 (right).

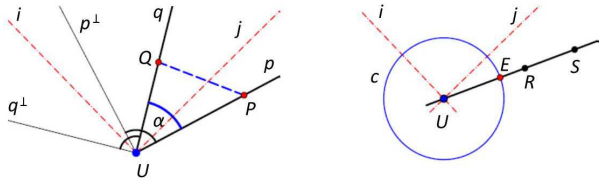


Figure 21: *qe-distance of non-parallel points (left) and of parallel points (right)*

(α) The dual of a Euclidean rectangle with sides a_1, b_1, a_2, b_2 is a quadrangle A_1, B_1, A_2, B_2 the diagonals e, f are orthogonal and intersect in U, (see Figure 22). For the dual of a Euclidean square the points $E := (A_1B_1) \cap (A_2B_2)$ and $F := (A_1B_2) \cap (A_2B_1)$ additionally are on orthogonal lines through U.

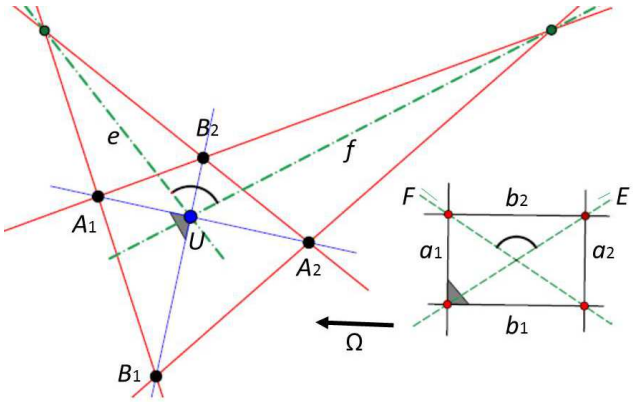


Figure 22: *qe-rectangles are quadrangles with orthogonal diagonals through U.*

So it is possible to transfer directly the Euclidean nested set of Golden Rectangles to the *qe*-plane, see Figure 23.

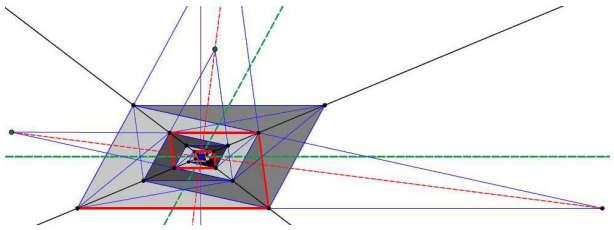


Figure 23: *Nested set of Golden qe-rectangles and a Golden qe-spiral polygon.*

(β) As a *qe*-circle is a conic section having U as one of its foci. Figure 24 shows that it is possible to construct a sequence of *qe*-circles, each touching the former two. Dealing with *radii*, however, would require a definition of *qe*-circles as a *qe*-distance set, which is not possible. This is why visualization method (β) does not work.

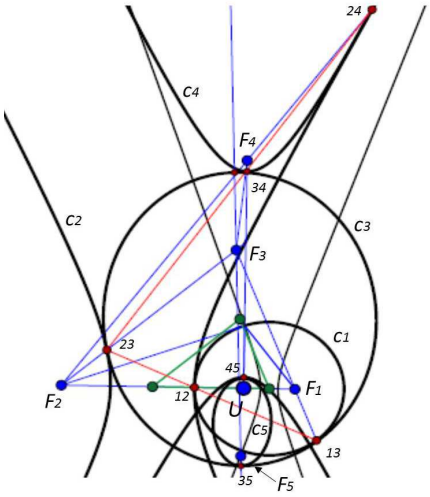


Figure 24: *A chain of qe-circles, one touching the two former qe-circles.*

(γ) To transfer the Euclidean Golden Root Spiral (Figure 8)) into the *qe*-plane model we have to construct a trilateral with angle ratio $\phi : \sqrt{\phi} : 1$. As Euclidean rotations with centre U are admitted, it becomes possible to also construct a *qe*-analog to the Golden Root Spiral, see Figure 25.

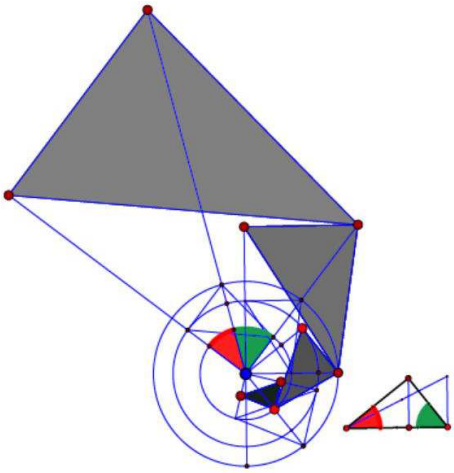


Figure 25: *A Golden quasi-elliptic root spiral.*

5 Conclusion and Outlook

The main subject of this paper is to show that visualizations of mathematical objects do not have to be performed in the classical Euclidean plane. Visualizations in other settings, like the Cayley-Klein planes or even spaces of higher dimensions, are justified as well. As an example we visualized the Golden Mean and the Fibonacci sequence in models of all possible Cayley-Klein planes. When doing so, visualisation methods based on typical Euclidean figures and properties have to be replaced by other methods. In this paper we propose three methods, which all work well in the Euclidean case. To transfer these methods in Cayley-Klein planes we use models conveniently adapted again to the plane of visual perception. Convenient means that constructions can be performed with available graphics software tools. In our paper we used Cinderella 2.8 for the figures. This CAD-software is distinguished by providing (planar) hyperbolic and elliptic geometry construction tools, too (see e.g. [15]). In some, but not all, Cayley-Klein planes we get visualizations of the Golden Mean and the Fibonacci sequence with modifications of the proposed Euclidean concepts.

Metallic Means generalize the Fibonacci sequence and the Golden Mean, see e.g. [13], [8]. The three presented methods could also be applied to visualize (generalized) Metallic Means. But as they all are defined as positive solutions of quadratic equations there will not occur essentially new results. Van der Laans and Rosenbuschs cubic generalizations of the Golden Mean have three-dimensional (Euclidean) visualizations by nested sets of boxes, thus generalizing the method (α) using squares and rectangles, see e.g. [5], [6] and [10]. Higher dimensional visualizations by nested sets of boxes are also known for Metallic Means, see [11]. For visualizations in non-Euclidean and Minkowski-spaces the method (β) (with hyper-spheres instead of circles, see [4]) seems to be natural and it is often the only possible method. Thereby one has to construct a chain of hyper-spheres, where the *n*th touches the former (*n*-1) hyper-spheres and their radii are proportional to Fibonacci numbers or elements of a geometric sequence. If we choose the radii according to a geometric sequence with proportionality factor ϕ or e.g. the Silver Mean one might call the occurring simplices of the centres of consecutive hyper-spheres Golden resp. Silver simplices.

References

[1] A. BEUTELSPACHER, B. PETRI, *Der goldende Schnitt*. Spektrum Akad. Verlag, B.I Mannheim 1996, ISBN 978-5860254042.

[2] H. BRAUNER, *Geometrie projektiver Räume*. Vol.II. B.I Mannheim 1976, ISBN 3-411-01513-6.

[3] O. GIERING, *Vorlesungen über höhere Geometrie*. Vieweg, Braunschweig 1982, ISBN 3-528-08492-8.

[4] S. MICK, G. WEISS, Fibonacci Triangles and Circle Chains and a Golden Bi-Arc Spiral in Non-Euclidean Planes. In H.-P. Schröcker and M. Husty, editors, *Proceedings of the 16th International Conference on Geometry and Graphics*, 415–422. Innsbruck University Press, 2014, ISBN 978-3-902936-46-2.

[5] R. PADOVAN, Dom Hans van der Laan and the Plastic Number. *Nexus IV, Architecture and Mathematics*, 181–193, 2002.

[6] L. ROSENBUSCH, *Räumliche Proportionen*. In O. Niewiadomski, editor, *Geometrie, Kunst und Wissenschaft 06*, Hauschild Verlag 2007, ISBN 978-38975-366-6.

[7] H. SACHS, *Ebene Isotrope Geometrie*. Vieweg, Braunschweig 1987, ISBN 978-3528084547.

[8] V. W. DE SPINADEL, The Family of Metallic Means, *Visual Mathematics* **1 (3)** (1999)

[9] H. WALSER, *The Golden Section* (transl. P. Hilton). The Mathem. Assoc. of America 2001, ISBN 0-8835-534-8.

[10] G. WEISS, V. W. DE SPINADEL, Remarks to classical cubic problems and the mean values of van der Laan and Rosenbusch, *Prooceedings of the 14th International Conference on Geometry and Graphics*, Article 235, 1–11, 2010.

[11] G. WEISS, From the Golden Rectangle to the Laan Box and more, *J. of Mathematics and Design*, **11** (2012), 14, ISBN 1515-7881.

[12] G. WEISS, V.W. DE SPINADEL, Bi-Arc Spirals an Minkowski Planes, In H.-P. Schröcker and M. Husty, editors, *Proceedings of the 16th International Conference on Geometry and Graphics*, 115–120. Innsbruck University Press, 2014, ISBN 978-3-902936-46-2.

[13] G. WEISS, V. W. DE SPINADEL, From George Odom to a new System of Metallic Means, *J. of Mathematics and Design* **12** (2013), 14, ISBN 1515-7881.

[14] `en.wikipedia.org/wiki/Golden-ratio`

[15] `http://www.heise.de/download/cinderella`

Gunter Weiss
e-mail: weissgunter@hotmail.com
University of Technology Vienna
Karlsplatz 13, 1040 Vienna, Austria

Sybille Mick
e-mail: mick@tugraz.at
University of Technology Graz
Kopernikusgasse 24, A-8010 Graz, Austria

Stručni rad
Prihvaćeno 24. 5. 2014.

ANDREJ NOVAK
ANDRIJA ŠTAJDUHAR

Primjena linearne i nelinearne jednadžbe provođenja topline u obradi digitalne slike

Application of Linear and Nonlinear Heat Equation in Digital Image Processing

ABSTRACT

We will explore the application of partial differential equations on digital images. We will show how to use the heat equation to eliminate noise in an image, highlight important elements and prepare it for possible further processing. We also show known heat equation's theoretical results in a methodical sequence and then derive simple numerical schemes based on the finite differences method. Guided by the idea of image structure preservation, for example edge preservation, the central part of this article introduces Perona-Malik equation as an example of a nonlinear heat equation. We conclude by comparing linear and nonlinear heat equation application on a couple of test images.

Key words: heat equation, Perona-Malik equation

MSC2010: 68U10

Primjena linearne i nelinearne jednadžbe provođenja topline u obradi digitalne slike

SAŽETAK

Istraživat ćemo primjenu parcijalnih diferencijalih jednadžbi u obradi digitalne slike. Primjenjujemo jednadžbu provođenja topline kako bismo na slici uklonili šum, istaknuli važne elemente i pripremili je za eventualnu daljnju obradu. Metodičkim slijedom dajemo teorijske značajke linearne difuzije, a zatim izvodimo jednostavne numeričke sheme temeljene na metodi konačnih razlika. Vođeni idejom očuvanja struktura na slici, primjerice rubova, u središnjem dijelu članka uvodimo Perona-Malikovu jednadžbu kao primjer nelinearne jednadžbe provođenja topline. Završavamo s usporedbom primjene linearne i nelinearne jednadžbe provođenja topline na testnim slikama.

Ključne riječi: jednadžba provođenja topline, Perona-Malikova jednadžba

Parcijalne diferencijalne jednadžbe uvele su novi pogled na obradu digitalne slike. Uspješnost metoda koje ih koriste nije iznenađujuća, budući da su takve jednadžbe polučile uspjeh i u drugim područjima, primjerice fizici, kemiji, elektrotehnici, graditeljstvu i drugdje. Dostupni su opsežni matematički rezultati, što omogućuje stvaranje jednostavnih numeričkih algoritama koje ćemo u ovom članku predstaviti. Metode temeljene na parcijalnim diferencijalnim jednadžbama jedne su od metoda u obradi slike koje imaju najbolje matematičke temelje, a razumijevanje ovih metoda vodilo je otkriću brojnih novih.

1 Linearna jednadžba provođenja topline

U nastavku će nam od interesa biti linearna jednadžba provođenja topline,

$$u_t - \Delta u = 0, \tag{1}$$

uz koju idu i prikladni početni i/ili rubni uvjeti koje ćemo kasnije navesti. Ovdje pretpostavljamo da je vremenska varijabla $t > 0$, a prostorna $x \in U$, gdje je $U \subset \mathbb{R}^n$ otvoren skup. Nepoznata je funkcija $u : U \times [0,\infty\rangle \to \mathbb{R}$, $u(x,t)$, a operator Δ djeluje na u obzirom na x, odnosno $\Delta u = \Delta_x u = \sum_{i=1}^{n} u_{x_i x_i}$. Ova jednadžba poznata je i pod nazivom *difuzijska jednadžba*.

1.1 Fizikalna interpretacija difuzijskog procesa

Poimanje fizikalnih procesa koji izjednačuju koncentraciju između povezanih područja prilično je intuitivno. Ono se može matematički formulirati Fickovim zakonom koji za jednu prostornu dimenziju glasi

$$\mathbf{F} = -A\frac{\mathrm{d}u}{\mathrm{d}x}, \tag{2}$$

gdje je tok $\mathbf{F}$ količina supstance po jedinici prostora i vremenu (npr. u $\frac{mol}{m^2 s}$), A je difuzijski koeficijent, u je koncentracija (npr. u $\frac{m^2}{s}$) te x varijabla. Općeniti Fickov zakon glasi

$$\mathbf{F} = -A\nabla u. \tag{3}$$

Tok $\mathbf{F}$ je uzrokovan gradijentom koncentracije u. Odnos između $\mathbf{F}$ i ∇u opisan je *difuzijskom matricom A*, koja je pozitivno definitna. Slučaj kada su $\mathbf{F}$ i ∇u paralelni naziva se *izotropan*, a u općem, *neizotropnom* slučaju, $\mathbf{F}$ i ∇u nisu paralelni.

Difuzija premješta masu bez gubitaka postojeće ili stvaranja nove. Ta činjenica može se opisati jednadžbom kontinuiteta

$$u_t = -\operatorname{div}\mathbf{F}.$$

Ako (3) umetnemo u jednadžbu kontinuiteta dobivamo *difuzijsku jednadžbu*

$$u_t = \operatorname{div}(A\nabla u),$$

što je upravo linearna jednadžba provođenja topline ako je $A = I$ jedinična matrica. U obradi slika koncentraciju iz ovih razmatranja možemo poistovjetiti s intenzitetom boje na određenom mjestu na slici (piksel).

1.2 Početno-rubni problem

Neka je $T > 0$, definiramo parabolički cilindar

$$U_T = U \times \langle 0, T],$$

i parabolički rub

$$\Gamma_T = \bar{U_T} - U_T.$$

U nastavku želimo riješiti početno-rubni problem

$$\begin{cases} u_t - \Delta u = 0 & \text{na } U_T \\ u = g & \text{na } \Gamma_T. \end{cases} \tag{4}$$

Iz klasične teorije parcijalnih diferencijalnih jednadžbi [2] poznat je sljedeći rezultat.

Teorem 1 (Jedinstvenost na ograničenim domenama) *Neka je $g \in C(\Gamma_T)$. Tada postoji najviše jedno rješenje $u \in C_1^2(U_T) \cap C(\bar{U_T})$ problema (4).*

Valja napomenti da je rješenje na neograničenim domenama, uz uvjet da je funkcija g neprekidna i ograničena, također jedinstveno ali i glatko zato što je dobiveno konvolucijom

$$u(x,t) = \frac{1}{(4\pi t)^{n/2}} \int_{\mathbb{R}^n} e^{-\frac{|x-y|^2}{4t}} g(y)\, dy \;\; (x \in \mathbb{R}^n, t > 0) \tag{5}$$

Vidimo da rješavanje jednadžbe provođenja topline zapravo znači konvoluiranje početne temperaturne distribucije, odnosno početnog uvjeta s Gaussovom funkcijom s parametrom standardne devijacije $\sigma = 2t$. Vrlo je zanimljivo te bismo to ovdje htjeli posebno istaknuti, da iz toga slijedi da je primjena jednadžbe provođenja topline ekvivalentna poznatoj tehnici Gaussovog izglađivanja digitalne slike. Takav postupak poznati je filtar koji se koristi za izglađivanje slika težinskim usrednjavanjem vrijednosti unutar određenog područja (npr. konvolucijom). Upravo iz toga dolazi ideja korištenja difuzije u obradi slike.

Zaključujemo uvodni teorijski dio, a u nastavku ćemo prethodna teorijska saznanja primjeniti u obradi digitalne slike.

1.3 Rješenje metodom konačnih razlika

Želimo početni problem (4) riješiti numerički. S obzirom na to da je slika zapravo matrica, u nastavku uzimamo da je $U_T = [0,p] \times [0,q] \times \langle 0,T]$. Tada (4) ima oblik

$$\begin{cases} u_t = u_{xx} + u_{yy} & \text{na } U_T \\ u = g & \text{na } \Gamma_T. \end{cases} \tag{6}$$

Početnu diferencijalnu jednadžbu ćemo primjenom Taylorovog teorema zamijeniti s diferencijskom jednadžbom. Diskretiziramo pravokutnik $[0,p] \times [0,q]$, gdje su p i q dimenzije slike, s ciljem uspostavljanja odnosa između čvorova u diskretnoj mreži i piksela na slici, između kojih je u obje prostorne dimenzije jedinična udaljenost. Primjenom Taylorovog teorema parcijalne derivacije zamjenjujemo konačnim razlikama

$$u_t \approx \tfrac{u_{i,j}^{k+1} - u_{i,j}^k}{\Delta t},$$
$$u_{xx} \approx \tfrac{u_{i+1,j}^k - 2u_{i,j}^k + u_{i-1,j}^k}{\Delta x^2},$$
$$u_{yy} \approx \tfrac{u_{i,j-1}^k - 2u_{i,j}^k + u_{i,j+1}^k}{\Delta y^2},$$

što uvrštavanjem u (6) daje

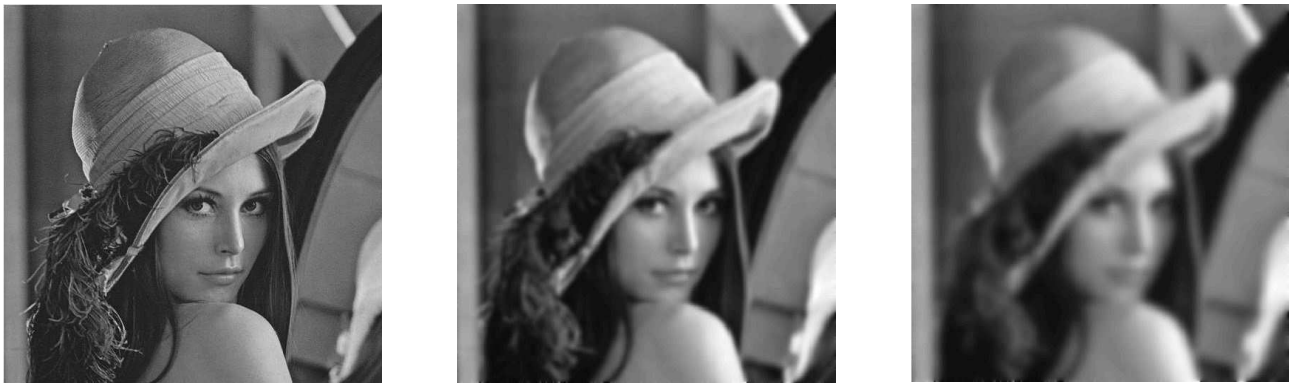
$$\frac{u_{i,j}^{k+1} - u_{i,j}^k}{\Delta t} = \frac{u_{i+1,j}^k - 2u_{i,j}^k + u_{i-1,j}^k}{\Delta x^2} + \frac{u_{i,j-1}^k - 2u_{i,j}^k + u_{i,j+1}^k}{\Delta y^2}.$$

Želimo promatrati evoluciju slike kroz vijeme u diskretnim vremenskim koracima, što znači da nas za danu sliku zanima vrijednost piksela na istom mjestu u idućem vremenskom koraku. S obzirom na to da je $\Delta x = \Delta y = 1$ dobivamo,

$$u_{i,j}^{k+1} = u_{i,j}^k + \Delta t \left[-4u_{i,j}^k + u_{i+1,j}^k + u_{i-1,j}^k + u_{i,j+1}^k + u_{i,j-1}^k \right].$$

Prethodna razmatranja možemo i jednostavno implementirati u programskom paketu SCILAB[1].

[1] SCILAB je besplatni program otvorenog koda za znanstveno računanje i numeričke simulacije koji se koristi u znanosti i industriji. SCILAB ima gotovo istu sintaksu kao i MATLAB, a više o programu se može pronaći na `www.scilab.org`



(a) Originalna slika (b) T = 5 (c) T = 20

(d) T = 50 (e) T = 100 (f) T = 300

Slika 1: *Otapanje početne slike (a) s vremenskim korakom* $\Delta t = 0.1$. *Na slikama (b), (c), (d), (e) i (f) T predstavlja proteklo vrijeme.*

1.4 Simulacije

Posljednji dio ovog odjeljka prikazuje korištenje dosad izvedenih zaključaka na testnoj slici Lena. U programskom kôdu učitavamo sliku i zatim primjenjujemo izvedenu shemu.

```
u0 = imread('Lena.eps');
[p,q]=size(u0);

for t = 0:dt:T
    for i=2:(p-1)
        for j=2:(q-1)
            u_xx =
                (u0(i+1,j)-2*u0(i,j)+u0(i-1,j));
            u_yy =
                (u0(i,j+1)-2*u0(i,j)+u0(i,j-1));
            u(i,j) = u0(i,j) + dt*(u_xx+u_yy);
        end
    end
    u0=u;
end
```

2 Nelinearna difuzija

Unatoč tome što je linearna difuzija jednostavna i primjenjiva, ima nekoliko mana. Očiti problem kod takvog, Gaussovog izglađivanja nalazi se u tome što ne samo da izglađuje šum, već zamućuje važne elemente slike poput rubova, čineći ih tako težim za pronalaženje i analiziranje. Željeli bismo razviti alat koji nam omogućuje uklanjanje šuma na način koji bi zatim olakšao pronalaženje rubova i ostalih elemenata slike. To znači da se difuzija treba odvijati samo unutar zasebnih područja koja se nalaze na slici, „poštujući“ njihove postojeće rubove. Potrebno je difuziju prilagoditi tako da njeno djelovanje nije jednoliko na cijeloj slici, već ovisi o pojedinim pikselima i njihovim okolinama.

2.1 Perona-Malikova jednadžba

Ideja koju su prvi uveli Perona i Malik u [4] je prilagodba jednadžbe provođenja topline tako da difuzivnost ovisi o promjenama na slici. Prisjetimo se difuzijske jednadžbe iz prethodnog odjeljka. Za difuzivnost A uzeli smo jediničnu

matricu i time dobili jednadžbu provođenja topline. No, A ne mora biti jedinična, čak niti konstantna. Vidjet ćemo da će nam upravo odabir prikladne difuzivnosti A omogućiti postizanje traženih svojstava difuzije. Kao što je spomenuto, želimo potaknuti izglađivanje unutar područja, za razliku od izglađivanja preko granica područja. To bismo mogli postići tako da vrijednost difuzivnosti bude 1 unutar područja, a 0 (ili barem blizu 0) na granicama. Difuzija će se tada odvijati unutar svakog od područja zasebno, ne prelazeći granice. Na žalost, ne možemo unaprijed znati gdje se na slici nalaze rubovi. Jedino što možemo jest ocijeniti koliko se, s obzirom na zadani piksel, njegova okolina mijenja.

Neka je funkcija $E = E(x,y)$ jedna takva ocjena definirana na slici. Bilo bi poželjno da E ima svojstva:

1. $E(x,y) = 0$ unutar svakog zasebnog područja
2. $E(x,y) = Ke(x,y)$ u svakoj rubnoj točki područja, gdje je e jedinični vektor normale na rub u točki (x,y), a K *lokalni kontrast*, odnosno razlika između intenziteta sive slijeva i zdesna od ruba.

Difuzivnost A možemo odabrati tako da bude funkcija koja ovisi o ocjeni koju smo upravo naveli, odnosno neka je $A = f(\|E\|_2)$. Prema dosadašnjim razmatranjima, A bi trebala biti nenegativna padajuća funkcija takva da je $f(0) = 1$. Na taj način difuzija će se odvijati uglavnom unutar područja i neće imati utjecaja na rubovima gdje je ocjena E velika. Srećom, vidjet ćemo da upravo najjednostavnija ocjena, gradijent $E = \nabla u$ daje izvrsne rezultate. Uzevši u obzir prethodno došli smo do Perona-Malikove jednadžbe,

$$\begin{cases} u_t = \operatorname{div}\left(f(|\nabla u|^2)\nabla u\right) & \text{na } U_T \\ u = g & \text{na } \Gamma_T \end{cases}. \tag{7}$$

Zbog traženih svojstava funkcije difuzivnosti koje se najčešće koriste su

$$f(s^2) = \frac{1}{1+\frac{s^2}{\lambda^2}} \tag{8}$$

i

$$f(s^2) = e^{-\frac{s^2}{2\lambda^2}}, \tag{9}$$

gdje je $\lambda > 0$. Obje funkcije su monotono padajuće i $f(s^2) = 1$ za $s = 0$, i $f(s^2) = 0$ kada $s \longrightarrow \infty$.

2.2 Teorijski rezultati za jednodimenzionalni model

Promotrimo sada jednodimenzionalni problem

$$\begin{cases} u_t = \operatorname{div}\left(f(u_x^2)u_x\right) & \text{na } U \times \langle 0,\infty\rangle \\ \dfrac{\partial u}{\partial x} = 0 & \text{na } \partial U \times \langle 0,\infty\rangle \\ u = g & \text{na } U \times \{t = 0\} \end{cases}. \tag{10}$$

Za f uzmimo funkciju

$f(s^2) = \frac{1}{1+\frac{s^2}{\lambda^2}}$.

Neka je funkcija toka dana s

$\Phi(s) = s \cdot f(s^2) = s \cdot \frac{1}{1+\frac{s^2}{\lambda^2}} = \frac{\lambda^2 s}{\lambda^2+s^2}$.

Derivacija ove funkcije je

$\Phi'(s) = \frac{\lambda^2(s^2-\lambda^2)}{(\lambda^2+s^2)^2}$.

Dakle, ova funkcija postiže ekstrem kada je $s^2 = \lambda^2$. Kako je $\lambda > 0$ imamo

$\Phi'(s) < 0,\ |s| > \lambda,$
$\Phi'(s) > 0,\ |s| < \lambda.$

Perona-Malikova jednadžba za jednu dimenziju glasi

$$\begin{aligned} u_t &= \left(f(s^2)\cdot u_x\right)_x \\ &= f'(u_x^2)\cdot 2u_x u_{xx} u_x + f(u_x^2)\cdot u_{xx} \\ &= \left(f(u_x^2) + f'(u_x^2)\cdot 2u_x\right)\cdot u_{xx} \\ &= \Phi'(u_x)\cdot u_{xx}. \end{aligned}$$

Za velike vrijednosti koeficijent $\Phi'(u_x)$ očito postaje negativan, što vodi na difuziju *unatrag*. U dvodimenzionalnom slučaju bismo, kao u [1], dobili

$$u_t = \Phi'(|\nabla u|^2)u_{\eta\eta} + f(|\nabla u|^2)u_{\xi\xi}.$$

Ovdje su η i ξ koordinate paralelne s ∇u i okomite na ∇u, redom, što znači da je difuzija *unaprijed* duž tangenti na rješenje u, a *unaprijed-unatrag* duž smjera ∇u. Očekujemo da će blizu ruba na slici ∇u biti velik, što znači difuziju unatrag u smjeru gradijenta, rezultat čega je izoštravanje rubova umjesto njihova zamućenja. Prethodno navedeno objašnjava kako Perona-Malikova jednadžba ne samo da čuva rubove, već ih i dodatno ističe.

Definicija 1 ([9])

Problem je dobro uvjetovan *ako ima jedinstveno rješenje koje neprekidno ovisi o početnim uvjetima. Za problem koji nije dobro uvjetovan kažemo da je* loše uvjetovan.

Poznato je da je difuzija unatrag loše uvjetovana. Prema tome, postojanje difuzije unatrag u Perona-Malikovoj jednadžbi sugerira da bi i ta jednadžba mogla biti loše uvjetovana.

Definicija 2 (Slabo rješenje za jednodimenzionalni slučaj, [9])

Za lokalno integrabilnu funkciju $u(x,t)$ kažemo da je slabo rješenje *Perona-Malikove jednadžbe ako je $\int_U (u^2+u_x^2)\,dx$ uniformno ograničen za ograničeni t i ako za svaku funkciju $\phi \in C_0^1(\mathbb{R}\times\mathbb{R}_+)$ vrijedi*

$$\int\int [\phi_t u - \phi_x f(u_x^2)u_x]\,dx\,dt = 0.$$

U [3] je pokazano da ako postoji slabo rješenje jednodimenzionalnog problema, početni uvjet mora biti beskonačno puta diferencijabilan u područjima gdje se odvija difuzija unatrag ($|u_x| > \lambda$). To pokazuje da je moguće da slabo rješenje uopće ne postoji. Očito su afine funkcije oblika $u(x,t) = ax+b$ rješenja, no one su nestabilne u smislu da ako promjenimo početni uvjet proizvoljno malo, rješenje možda neće postojati.

Teorem 2 (Nepostojanje rješenja za jednodimenzionalni slučaj)

Za dani g takav da je $g'(x) = 0$ na ∂U i $g'(x) > \lambda$ na samo jednom kompaktu u U, na primjer $g'(x) > \lambda$ na $Q = \langle x_0, y_0\rangle \subset\subset U$ i $|g'(x)| < \lambda$ na $U\backslash\bar{Q}$, jednodimenzionalni problem (10) nema globalnog slabog rješenja u $C^1(U)$.

Teorem 3 (Jedinstvenost lokalnog slabog rješenja za jednodimenzionalni slučaj)

Pretpostavimo da su u i v lokalna slaba rješenja (10) na U_T s jednakim početnim uvjetima g, gdje je g analitička funkcija, $(g')^2 - \lambda^2$ ima samo jednostruke nultočke i difuzivnost f je analitička funkcija. Tada je $u(x,t) \equiv v(x,t)$ na U_T.

Kao što vidimo iz prethodnih osnovnih teorijskih rezultata, teorija postaje prilično složena čak i za jednodimenzionalni slučaj, što nije ni čudno s obzirom da se radi o nelinearnoj parcijalnoj diferencijalnoj jednadžbi.

Pogledajmo sada praktičnu primjenu Perona-Malikove jednadžbe, kao što smo učinili i s linearnom jednadžbom.

2.3 Rješenje metodom konačnih razlika

Programska implementacija analogna je onoj za linearnu difuziju iz prethodnog odjeljka. Iako nije odmah vidljivo kako ona izgleda, budući da koristimo difuzivnost koja uključuje funkciju koja je rješenje jednadžbe. Zbog toga je potrebno pojednostavniti izraz iz problema (7). Računamo:

$$\begin{aligned} u_t &= \operatorname{div}\Big(f(|\nabla u|^2)\cdot\nabla u\Big) \\ &= \operatorname{div}\Big(f(u_x^2+u_y^2)\cdot[u_x,u_y]\Big) \\ &= \Big(f(u_x^2+u_y^2)\cdot u_x\Big)_x + \Big(f(u_x^2+u_y^2)\cdot u_y\Big)_y. \end{aligned} \tag{11}$$

Računanjem parcijalnih derivacija te primjenom Schwarzovog teorema slijedi

$$\begin{aligned} u_t =& f'(u_x^2+u_y^2)\Big(2u_xu_{xx}+2u_yu_{yx}\Big)u_x + f(u_x^2+u_y^2)u_{xx} \\ &+ f'(u_x^2+u_y^2)\Big(2u_xu_{xy}+2u_yu_{yy}\Big)u_y + f(u_x^2+u_y^2)u_{yy} \\ =& 2f'(u_x^2+u_y^2)\left(u_x^2u_{xx}+2u_xu_yu_{xy}+u_y^2u_{yy}\right) \\ &+ f(u_x^2+u_y^2)(u_{xx}+u_{yy}). \end{aligned}$$

Pojednostavljena Perona-Malikova jednadžba je jednostavna za programsku implementaciju.

```
T = 20;
dt = 0.2;
lambda = 1;
f =@(s) 1/(1+s/lambda^2);
df =@(s) -((1/lambda^2)/((1+s/lambda^2)^2));

for t = 0:dt:T
  for i=2:(p-1)
    for j=2:(q-1)
        ux=(u0(i+1,j)-u0(i-1,j))/2;
        uy=(u0(i,j+1)-u0(i,j-1))/2;
        uxx=(u0(i+1,j)-2*u0(i,j)+u0(i-1,j));
        uyy=(u0(i,j+1)-2*u0(i,j)+u0(i,j-1));
        uxy=(u0(i+1,j+1)-u0(i+1,j-1)-u0(i-1,j+1)
             +u0(i-1,j-1))/4;

        u(i,j)=u0(i,j)+2*dt*...
        ((df(ux^2+uy^2)*(ux*uxx*ux+uy*uyy*uy+2*ux*uy*uxy))
          + f(ux^2+uy^2)*(uxx+uyy));
    end
  end
  u0=u;
end
```

Uz dulji protok vremena dobivamo jači utjecaj na početnu sliku. Rezultati dobiveni eksperimentima prikazani su na slici 2. Primjećujemo očuvanje rubova i uklanjanje „zrnatih“ elemenata na slici.

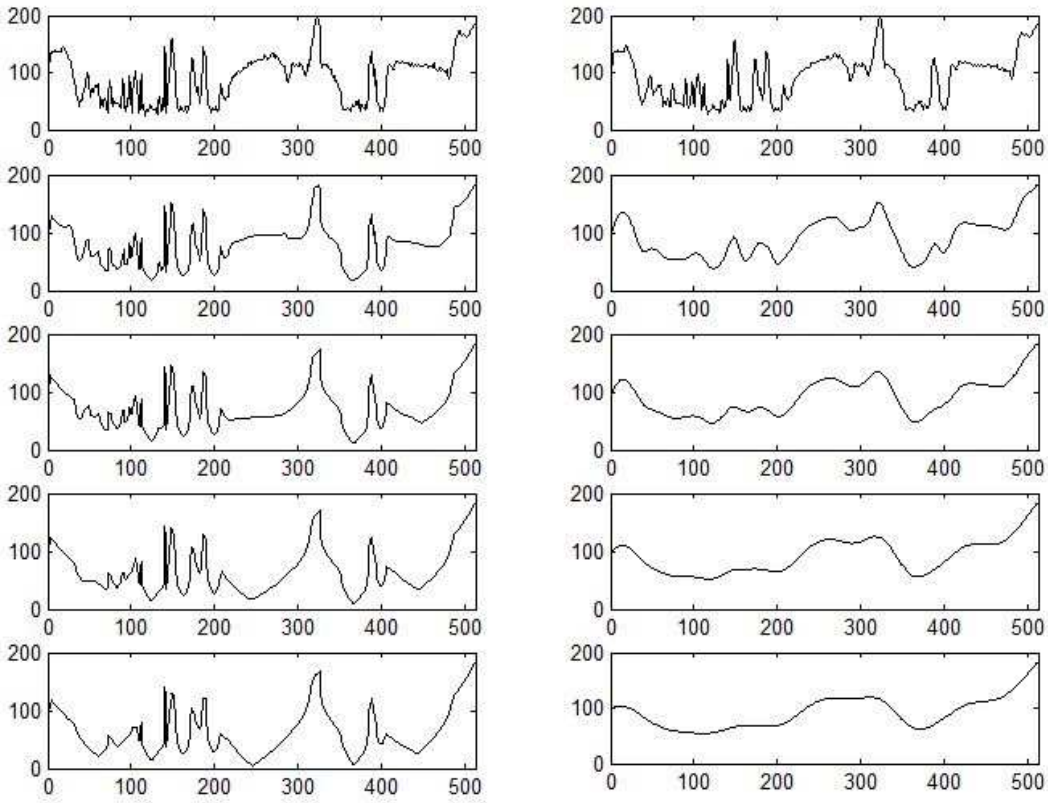
(a) Originalna slika (b) T = 20 (c) T = 50 (d) T = 100

Slika 2: *Otapanje početne slike (a) koristeći Perona-Malikovu jednadžbu s vremenskim korakom* $\Delta t = 0.2$. *Na slikama (b), (c) i (d) T predstavlja proteklo vrijeme,* $\lambda = 1$.

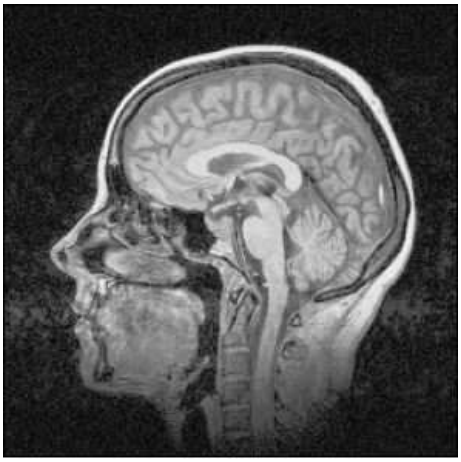
Unatoč tome što je algoritam osjetljiv na rubove, nakon dovoljno vremena neki rubovi, inače nedovoljno jasni, nestaju u stapanju s okolinom.

2.4 Usporedba s linearnom difuzijom

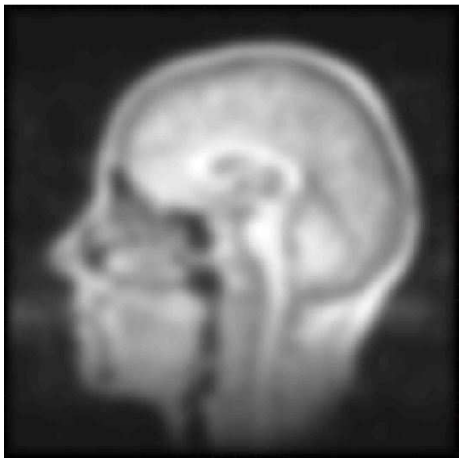
Za kraj dajemo usporedbu djelovanja linearne i nelinarne difuzije. Prvo koristimo jednodimenzionalne podatke da bismo vidjeli kako nelinearna difuzija čuva rubove, za razliku od linearne koja nekritično ”otapa” cijelu sliku. Testni podaci dobiveni su presjekom slike Lene u smjeru *x* osi, dakle graf predstavlja intenzitet sive duž jedne od horizontalnih linija na testnoj slici. Usporedbu ova dva modela vidimo na slici 3. Na slici 4 dajemo usporedbu modela na dvodimenzionalnim podacima, odnosno na slici glave dobivenoj magnetskom rezonancom.



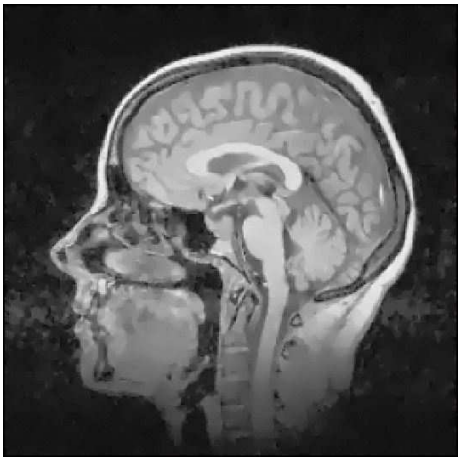
Slika 3: *Usporedba Perona-Malikove jednadžbe (lijevo) i linearnog modela (desno).*



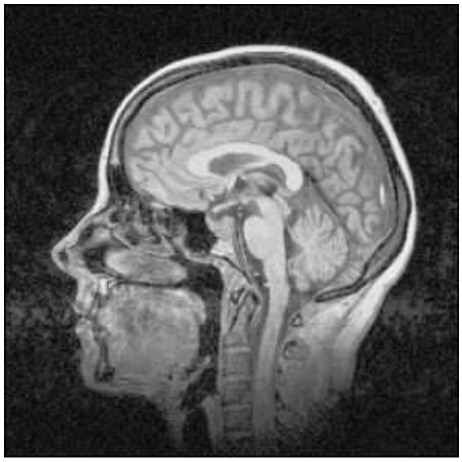
(a) Originalna slika



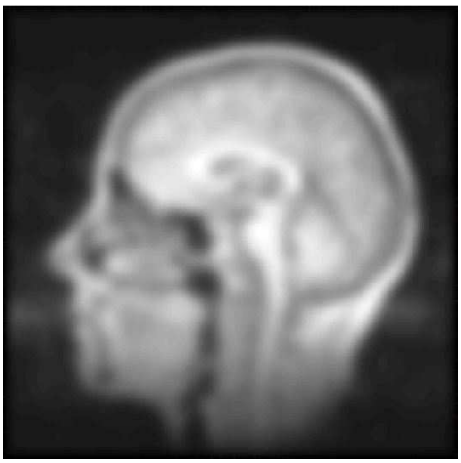
(b) Linearna difuzija, T = 10



(c) Nelinearna difuzija, T = 25



(d) Linearna difuzija, T = 200



(e) Nelinearna difuzija, T = 200

Slika 4: *Usporedba linearne i nelinearne difuzije na sliku dobivenu magnetskom rezonancom,* $\Delta t = 0.2$*, a T predstavlja proteklo vrijeme.*

Literatura

[1] L. ALVAREZ, F. GUICHARD, P. L. LIONS, J. M. MOREL, Axioms and fundamental equations of image processing, *Archive for rational mechanics and analysis* **123 (3)** (1993), 199−257.

[2] L. C. EVANS, Partial Differential Equations, *Providence, RI: American Mathematical Society*, Teh. izv., ISBN 0-8218-0772-2, 1998.

[3] S. KICHENASSAMY, The Perona–Malik Paradox, *SIAM Journal on Applied Mathematics* **57 (5)** (1997), 1328−1342.

[4] P. PERONA, J. MALIK, Scale-space and edge detection using anisotropic diffusion, *Proceedings of IEEE Computer Society Workshop on Computer Vision*, 16−22, 1987.

[5] A. QUARTERONI, R. SACCO, F. SALERI, *Numerical Mathematics*, Springer, 2000.

[6] A. ŠTAJDUHAR, *Primjena parcijalnih diferencijalnih jednadžbi u obradi digitalne slike*, Diplomski rad, Sveučilište u Zagrebu, 2013.

[7] J. WEICKERT, A review of nonlinear diffusion filtering, *Scale-space theory in computer vision*, Springer, 1−128, 1997.

[8] J. WEICKERT, *Anisotropic diffusion in image processing*, sv. 1, Teubner Stuttgart, 1998.

[9] M. WIELGUS, *Perona-Malik equation and its numerical properties*, Disertacija, Uniwersytet Warszawski, 2010.

Andrej Novak
andrej.novak@unidu.hr
Sveučilište u Dubrovniku
Odjel za elektrotehniku i računarstvo
Ćire Carića 4, 20000 Dubrovnik

Andrija Štajduhar
astajd@astajd.com
Sveučilište u Zagrebu
Hrvatski institut za istraživanje mozga
Šalata 12, 10000 Zagreb

Professional paper
Accepted 26. 12. 2014.

ANDREA K. MYERS-BEAGHTON
ALAN L. MYERS

The Moon Tilt Illusion

The Moon Tilt Illusion

ABSTRACT

The moon tilt illusion is the startling discrepancy between the direction of the light beam illuminating the moon and the direction of the sun. The illusion arises because the observer erroneously expects a light ray between sun and moon to appear as a line of constant slope according to the positions of the sun and the moon in the sky. This expectation does not correspond to the reality that observation by direct vision or a camera is according to perspective projection, for which the observed slope of a straight line in three-dimensional space changes according to the direction of observation. Comparing the observed and expected directions of incoming light at the moon, we derive a quantitative expression for the magnitude of the moon tilt illusion that can be applied to all configurations of sun and moon in the sky.

Key words: moon tilt, perspective projection, illusion

MSC 2000: 51N05

Iluzija nagiba mjeseca

SAŽETAK

Iluzija nagiba mjeseca zapanjujući je raskorak između svjetlosne zrake koja osvjetljava mjesec i smjera sunca. Ona se povećava, jer promatrač pogršno očekuje da zraka svjetla između sunca i mjeseca bude pravac konstantnog koeficijenta smjera s obzirom na položaj sunca i mjeseca na nebu. Ovakvo očekivanje ne odgovara stvarnosti kod koje je promatranje s direktnom osi pogleda, ili s kamerom, u skladu s perspektivom (centralnim projiciranjem) za koju se promatrani koeficijent smjera pravca u trodimenzionalnom prostoru mijenja s obzirom na os pogleda. Uspoređujući promatrane i očekivane smjerove zrake usmjerene na mjesec, izvodimo kvantitativan izraz za veličinu iluzije nagiba mjeseca koji se može primijeniti na sve položaje sunca i mjeseca na nebu.

Ključne riječi: nagib mjeseca, perspektiva, iluzija

1 The Nature of the Illusion

The photograph in Figure 1 provides an example of the moon tilt illusion. The moon's illumination is observed to be coming from above, even though the moon is high in the sky and the sun had set in the west one hour before this photo was taken. The moon is 45° above the horizon in the southeast, 80% illuminated by light from the sun striking the moon at an angle of 17° above the horizontal, as shown by the arrow drawn on the photograph. Our intuition (i.e., the incorrect perception that creates the illusion) is that given the relative positions of the sun and the moon, the light from the sun should be striking the moon from below. The moon tilt illusion is thus the perceived discrepancy between the angle of illumination of the moon that we observe (and can capture photographically with a camera pointed at the moon) and the angle that we expect, based on the known locations of the sun and the moon in the sky.

Figure 1: *Photograph of the moon tilt illusion. Picture taken one hour after sunset, with the moon in the southeast and sun already set in the west. Camera pointed upwards* 45° *from the horizon with bottom of camera parallel to the horizon.*

Rather surprisingly, little mention of the moon tilt illusion (much less a detailed explanation of why it occurs) can be found in astronomy books. Minnaert [1] gives a passing reference: "...the line connecting the horns of the moon, between its first quarter and full moon, for instance, does not appear to be at all perpendicular to the direction from sun to moon; we apparently think of this direction as being a curved line. Fix this direction by stretching a piece of string taut in front of your eye; however unlikely it may have seemed to you at first you will now perceive that the condition of perpendicularity is satisfied". A photograph taken by Lodriguss [2] shows a waxing moon and the setting sun in the same photo. The angle of 23° between the direction of the moon's illumination and the direction of the sun provides a striking illustration of the moon tilt illusion. An article by Schölkopf [3] documents the illusion in an experiment involving 14 subjects by having them indicate their expectation of how the moon's illumination should be oriented with respect to the position of the (visible) sun. He reports that an average discrepancy of 12° is perceived by the subjects between the observable versus expected orientation of the moon's bright limb. Schott's website entitled " 'Falsche' Mondneigung" ('False' Moon Tilt) [4] is devoted to the moon tilt illusion, and features illustrations and useful links. A paper by Glaeser and Schott [5], approaching the phenomenon via the principles of photography, shows that the magnitude of the illusion could in theory be measured through comparison of a close-up shot of the moon with a photograph containing both sun and moon, with the camera directed in a specified direction between them (although no equations are given). However, as they point out, in practice it is not feasible since even a wide-angle lens cannot capture both sun and moon in a photo with azimuth differences for which the illusion can be most clearly observed (between 90° and 180°). Berry [6] proposed a zenith-centered stereoscopic projection of the celestial sphere onto a flat surface with the moon tilt illusion defined as the angle between the projected great circle and a straight moon-sun line drawn on the flat surface "mimicking how we might see the sky when lying on our back looking up". Apparently there still persists a lack of consensus in the literature about the explanation of the moon tilt illusion and disagreement about the best way to measure it.

In this paper, our aim is to derive a quantitative expression for the magnitude of the moon tilt illusion experienced by an upright observer that is straightforward to apply to all configurations of sun and moon in the sky. We model the viewer's expectation of the direction of incoming light using vector geometry, which is appropriate for treating 3D straight lines such as the sun-moon light ray.

2 System of Coordinates and Definitions

Our analysis of the moon tilt illusion is based upon the known locations of the sun and moon in the sky. We adopt topocentric coordinates (instead of right ascension and declination) for the sun and moon, denoted by azimuth (ϕ) [1] and altitude (η). The altitude η is the angle between the sun (or moon) and the observer's local horizon. Recognizing that the altitude angle (η) is the complement of the polar angle (θ), we may rewrite azimuth and altitude (ϕ, η) as spherical coordinates (ϕ,θ). Spherical coordinates for the sun and moon are converted to Cartesian coordinates to allow vector manipulations such as dot and cross products.

2.1 Moon Pointer and Moon Tilt Angle

The moon pointer is defined as the vector *CP* in Figure 2, where *C* is the center of the moon and the vector *CP* has the observed slope of the moon-sun line at point *C*. The demarcation between illuminated and dark portions of the moon is called the terminator. Line *AB* connects the two "horns" of the terminator through the moon's center *C*. The moon pointer *CP* is the perpendicular bisector of line *AB*.

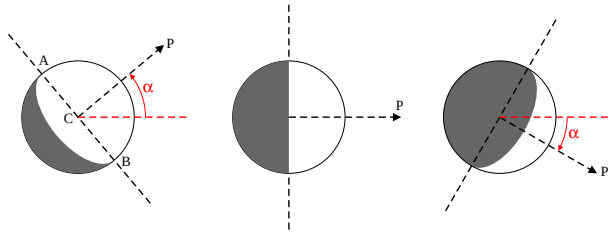


Figure 2: *Definition of moon pointer with* α *angle. From left to right,* $\alpha = 40°$ *(75% illumination),* $\alpha = 0°$ *(50% illumination),* $\alpha = -30°$ *(25% illumination).*

The moon tilt α is the signed angle of the moon's pointer with the horizontal, positive upward and negative downward. An equation for calculating this angle from the locations of the sun and moon is given in Section 4. Using the construction in Figure 2, the angle α may be found experimentally by taking a picture of the moon with the optical axis of the camera pointed at the moon and the bottom of the camera oriented horizontally. For example, for the photo in Figure 1, $\alpha = 17°$.

[1] In physics, the azimuthal angle is defined as positive for counter-clockwise (CCW) rotation from due north (*x*-direction), with the Cartesian coordinates satisfying the right-hand rule. In navigation, azimuth is defined as positive in the clockwise (CW) direction. We will use the CCW notation for calculations but revert to the more familiar navigational CW direction for the presentation of results in Section 5.

3 Cause of Moon Tilt Illusion

When we view the light ray at the moon, which is the only place we can photograph its direction, the slope with the horizontal (α) that we observe is exactly what one would expect from the principles of perspective projection that form the basis of human vision or photography.

The cause of the moon tilt illusion is simply that the observer is not taking into account the rules of perspective that dictate that the observed slope of the light ray will change when he turns his head to observe the moon and sun. This perceptual disconnect occurs because the observer cannot see the light ray itself, but only its starting position at the sun and the angle at which it strikes the moon. Without any other visual cues to provide more information, he is perceptually unable to envision how the slope of a visible line overhead changes with viewing angle due to perspective projection.

The changing-slope effect due to perspective projection is apparent in a video [7] which scans a long, straight string of lights along the Thames near London's Tower Bridge. All of the lights are at roughly the same distance from the ground. The moving video camera shows the observed slope of the string of lights varying continuously with camera motion: first sloping upwards from the ground on the left, then with zero slope in the middle, and finally sloping downwards to the right. For the moon illusion, the path of the light ray is invisible and we can observe its slope only at one end. If the sun-moon light ray were visible, we would see a straight line of varying slope just like the video and the illusion would vanish.

Knowing that light travels in straight lines in space but 'forgetting' that slope changes as the head turns along a line, the observer expects that when he scans from sun to moon he would see a straight line of constant slope, even though his head has moved. On the basis of this explanation, we calculate the observed angle of the moon tilt (α) and compare it with the expected angle (β) of the moon tilt based upon the known positions of the moon and the sun in the sky. The difference between the observed and expected angles (δ) quantifies the moon tilt illusion.

4 Observed and Expected Slope of Incoming Light

4.1 Observed Moon Tilt (α)

The principles of 2D perspective projection govern the viewing of a 3D line between two objects overhead by the human eye or a camera. The light ray from the sun that illuminates the moon is invisible in the sky, but we can observe its slope with the horizontal where it intersects the moon from the direction of the moon's illumination. The derivation of the tilt angle (α) between the observed incoming direction of light and the horizontal is straightforward but lengthy and is not given here because the equation for the observed tilt from the vertical (χ) is already well known.

Let ϕ_m be the azimuth of the moon and ϕ_s the azimuth of the sun; let $\Delta\phi = |\phi_s - \phi_m|$; let η_m be the altitude of the moon and η_s the altitude of the sun. The angle of the moon's tilt from the horizontal may be derived from an equation for the position angle of the moon's bright limb [8], [9]:

$$\tan\chi = \frac{\cos\eta_s \sin\Delta\phi}{\cos\eta_m \sin\eta_s - \cos\eta_s \sin\eta_m \cos\Delta\phi} \tag{1}$$

χ in this equation is called the position angle of the midpoint of the moon's bright limb measured from the north point of the disk. This may be written:

$$\tan\chi = \frac{\sin\Delta\phi}{\cos\eta_m \tan\eta_s - \sin\eta_m \cos\Delta\phi} \tag{2}$$

The desired angle with the horizontal (α) is the complement of χ so:

$$\boxed{\tan\alpha = \frac{\cos\eta_m \tan\eta_s - \sin\eta_m \cos\Delta\phi}{\sin\Delta\phi}} \tag{3}$$

4.2 Expected Moon Tilt (β)

An observer bases his expectation of the incoming direction of light at the moon on his knowledge of the 3D positions of the sun and moon as they appear to him in the sky, i.e., according to their height difference and horizontal distance apart. For example, in Figure 1, the upright viewer sees the light illuminating the moon from above, but he expects the light to come from below the horizontal, since the moon is higher than the sun. In the sky, there is an absence of visual cues by which the viewer could evaluate the distance of an object; thus the direction of light from sun to moon is assessed from their relative altitudes and azimuths as though sun and moon were equidistant [2] from the viewer. We represent this expected direction of light as a 3D vector $\mathbf{v}$, given by the difference of the unit vectors from the observer to the sun ($\hat{\mathbf{s}}$) and the moon ($\hat{\mathbf{m}}$):

$$\mathbf{v} = \hat{\mathbf{s}} - \hat{\mathbf{m}} \tag{4}$$

[2] This assumption is a natural consequence of the 2D perspective-projection basis of human vision. Since objects are projected bigger or smaller when closer or farther away, objects of apparent equal size will be judged as equidistant, in the absence of additional visual cues such as clarity or brightness. We note that even if observers take into account that the sun is much farther away from the earth than the moon, they will still experience an illusion by not considering perspective distortion. For example, for a setting sun they would expect the moon (in any position) to be illuminated from the horizontal, leading to an illusion equal to the observed α tilt.

The observer naively expects to view $\mathbf{v}$ without any perspective distortion. If the observer faced the vertical plane containing the sun and moon directly, the slope of $\mathbf{v}$ in this plane is simply the height difference of sun and moon divided by the horizontal distance between them. However, the observer must face the moon for his observation of the illusion. (If our eyes deviate from the azimuth of the moon, the *observed* angle α of the moon tilt would change). With knowledge of the position of the sun and the moon relative to his orientation facing the moon, the observer expects his view of $\mathbf{v}$ as it strikes the moon to be determined by this orientation. This is simply the orthogonal projection of $\mathbf{v}$ on the vertical plane at the moon. The vector $\mathbf{n}$ normal to the vertical projection plane is:

$$\mathbf{n} = m_x\,\hat{\mathbf{x}} + m_y\,\hat{\mathbf{y}} \tag{5}$$

where m_x and m_y are the x and y components of the unit $\hat{\mathbf{m}}$ vector. The unit normal vector is:

$$\hat{\mathbf{n}} = \frac{\mathbf{n}}{|\mathbf{n}|} = \frac{\mathbf{n}}{\sqrt{m_x^2 + m_y^2}} \tag{6}$$

The projection $\mathbf{v_p}$ on the vertical plane is:

$$\mathbf{v_p} = \mathbf{v} - \mathbf{v_n} \tag{7}$$

where $\mathbf{v_n}$ is perpendicular to the vertical plane with:

$$\mathbf{v_n} = (v \cdot \hat{\mathbf{n}})\hat{\mathbf{n}} \tag{8}$$

The horizontal unit vector lying in the vertical plane is $\hat{\mathbf{h}} = \hat{\mathbf{n}} \times \hat{\mathbf{z}}$. Since the tangent of an angle between two vectors is equal to the ratio of the cross and dot products, it follows that the desired angle β between $\mathbf{v_p}$ and $\hat{\mathbf{h}}$ is given by:

$$\tan\beta = \frac{|\mathbf{v_p} \times (\hat{\mathbf{n}} \times \hat{\mathbf{z}})|}{\mathbf{v_p} \cdot (\hat{\mathbf{n}} \times \hat{\mathbf{z}})} \tag{9}$$

This formula for β was chosen to avoid having to normalize $\mathbf{v_p}$. As shown in Appendix A, Eq. (9) may be written in terms of the Cartesian components of the unit moon vector $\hat{\mathbf{m}}$ and the unit sun vector $\hat{\mathbf{s}}$:

$$\tan\beta = \frac{|s_z - m_z|\sqrt{m_x^2 + m_y^2}}{s_x m_y - s_y m_x} \tag{10}$$

As shown in Appendix B, conversion of the Cartesian components of the moon and sun vectors to angles yields:

$$\tan\beta = -\frac{|\sin\eta_m - \sin\eta_s|}{\cos\eta_s \sin(\Delta\phi)} \tag{11}$$

This equation for β applies to waxing and waning moons in both hemispheres. The nuisance of insuring that the angle is in the right quadrant can be avoided by writing Eq. (11) in the form:

$$\boxed{\tan\beta = -\frac{(\sin\eta_m - \sin\eta_s)}{\cos\eta_s \sin(\Delta\phi)}} \tag{12}$$

where it is understood that $\Delta\phi = |\phi_s - \phi_m|$ and $|\Delta\phi| \leq 180^\circ$. The sign convention for the β pointer is the same as for the α pointer: a positive value for β corresponds to a direction upward from the horizontal and a negative value corresponds to a direction downward from the horizontal, pointing east or west depending on the location of the sun. Typically the altitude of the moon is higher than that of the sun and β is negative.

4.3 Magnitude of Moon Tilt Illusion

The moon tilt illusion is defined as the difference (δ) between the slope angle of the observed moon-sun line (α) and slope angle of the expected moon-sun line (β):

$$\boxed{\delta = \alpha - \beta} \tag{13}$$

We may apply this equation to the photograph in Figure 1. The locations of the sun and moon are the altitudes $\eta_m = 45^\circ$, $\eta_s = -15^\circ$, and an azimuth difference $\Delta\phi = 128^\circ$. The illumination of the moon in the photograph is 80%, which agrees with the calculated value [9]. From Eq. (3), $\alpha = 17^\circ$, which is confirmed by the photograph. Eq. (12) gives $\beta = -52^\circ$ and from Eq. (13), $\delta = 17 - (-52) = 69^\circ$, consistent with the viewer's expectation that the incoming light should be strongly angled from below the horizontal.

5 Discussion

We have presented a method for calculating the magnitude of the moon tilt illusion as the degree difference (δ) between the observed direction of the incoming light and the expected direction of incoming light. The model identifies sun/moon configurations ranging from no illusion (when the sun and moon are either close together or both on the horizon) to the strongest illusion (at the equator, when the moon is above the horizon and the azimuthal difference between moon and sun is 180°).

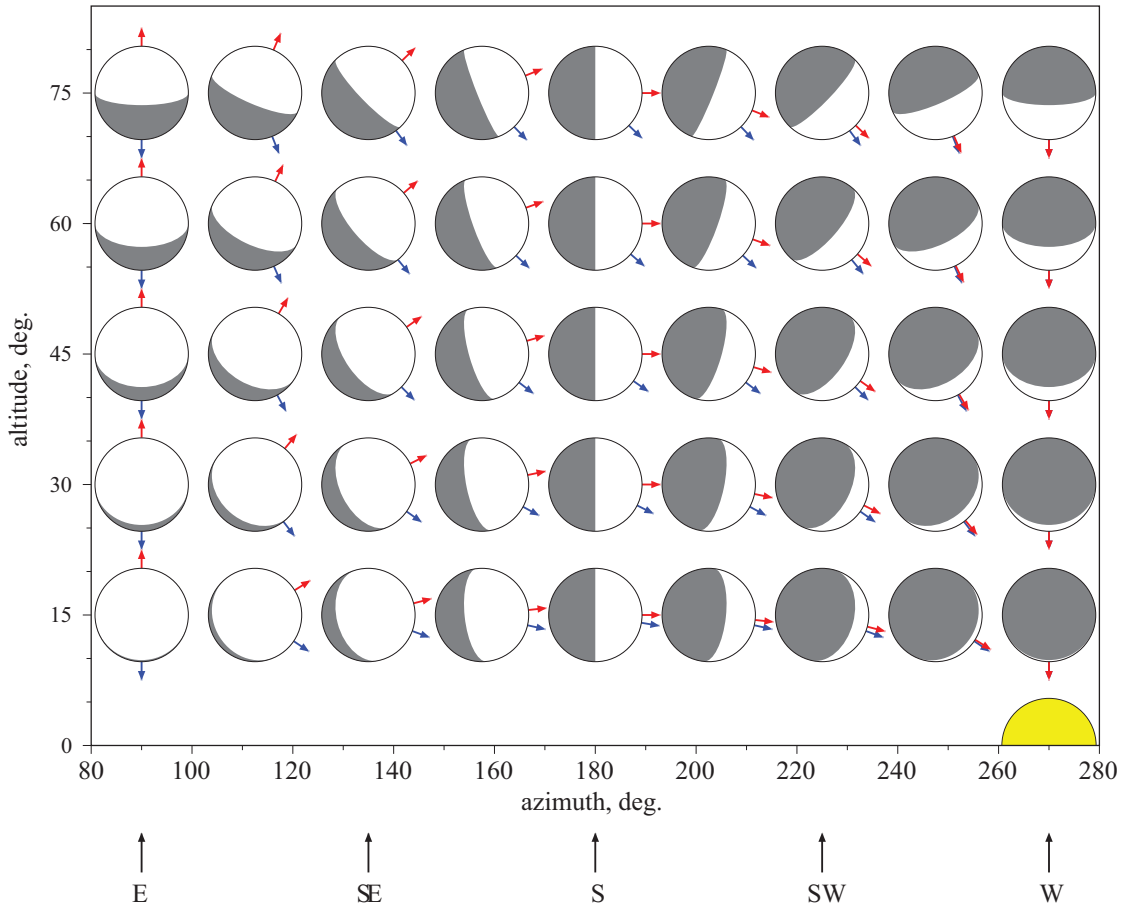


Figure 3: *Moon tilt illusion for waxing phases in northern hemisphere. Sun is setting due west. Red line is observed slope and blue line is expected slope of moon-sun line. Azimuth measured CW from north.*

We focus on cases where the sun and the moon are both visible in the sky, as this allows the observer to evaluate the positions of each. The moon is visible at twilight. Shown in Figure 3 is a chart for a waxing moon with a setting sun in the northern hemisphere. The magnitude of the moon tilt illusion is δ, the degree difference between the observed (red) arrow and the expected (blue) arrow. A set of four charts for waxing or waning moon in the northern and southern hemisphere could be constructed to cover all similar situations. Whether or not a particular configuration is visible depends on the latitude of the observer. For example in Figure 3 for a waxing moon, the horizontal "boat" crescent moon at high altitude in the west is observed near the equator but not in temperate zones. The chart is for the sun setting due west, which occurs at all latitudes during the spring and fall equinoxes. Since the moon tilt (α or β) depends on the *difference* of azimuths ($\Delta\phi$), corrections can be made for the sun setting at azimuths other than 270° by translating the entire set of images horizontally to the right or left.

The limits of δ, the magnitude of the illusion, are 0° for a new moon and 180° for a full moon. Near new moon, the δ angle is too small to be visible with the naked eye. For the crescent moon with under 90° azimuth difference between sun and moon, the magnitude of the illusion (δ) is small and the illusion is unimpressive, since the observed (red) and expected (blue) light directions are both below the horizontal. At half moon (sun-moon azimuth difference of 90°, moon at 180° on the chart), the discrepancy between the observed and expected directions becomes very noticeable since the observed light direction (red) is horizontal but the expected light direction(blue) is from below. For the gibbous moon at sunset or sunrise with azimuth difference greater than 90°, the illusion becomes striking since the moon is unambiguously lit from above the horizontal and the position of the sun is below the horizontal. The illusion is particularly impressive at sunset when the gibbous moon is at high altitude in the southwest or at sunrise when the gibbous moon is at high altitude in the southeast (both cases for the northern hemisphere). If illumination

exceeds about 90 percent, the direction of the red moon pointer may become difficult for the observer to discern.

In addition to the setting sun configuration in Figure 3, another interesting case occurs when the sun and moon are at the same (non-zero) altitude. Although the moon is lit from above the horizontal, the observer would expect to see the light travel horizontally from the sun to the moon. Our model gives $\beta = 0$ and $\delta = \alpha$.

In Figure 3, we note that for a particular elevation and setting sun at 270°, the expected beta is the same at moon azimuths of (180 + x) and (180 - x) degrees. For example, at 60° elevation, the angle of the blue sun-moon arrow is $-50.8°$ at moon azimuths of 135° and 225°. Moving from right to left at fixed altitude on Figure 3, the blue arrow moves CCW at first but switches to a CW movement after passing the 180° azimuth. Looking at Figure 3, instead of symmetry about 180°, one might expect the blue arrow indicating the direction of the sun to continue turning CCW when moving right to left at constant altitude. However, Figure 3 is a 2D representation of the 3D position of the sun relative to the moon. As the moon-sun azimuth difference increases beyond 90° from right to left at constant altitude, the sun moves *behind* the observer, causing the projected slope of the moon-sun vector to move in a CW direction. Facing the elevated moon and with the setting sun directly behind him, the observer would expect the light to illuminate the moon from below. The actual illumination is directly from above. Thus on the equator at sunset and particularly at high moon altitudes for which the moon is lit from above, observers experience a spectacular moon tilt illusion of magnitude 180°.

Eq. (12) for the calculation of the expected angle β depends upon the locations (azimuth and altitude) of the moon and sun in the sky. β is the angle of the sun-moon vector with the horizontal as projected upon a vertical plane perpendicular to the azimuth of the moon. Consider some of the limits for β which are independent of the projection plane used for its calculation and depend only on the geometry of the configuration. When the moon and sun have the same altitude, $\beta = 0$. When the moon and sun have the same azimuth with the moon above the sun, $\beta = -90°$. In the limit as the moon approaches the sun (new moon), $\beta = \alpha$ for any angle of approach. When the moon has a non-zero altitude and the moon-sun azimuth difference is 180°, the moon-sun vector strikes the moon from below so that $\beta = -90°$. Values of β from Eq. (12) conform to these limits.

Acknowledgement

The authors are grateful to Professor Benjamin Shen and Lecturer Mitchell Struble at the University of Pennsylvania, Professor Michael Berry at Bristol University, and Professor Zeev Vager at the Weizmann Institute of Science for valuable discussions and advice.

Notation

$\mathbf{m}$	moon vector OM
$\mathbf{n}$	normal vector to vertical plane
$\mathbf{s}$	sun vector OS
$\mathbf{x,y,z}$	Cartesian coordinate vectors
$\mathbf{v}$	moon-sun direction, Eq. (4)
$\mathbf{v_n}$	projection of $\mathbf{v}$ on $\mathbf{n}$
$\mathbf{v_p}$	projection of $\mathbf{v}$ on vertical plane
α	observed angle of moon pointer with horizontal
β	expected angle of moon pointer with horizontal
δ	difference of observed and expected angles of moon pointer with horizontal
η	altitude of moon or sun
θ	polar angle of moon or sun in spherical coordinates
ϕ	azimuth of moon or sun
χ	position angle of moon's bright limb
$\hat{}$	"hat" symbol for unit vector

References

[1] M. MINNAERT, *The Nature of Light and Colour in the Open Air*, Dover Publications Inc., 1954.

[2] J. LODRIGUSS, *A Different Moon Illusion*, `http://www.astropix.com/HTML/L_STORY/MOONILL.HTM`

[3] B. SCHÖLKOPF, The Moon Tilt Illusion, *Perception* **27** (1998), 1229–1232. `http://www.perceptionweb.com/abstract.cgi?id=p271229`

[4] K. SCHOTT, *"Falsche" Mondneigung - Wohin zeigt die Mondsichel?*, `http://falsche-mondneigung.jimdo.com/`

[5] G. GLAESER, K. SCHOTT, Geometric Considerations about Seemingly Wrong Tilt of Crescent Moon, *KoG* **13** (2009), 19–26,

[6] M. V. BERRY, Nature's optics and our understanding of light, *Contemp. Phys.*, in press, (2015).

[7] A. MYERS-BEAGHTON, *Tower Bridge Video*, `https://www.youtube.com/watch?v=Pt9iY-2XtDI`

[8] J. MEEUS, *Astronomical Algorithms*, 2nd Edition, William-Bell, Inc., Richmond, Virginia, USA, 1998, p. 346.

[9] D. B. TAYLOR, S. A. BELL, J. L. HILTON, *Computation of the Quantities Describing the Lunar Librations in The Astronomical Almanac*, Appendix C. `http://astro.ukho.gov.uk/data/tn/naotn74.pdf`

Andrea K. Myers-Beaghton
e-mail: a.beaghton@imperial.ac.uk

Department of Life Sciences, Imperial College London, U.K.

Alan L. Myers
e-mail: amyers@seas.upenn.edu

Chemical and Biomolecular Engineering, University of Pennsylvania,
Philadelphia, Pennsylvania, U.S.A.

Appendix A. β from vector components

Reduce the vector equation:

$$\tan\beta = \frac{|\mathbf{v_p}\times(\hat{\mathbf{n}}\times\hat{\mathbf{z}})|}{\mathbf{v_p}\cdot(\hat{\mathbf{n}}\times\hat{\mathbf{z}})}$$

to component form.

$$\mathbf{v_p} = \mathbf{v} - (\mathbf{v}\cdot\hat{\mathbf{n}})\hat{\mathbf{n}}$$

$$\mathbf{v_p}\times(\hat{\mathbf{n}}\times\hat{\mathbf{z}}) = (\mathbf{v_p}\cdot\hat{\mathbf{z}})\hat{\mathbf{n}} - (\mathbf{v_p}\cdot\hat{\mathbf{n}})\hat{\mathbf{z}}$$

But $\mathbf{v_p}$ is perpendicular to $\hat{\mathbf{n}}$ so:

$$\begin{aligned}\mathbf{v_p}\times(\hat{\mathbf{n}}\times\hat{\mathbf{z}}) &= (\mathbf{v_p}\cdot\hat{\mathbf{z}})\hat{\mathbf{n}}\\ &= (\mathbf{v}\cdot\hat{\mathbf{z}})\hat{\mathbf{n}} - (\mathbf{v}\cdot\hat{\mathbf{n}})(\hat{\mathbf{n}}\cdot\hat{\mathbf{z}})\hat{\mathbf{n}}\\ &= (\mathbf{v}\cdot\hat{\mathbf{z}})\hat{\mathbf{n}}\end{aligned}$$

because $\hat{\mathbf{n}}$ is perpendicular to $\hat{\mathbf{z}}$.

$$\mathbf{v} = \hat{\mathbf{s}} - \hat{\mathbf{m}} = (s_x - m_x)\hat{\mathbf{x}} + (s_y - m_y)\hat{\mathbf{y}} + (s_z - m_z)\hat{\mathbf{z}}$$

$$|\mathbf{v_p}\times(\hat{\mathbf{n}}\times\hat{\mathbf{z}})| = |\mathbf{v}\cdot\hat{\mathbf{z}}| = |v_z| = |s_z - m_z|$$

The denominator is the scalar triple product:

$$\begin{aligned}\mathbf{v_p}\cdot(\hat{\mathbf{n}}\times\hat{\mathbf{z}}) &= \mathbf{v}\cdot(\hat{\mathbf{n}}\times\hat{\mathbf{z}}) - (\hat{\mathbf{v}}\cdot\hat{\mathbf{n}})[\hat{\mathbf{n}}\cdot(\hat{\mathbf{n}}\times\hat{\mathbf{z}})]\\ &= \mathbf{v}\cdot(\hat{\mathbf{n}}\times\hat{\mathbf{z}})\end{aligned}$$

because $\hat{\mathbf{n}}\cdot(\hat{\mathbf{n}}\times\hat{\mathbf{z}}) = (\hat{\mathbf{n}}\times\hat{\mathbf{n}})\cdot\hat{\mathbf{z}} = 0$.

$$\mathbf{n}\times\hat{\mathbf{z}} = (m_x, m_y, 0)\times(0,0,1) = (m_y, -m_x, 0)$$

$$\begin{aligned}\mathbf{v}\cdot(\mathbf{n}\times\hat{\mathbf{z}}) &= [(s_x - m_x),(s_y - m_y),(s_z - m_z)]\cdot(m_y, -m_x, 0)\\ &= s_x m_y - s_y m_x\end{aligned}$$

In terms of the normalized vector $\hat{\mathbf{n}}$:

$$\mathbf{v}\cdot(\hat{\mathbf{n}}\times\hat{\mathbf{z}}) = \frac{s_x m_y - s_y m_x}{\sqrt{m_x^2 + m_y^2}}$$

Substituting results for the numerator and denominator of $\tan\beta$:

$$\tan\beta = \frac{|s_z - m_z|\sqrt{m_x^2 + m_y^2}}{s_x m_y - s_y m_x}$$ □

Appendix B. β from altitude and azimuth angles

Convert the component formulation for β to altitude and azimuth angles of the sun and the moon:

$$\tan\beta = \frac{|s_z - m_z|\sqrt{m_x^2 + m_y^2}}{s_x m_y - s_y m_x}$$

Cartesian coordinates of the observer-moon and observer-sun unit vectors are:

$$\begin{aligned}\hat{\mathbf{m}} &= m_x\hat{\mathbf{x}} + m_y\hat{\mathbf{y}} + m_z\hat{\mathbf{z}}\\ \hat{\mathbf{s}} &= s_x\hat{\mathbf{x}} + s_y\hat{\mathbf{y}} + s_z\hat{\mathbf{z}}\end{aligned}$$

In terms of altitudes and azimuths:

$$\begin{aligned}m_x &= \cos\eta_m\cos\phi_m;\quad m_y = \cos\eta_m\sin\phi_m;\quad m_z = \sin\eta_m\\ s_x &= \cos\eta_s\cos\phi_s;\quad s_y = \cos\eta_s\sin\phi_s;\quad s_z = \sin\eta_s\end{aligned}$$

so:

$$m_x^2 + m_y^2 = \cos^2\eta_m\cos^2\phi_m + \cos^2\eta_m\sin^2\phi_m = \cos^2\eta_m$$

$$\sqrt{m_x^2 + m_y^2} = \cos\eta_m$$

$$(s_z - m_z) = \sin\eta_s - \sin\eta_m$$

$$\begin{aligned}s_x m_y - s_y m_x &= \cos\eta_s\cos\phi_s\cos\eta_m\sin\phi_m\\ &\quad - \cos\eta_s\sin\phi_s\cos\eta_m\cos\phi_m\\ &= \cos\eta_s\cos\eta_m(\sin\phi_m\cos\phi_s - \cos\phi_m\sin\phi_s)\\ &= \cos\eta_s\cos\eta_m\sin(\Delta\phi)\end{aligned}$$

where $\Delta\phi = (\phi_m - \phi_s)$.

$$\begin{aligned}\tan\beta = -\frac{|s_z - m_z|\sqrt{m_x^2 + m_y^2}}{s_x m_y - s_y m_x} &= -\frac{\cos\eta_m\,|\sin\eta_s - \sin\eta_m|}{\cos\eta_s\cos\eta_m\sin(\Delta\phi)}\\ &= -\frac{|\sin\eta_m - \sin\eta_s|}{\cos\eta_s\sin(\Delta\phi)}\end{aligned}$$ □

How to get KoG?

The easiest way to get your copy of KoG is by contacting the editor's office:

Marija Šimić Horvath
msimic@arhitekt.hr
Faculty of Architecture
Kačićeva 26, 10 000 Zagreb, Croatia
Tel: (+385 1) 4639 176
Fax: (+385 1) 4639 465

The price of the issue is €15 + mailing expenses €5 for European countries and €10 for other parts of the world.

The amount is payable to:

ACCOUNT NAME:	Hrvatsko društvo za geometriju i grafiku Kačićeva 26, 10000 Zagreb, Croatia
IBAN:	HR862360000-1101517436

Kako nabaviti KoG?

KoG je najbolje nabaviti u uredništvu časopisa:

Marija Šimić Horvath
msimic@arhitekt.hr
Arhitektonski fakultet
Kačićeva 26, 10 000 Zagreb
Tel: (01) 4639 176
Fax: (01) 4639 465

Za Hrvatsku je cijena primjerka 100 KN + 10 KN za poštarinu.

Nakon uplate za:

HDGG (za KoG), Kačićeva 26, 10000 Zagreb
žiro račun broj **2360000-1101517436**

poslat ćemo časopis na Vašu adresu.

Ako Vas zanima tematika časopisa i rad našeg društva, preporučamo Vam da postanete članom HDGG-a (godišnja članarina iznosi 150 KN). Za članove društva časopis je besplatan.

ISSN 1331-1611
9 771331 161005