Res. Lett. Inf. Math. Sci., (2002) **3**, 25–35
Available online at http://www.massey.ac.nz/∼wwiims/research/letters/

Constructing an Ellipse in a Rosette Section

Bruce Mills
I.I.M.S., Massey University Albany Campus, Auckland, New Zealand
b.i.mills@massey.ac.nz

Abstract
Geometry is a craft, an amalgam of science and art. A good development should be pleasing to the mind, kin to poetry. When, after sedentation of a century, a desk was moved in the Laurentian library[1], it was discovered that the architect, Michaelangelo, had inscribed an ellipse in a rosette section. Clarifying the nature of this ellipse eluded naive automated algebra. Herein I present a development intended for human appreciation. It is an introduction to the geometry of the rosette diagram. A sketch of the structure of the space of inscribed ellipses. A beginning to a pragmatic description of the meaning and relation between some algebraic equations. A plan of attack on a geometric problem with historical connections. This paper expresses core groundwork I have been conducting in order to embark on the search for a more classical construction of the complete elliptically decorated rosette diagram.

1 Introduction

A rosette is a geometrical figure, used for centuries as a decoration. It is a fan of congruent circles, each passing through a central pinion, in which no circle is distinguished. It is apparent that a congruent circle centred on the pinion would pass through the centre of each circle, and be divided by the centres into equal parts. To construct a rosette, we begin with a circle evenly divided and draw a congruent circle centred at each division. There must be at least a trinity, or the aesthetic intent is frustrated. The number chosen is the order. A complete rosette of order 8 is shown in Figure 1. The whole is more evocative of an artichoke than a rose.

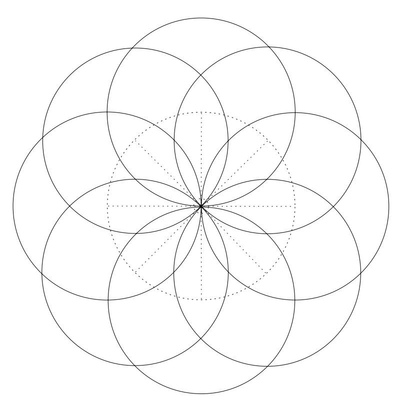

Figure 1: The order eight rosette

[1]http://www.kfki.hu/ arthp/html/m/michelan/5archite/early/3bibliot.html

1.1 The placement of the ellipse

A rosette diagram induces a natural tiling of a botryoid planar region, the tiles of which may be classified as types according to congruency. We see that each type of tile forms a single ring around the pinion, the tiles being joined together in the manner of paper cut-out dolls. An entire rosette tiling is a nesting of mono-type rings. The number of rings depends on the order of the rosette. There are always at least two. The innermost ring composed of needle tiles, and the outermost ring of eskimo ulu-knife blade tiles.

Between these two are any number of rings of quad-arc delimited tiles which I will call *darts* (See Figure 2 for a generic illustration). Each dart is embedded in a symmetric quadrilateral kite (known to be a rhombus). The radial line passing through two of the corners of the kite will be said to be the major axis of the dart, and the line orthogonal to this through the other two corner points, the minor axis. The major axis is also an axis of mirror symmetry of the dart (and the kite).

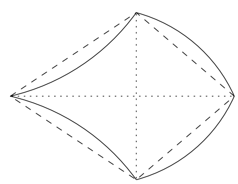

Figure 2: A generic kite tile

Our task is to add to this design by inscribing each dart tile with a maximal sized axis aligned ellipse (see Figure 3). (The problem of inscribing an ellipse in the kite has a known, though complicated, solution). The ellipse in question has its major axis aligned with the axis of symmetry of the dart, its interior is entirely contained in the interior of the dart and it is tangent to each of the four edges of the dart. The task of inscribing either the needle or blade tiles with a like ellipse is not entirely trivial, but is not the subject of this paper.

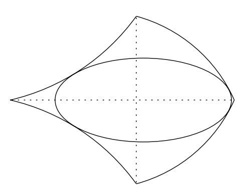

Figure 3: A maximal ellipse in a dart

In order to examine the problem we begin by considering the left diagram in Figure 4. This shows four circles extracted from the rosette. Sections of these circles form the four edges of the dart. The centre of the rosette is marked with a dot, and the horizontal radial line through this centre is also the axis of symmetry of the dart. The entire figure has a mirror symmetry about this axis. The four circles can be grouped as two pairs, one pair providing the two rear edges of the dart, and the other providing the two forward edges. The two rear circles are mirror images of each other across the axis of symmetry, as are the two forward circles.

Imagine that the major axis of the dart is a wire, and the boundary is a unyielding wall. A prolate ellipsoidal bead, pierced along its rotational axis, could be placed on the wire. If this ellipsoid should expand (keeping its shape), as it reached the wall at either end, it would be pushed

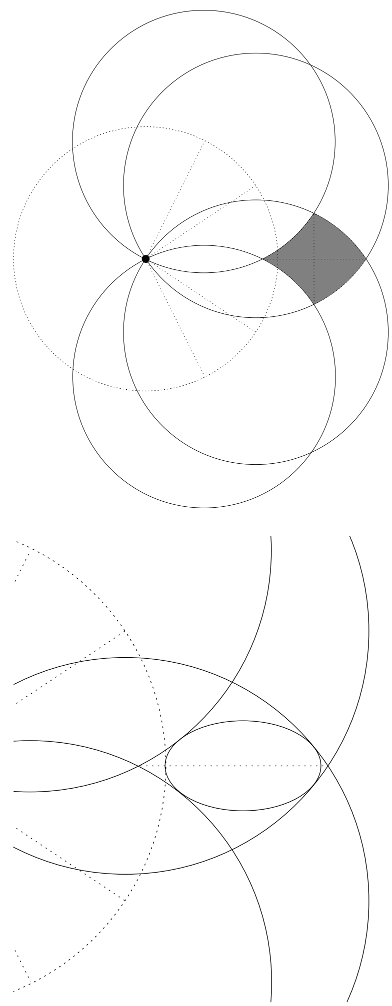

Figure 4: The dart in context, and ellipse close up

along the wire until it made contact at the other end. Thus it is made intuitively clear that there must be a maximal ellipse tangent to the dart (in the qualitative manner indicated in the right diagram of Figure 4) for each aspect ratio. Inspection of the shaded dart in Figure 4 will show that the intersection of the dart with its major axis is a degenerate case with an aspect ratio of zero. Likewise the minor axis, but with an infinite aspect ratio. It is clear that the question not simply of how to, but how many, must be answered.

1.2 An exact circular solution

A diagram of two congruent circles intersecting at two distinct points is also a diagram of two crescents touching at their end points (see Figure 6). In the general case, suppose that (c, r) are the centre and radius of the inscribed circle, and (c_i, r_i) the centre of the ith other circle $(i = 1, 2)$. A maximal circle inscribed in a crescent will be tangent to both other circles, laying wholly inside one and outside the other. Suppose it is outside the first other circle. When two circles are tangent, the line between the two centres passes through the point of tangency. Thus the distance between the two centres is either the sum or the difference of the radii. In particular $\|p - p_1\| = r_1 + r$ and $\|p - p_2\| = r_2 - r$. Thus the sum of these two distances is $r_1 + r_2$, the sum of the two radii, regardless of exactly where is the inscribed circle.

Thus, the centre of the inscribed circle is on an ellipse whose foci are the two centres, and the constant distance is the sum of the two radii (see Figure 5). Conversely, each point on this ellipse is the centre of a maximal inscribed circle (allowing a single point as a degenerate circle).

Drawing up the major and minor axial lines of this ellipse, we see that the major axis line is the line through the centres of the two circles, and the minor axis line must be the perpendicular bisector of the line segment determined by the two centres. The minor axis passes through the centre of the rosette. We have, therefore the two foci, the centre of the ellipse, and a minor radius drawn in, and of course we know the major axial line. The major axis length is obtained by finding the point mid way between the intersection of the major axis and the two circles. This is the point at which the ellipse crosses the major axis and numerically it is $(r_1 + r_2)/2$, the average of the two radii.

The maximal circle in the dart (again see Figure 5) is maximal in two crescents. The intersection between the two ellipses, generated as per the preceding discussion, give the centre point of the maximal circle. Symmetry, of the diagram about the major axis of the dart, dictates that this intersection will be on the major axis. Thus the point is the intersection of one ellipse with the major axis. To determine the radius of the maximal circle, we note that the line through the two centres passes through the tangency point. The minor axis length is the distance from the major axis to the intersection of the two circles.

We now have ample information to construct the ellipse by various processes. Simply knowing the the two foci an one point on the ellipse is enough for the mechanical two studs and a loop of string method. This is expressed physically as the two centres, and the intersection point of the circles.

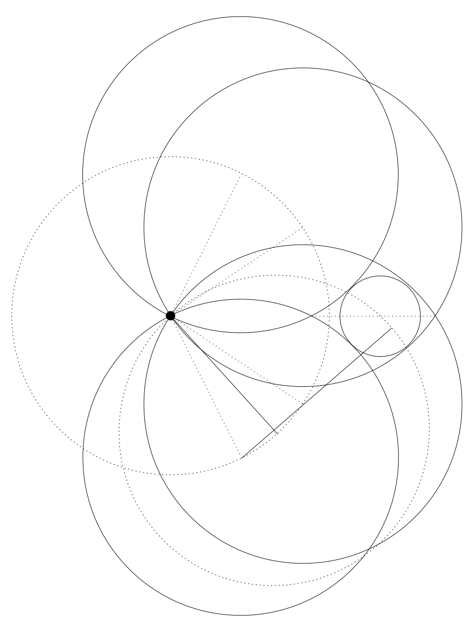

Figure 5: Construction of a circle in a dart

2 Some algebraic expressions

The dart itself is bordered by four circles. Due to the major axis symmetry any ellipse with is tangent to an unmatched pair of these must be tangent to the other two. Thus the task can be restated. Under the assumption of general position, given two congruent circles and an axis with a common intersection, find the collection of ellipses with that axis and tangent to both circles. Although the geometry of the diagram may vary, the topology is thus fixed (see Figure 6).

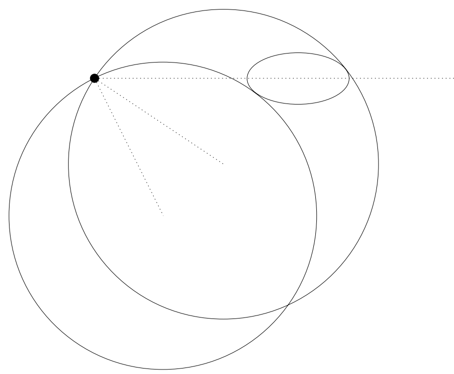

Figure 6: A simplified diagram

We establish a Cartesian frame using the major axis of the dart as the ordinate with the coordinate axis passing through the pinion. Let $(c, 0)$ be the centre of the inscribed ellipse. The centre points of the two circles are (α_1, β_1) and (α_2, β_2). The point of intersection of the ellipse and circle 1 is (x_1, y_1), and with circle 2 is (x_2, y_2). It is assumed that the circles are of radius 1. The lengths of the two extreme radii of the ellipse are taken as $1/\sqrt{a}$ and $1/\sqrt{b}$. The ellipse is the locus of solutions of $a(x-c)^2 + by^2 = 1$. Several algebraic identities between these quantities are apparent (see Figure 7).

1.	Point 1 is on the ellipse	so	$a(x_1 - x)^2 + by_1^2 = 1$
2.	Point 1 is on circle one	so	$(x_1 - \alpha_1)^2 + (y_1 - \beta_1)^2 = 1$
3.	Point 2 is on the ellipse	so	$a(x_2 - c)^2 + by_2^2 = 1$
4.	Point 2 is on circle two	so	$(x_2 - \alpha_2)^2 + (y_2 - \beta_2)^2 = 1$
5.	Tangent at point 1	so	$a(x_1 - c)(y_1 - \beta_1) - by_1(x_1 - \alpha_1) = 0$
6.	Tangent at point 2	so	$a(x_2 - c)(y_2 - \beta_2) - by_2(x_2 - \alpha_2) = 0$

Figure 7: The meaning of the equations

Each equation has a logical geometric interpretation. We have four equations expressing the idea that each of the tangency points are on the ellipse, and one of the circles. The tangency constraints may be determined by implicit differentiation of the ellipse and circle equations and requiring that the resulting differential vectors be parallel. However, the resulting equation is has a broader geometric significance.

In determining equations to model logical conditions we have to be careful to make sure that they are equivalent. It is more illuminating to derive the tangency equations by the method of Lagrange multipliers. We note that the tangency point is also an extremum point for distance from the centre of a circle. Thus we optimise $(x-\alpha)^2+(y-\beta)^2$ subject to the constraint that $a(x-c)^2+by^2=1$, producing exactly the same equation.

This makes a total of 6 independent equations in 7 unknowns. We can expect that there is a one dimensional continuum of solutions. Intuition, and a few trial constructions of ellipses with a given centre will convince that this continuum of solutions does exist. Note should be taken that the given algebraic equations indicate only that the ellipse is symmetrically placed, and tangent to all four circles. There are more ways to do this than the ellipse being in the interior of the dart (see Figure 8).

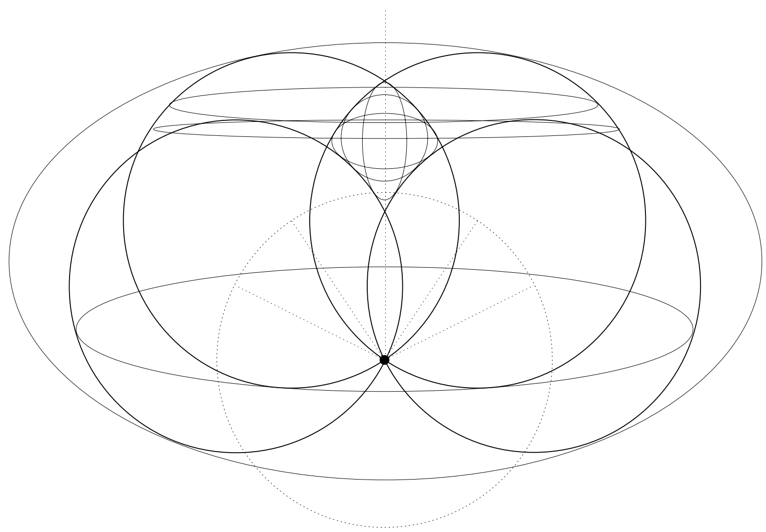

Figure 8: Several types of tangent ellipse

Each of the ellipses determines a solution to the six basic equations. It is unclear how the requirement that the ellipse be inside the dart can be expressed neatly with a low order equation. At this point the task enters the realm of *semi* algebraic geometry, which has far more geometric significance to humans, but is much more difficult to work with.

The ways in which the ellipse can be placed may be equated with the path connected components of the solution locus in the a, b, c space. It is desirable to determine a parameterisation of each component by a single real parameter. In this way any number of such ellipses could be generated. However, such a parameterisation using simple expressions might not be possible at all, and appears to be difficult if possible.

For the moment we consider in more detail the problem of determining two equations that relate a, b and c, without any reference to x_1, y_1, x_2 or y_2.

3 Some remarks on strategy

The sextuplet of equations is naturally divided into two independent triplets, according as they refer to one tangent point or the other. The generic form of each triplet of equations is:

$$\begin{aligned} &\text{e:} \quad a(x-c)^2+by^2=1 \\ &\text{c:} \quad (x-\alpha)^2+(y-\beta)^2=1 \\ &\text{t:} \quad a(x-c)(y-\beta)-by(x-\alpha)=0 \end{aligned}$$

which expresses the problem of finding an ellipse with a given axis and tangent to a given circle cut by that axis. For a non degenerate ellipse, equation e implies

$$y^2 = \frac{1 - a(x-c)^2}{b} \tag{1}$$

and equation t implies

$$y = \frac{a\beta(x-c)}{(a-b)(x-c) + b(\alpha - c)} \tag{2}$$

finally equation c tells us that

$$(x-\alpha)^2 + y^2 - 2\beta y + \beta^2 = 1 \tag{3}$$

Substituting for y from (2) and y^2 from (1) into (3) produces the cubic equation:

$$((a-b)x - ac + b\alpha)(1 - a(x-c)^2 - 2\alpha bx + bx^2) = 2\beta^2 ab(x-c) \tag{4}$$

For any specific ellipse inscribed in the dart, the value of (a, b, c) determines a specific cubic whose roots include the horizontal coordinate of the point of contact with the circle. It would be nice to just pick a value of (a, b, c) and find a value for x. But, there is no obvious interpretation of the value of x obtained when (a, b, c) does not represent an inscribed ellipse. Alternatively, since there will be no real valued (x, y) that solves the three equations when (a, b, c) does not represent an inscribed ellipse, we might consider looking at the conditions under which the cubic has real solutions. But, there will always be at least one real root, it is just that either the y value computed via the ellipse, or the circle equation, or both, will be complex if the a, b, c does not represent an inscribed ellipse.

I conjecture, however, that the cubic has a repeated real root when (a, b, c) represents an inscribed ellipse. If this is true, then the approach described in Section 4 would yield one polynomial algebraic equality between a, b and c, which, for each specified c, would describe exactly all the ellipses with the designated centre that are tangent to the circle. For a fixed topology this would be some closed (but perhaps not simple) loop in the plane. The intersection points between two of these loops would give an ellipse tangent to two circles, thus providing a fairly simple relation, independent of x_i and y_i, for a, b and c.

Alternatively, equating the square of the RHS of (2) with the RHS of (1) we obtain:

$$a^2 b\beta^2 (x-c)^2 = (1 - a(x-c)^2)(a(x-c) - b(x-\alpha))^2$$

For a specific choice of the two circles, this is a quartic in x, in which the coefficients are functions of (a, b, c). Thus, for example, once a valid value of these three ellipse parameters are known, then x can be determined. Geometrically, this is apparent, since under the given conditions the ellipse will be tangent to each circle at no more than a single point. In particular, this means that for valid ellipse parameters, the above quartic must have exactly one real root. However, the complex roots of a real coefficient quartic come in complex conjugate pairs, so if there is exactly one real solution for y, then this must be a repeated root. The condition that the quartic has a repeated real root means that the GCD of the quartic and its derivative must be either a linear or quadratic term. By proceeding through Euclid's algorithm for the GCD, using the above quadratic, and its cubic derivative, we find an algebraic condition on the ellipse parameters which is required for Euclid's algorithm to return a linear or quadratic term. This condition also includes the α and β and we have two circles that the ellipse must be correctly related to. Thus, there are actually two equations in the three variables, a, b and c. This defines a one dimensional solution space in the ellipse space for the two specified circles.

3.1 Expressed more generally

If we have two polynomial equations: $y = P(x)$ and $Q(x,y) = 0$, then by substituting, we get $Q(x, P(x)) = 0$. If we solve this equation, then we have a pair (x, y) that is a solution to the original equation set. But, if we have three rational polynomial equations: $y = P_1(x)$, $y^2 = P_2(x)$ and $Q(x, y, y^2) = 0$, (which is our situation) then a solution to the equation $Q(x, P_1(x), P_2(x))$ might not lead to a solution to the original equation set, we still need to check that $P_1(x)^2 = P_2(x)$. If this is so, then we have a solution. Thus, strictly, the substitution has produced two equations in x, the solution of which leads to the solution of the original three equations in x and y.

In our particular case we have:

$$P_1(x) = \frac{a(x-c)\beta}{(a-b)(x-c)+b(\alpha-c)}$$

$$P_2(x) = \frac{1-a(x-c)^2}{b}$$

$$Q(x, y, y^2) = (x-c)^2 + y^2 - 2\beta y + \beta^2 - 1$$

The equation $Q(x, P_1(x), P_2(x)) = 0$ is the cubic equation that we derived above. The equation $P_1(x)^2 = P_2(x)$ is actually the equation:

$$\left(\frac{a(x-c)\beta}{(a-b)(x-c)+b(\alpha-c)}\right)^2 = \frac{1-a(x-c)^2}{b}$$

whose solution set is identical to that of the quartic

$$a^2 b\beta^2 (x-c)^2 = (1-a(x-c)^2)((a-b)(x-c)+b(\alpha-c))^2$$

Thus, all solutions to this quartic give solutions to the original two equations, and y is expressed as a rational polynomial of x with real coefficients, so if x is real, then y is real. From geometric intuition we may argue that there might be none, one, two, three or four suitable real values of x. However since the roots move continuously with the ellipse, as two intersection points merge into one the root must be repeated. Thus, in order to constrain the equation to correspond to an ellipse with one intersection point, and that being tangential, we may equate the quartic with the form $(x^2 - 2px + p^2 + q^2)(x^2 - 2rx + r^2)$, from which polynomial equations in a,b and c may be derived, not involving the coordinates of the tangential points.

If we find (a, b, c) such that a root of the cubic is equal to a root of the quartic, then it must be a solution to the original equation set. There can only ever be one such real value for any given (a, b, c). Further we know that the set of values for which it does occur is thin in a, b, c space. In order for the cubic and quartic to have a common solution they have to have a non trivial common divisor, and we may be able to determine the condition for this, in terms of a, b, c, by using Euclid's Algorithm.

3.2 Trigonometric approach

From the tangency condition we may also produce the equation that

$$\frac{a}{b} = \frac{(x-c)(y-\beta)}{y(x-\alpha)}$$

Since the ellipse is the same in each case, we see that

$$\frac{(x_1 - c)(y_1 - \beta_1)}{y_1(x_1 - \alpha_1)} = \frac{x_2(y_2 - \beta_2)}{(y_2 - c)(x_2 - \alpha_2)}$$

Which can also be expressed as,

$$\frac{x_1}{(y_1 - c)} \frac{(y_1 - \beta_1)}{(x_1 - \alpha_1)} = \frac{x_2}{(y_2 - c)} \frac{(y_2 - \beta_2)}{(x_2 - \alpha_2)}$$

Each side of the equation is a ratio of tangents of geometrically significant angles. The first is the angle of the point from the centre of the ellipse, and the second is the angle of the point from the centre of the circle. By parameterising the ellipse using an angle, and the cos/sin trig formulas, we could obtain a trigonometric expression, whose solution would yield dividends. However, this approach will not be pursued further in this paper.

4 The effect of repeated roots

If the conjecture of the section 3 is correct, and the cubic has repeated roots, then the following approach should generate the required relation between a, b and c.

Suppose that $x^3 - Ax^2 + Bx - C$ is a cubic with one repeated root. The A, B and C are intended as polynomials in a, b, c. Thus, all its roots are real, but further it is of the form $(x - p)(x - q)^2 = x^3 - (2q + p)x^2 + (q^2 + 2pq)x - pq^2$. From which we get a basic set of equations:

$$\begin{aligned} &2q + p = A \\ &q^2 + 2pq = B \\ &pq^2 = C \end{aligned}$$

Now, it happens that we can isolate p in each equation:

$$\begin{aligned} &p = A - 2q \\ &p = (B - q^2)/2q \\ &p = C/q^2 \end{aligned}$$

The complete information contained in these equations is preserved in the following:

$$A - 2q = (B - q^2)/2q = C/q^2$$

Re-expressing this as the three required equalities we get

$$\begin{aligned} &A - 2q = (B - q^2)/2q \\ &A - 2q = C/q^2 \\ &(B - q^2)/2q = C/q^2 \end{aligned} \qquad \text{which is} \qquad \begin{aligned} &2qA - 4q^2 = B - q^2 \\ &2qA - 4q^3 = 2C \\ &Bq - q^3 = 2C \end{aligned}$$

Which is also

$$5q^2 - 2Aq - B = 0$$
$$4q^3 - 2Aq + 2C = 0$$
$$q^3 - Bq + 2C = 0$$

This is an homogeneous simultaneous equation set, and the three polynomials involved must have a root in common. We can proceed by symbolically determining the greatest common factor of the quadratic with each of the cubics. If the cubics are not multiple of the quadratic then they must have precisely a linear term in common. Proceeding symbolically on this assumption obtains two linear terms that must be equivalent. Thus we have a polynomial in A, B and C which must have a zero value for a valid ellipse.

5 The construction loop

Each ellipse can be referred to an aligned rectangle to which it is tangential. For the family of ellipses that we are talking about we can plot the corners of these rectangles. The locus of all these points is a one dimensional loop, which is moderately close to the ellipse plotted through the centre of the dart, and passing through the four corners. However, if this where the real shape then the centre of the ellipse would be constant. In fact near the corners the loop will be much like the circumscribed ellipse, but between the corners at the front it will bulge somewhat, causing the centre point of the inscribed ellipse to move forward of the centre of the dart. The rear edge of the loop either is inside the circumscribed ellipse, or at least does not bulge so much. A rough image of this is shown in Figure 9.

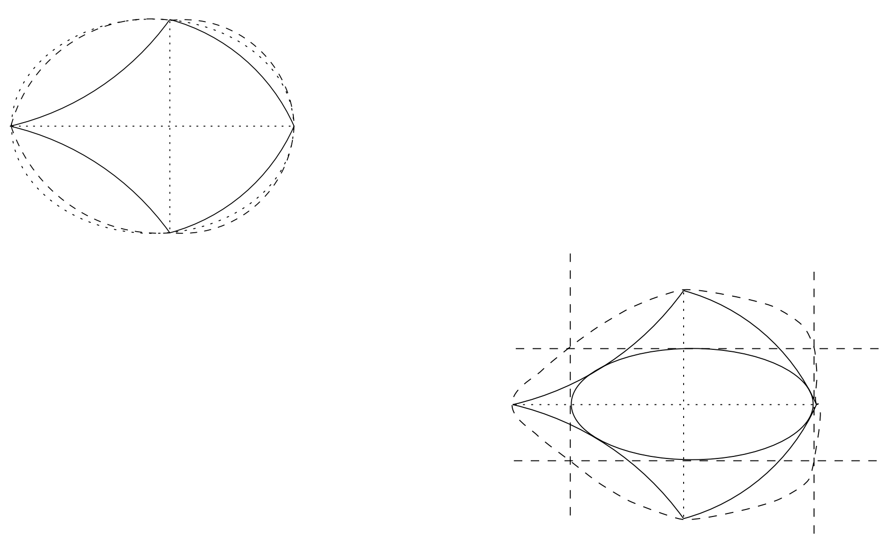

Figure 9: The construction loop and its use

If we had this loop plotted around the dart we could construct an ellipse by choosing minor or major radius, and drawing in a pair of lines perpendicular or parallel, respectively, to the main axis, to obtain the bounding rectangle for the ellipse. This construction is illustrated in an approximate fashion in Figure 9. The relationship required to define this loop is one between $c + a$ and b.

6 Future Research

The exact meaning of the tangency condition is actually quite complex, it expresses the notion that at the point of intersection the direction of the ellipse and the circle are the same. If the minimum curvature of the ellipse is less than that of the circle, then it is possible to place it so that it crosses the ellipse, rather than is tangent to it. More generally, for an ellipse in a circle, there are potentially four critical points of the distance from the centre (on the ellipse), the maximum, the minimum, and two points of inflection. The points of inflection cannot be the same as the tangency point. If we modify the ellipse to attempt this it drives the tangency point toward the pole or the equator. If we modify the ellipse so that its minimum curvature is greater than the circles curvature, then the inflection points move so that they have a common x coordinate value. Any further increase in the maximum curvature of the ellipse results in the launching of two complex valued solutions to the corresponding quartic. Investigation of the topology of the space of these diagrams should produce some constraints on the nature of the quartic. It is in this direction that the author intends to focus.

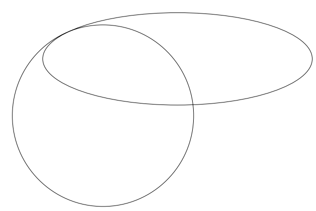

Figure 10: Ellipse crossing a circle

It might also be interesting to consider what the shape of the construction loop is for the case of a kite. What sort of shape has an ellipse for this loop? You can find the loop from the shape, can you find the shape from the loop? When is it the envelope of all the rectangles circumscribed on inscribed ellipses? If shape to loop is not a one to one function, what sort of variation can you have? For what sort of shape might the construction loop come inside the shape. Due to the tangential nature of the ellipse, the corners of the rectangle would typically be outside the shape. So the loop would be essentially outside the shape. Are there any shapes at all for which the loop comes inside? There obviously are, for example we could stick out a thin spire, that projects through the otherwise unmodified loop. But, can it occur for a convex shape? The shape most likely has to be very non convex, and perhaps even have multiple solutions for maximal inscribed axially aligned ellipses of a given aspect ratio.

In all of this, a distinct question still remains. This discussion was inspired by a diagram of a pattern by Michelangelo, in a library floor panel. The pattern was a rosette in which each dart was inscribed with an ellipse. So, for reasons of historical and mathematical interest, how did *they* inscribe the ellipses? Was it trial and error, this could be done, or maybe an approximation, several suggest themselves. Or perhaps like the circle, there is a particular ratio ellipse for which there is a simple construction. The construction given for the circle is not very practical, due to the distance from the centre of the ellipse to the intersection point, and the extreme delicacy of the determination of the centre of the circle, it is easier in practice just to fit the circle by guess work. Is there a more robust method for some ellipse? Admittedly, the ellipses in the diagram are clearly not a tight fit, so it was mostly likely an empirical process. One could wander for ages through this problem in all sorts of guises. It incorporates questions from ancient Greece, through Renascence Italy, painting and architecture, mysticism, and all the way up to modern computer algebra techniques. This all goes to make it a lively relevant, and thoroughly human study which brings together in an unforced way, many aspects of what it is to be human.