

# Morikawa’s Unsolved Problem

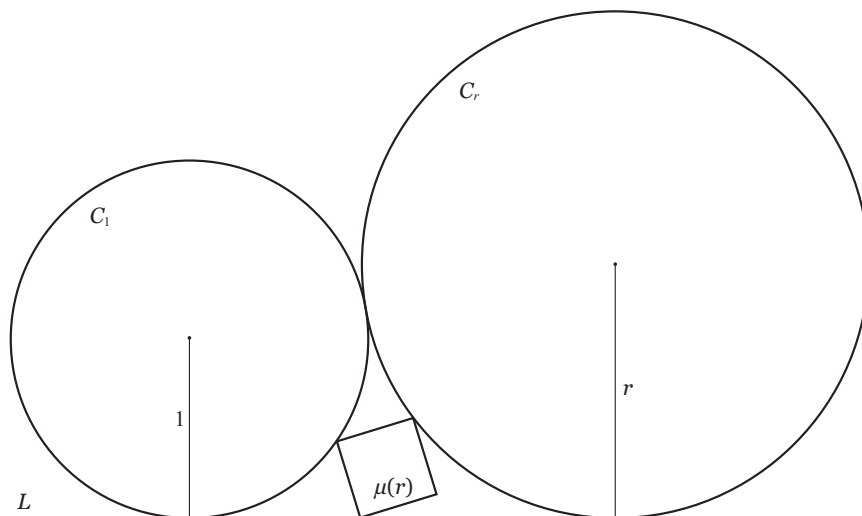
Jan E. Holly and David Krumm

**Abstract.** By combining theoretical and computational techniques from geometry, calculus, group theory, and Galois theory, we prove the nonexistence of a closed-form algebraic solution to a Japanese geometry problem first stated in the early nineteenth century. This resolves an outstanding problem from the *sangaku* tablets which were at one time displayed in temples and shrines throughout Japan.

**1. INTRODUCTION.** During the Edo Period of Japanese history (1603–1867) there developed a curious practice of hanging wooden tablets with mathematical content from the eaves of Buddhist temples and Shinto shrines. Many of these tablets, known as *sangaku*, have been lost to history, but close to 900 of them have been preserved [8]. The problems inscribed on the surviving *sangaku* are mostly of a geometric nature, and they range in difficulty from trivial to unsolved. Solutions to many of these problems can be found in the books by Fukagawa and Pedoe [4] and Fukagawa and Rothman [5]; the latter reference also discusses various historical aspects surrounding the mathematics of the Edo Period.

Unsolved *sangaku* problems seem to be rare; in fact, we are aware of only two such problems listed in the literature, both in Fukagawa and Rothman [5, Chapter 7]. One of the problems mentioned in that text was originally proposed in 1821 and has recently been solved [6]. The present article concerns the other unsolved problem, which was proposed by Jihei Morikawa during the same time period.

The main objects involved in Morikawa’s problem are illustrated in Figure 1. Given a line  $L$  and circles  $C_1$  and  $C_r$  of radii 1 and  $r \geq 1$ , respectively, such that  $C_1$  and  $C_r$  are tangent to each other and to  $L$ , the problem asks us to express, in terms of  $r$ , the minimum side length  $\mu(r)$  of a square that can be inscribed in the region between  $C_1$ ,  $C_r$ , and  $L$ . Here, “inscribed” means touching all three of  $C_1$ ,  $C_r$ , and  $L$ .



**Figure 1.** The circles  $C_1$  and  $C_r$ , the line  $L$ , and the minimal inscribed square with side length  $\mu(r)$ . The original statement of the problem involves circles of radii  $a$  and  $b$  with  $b \geq a$ , but a scaling of the plane reduces this general case to the case  $a = 1$ .

A surviving travel diary of mathematician Kanzan Yamaguchi, a contemporary of Morikawa, includes an entry with some additional information about Morikawa’s problem and the *sangaku* containing it. Based on that entry, Fukagawa and Rothman [5, p. 265] report the following.

The tablet contained no solution, but Morikawa had written, “I will be very happy if someone can solve this problem.” And so, says Yamaguchi, “I went to Morikawa’s home with my friend Takeda and asked him what the answer is. He said that he could not solve the problem yet.” Neither does Yamaguchi’s diary contain a solution and, like Morikawa, we would be very happy if someone solves this problem. [5, p. 265]

It seems surprising that Morikawa’s problem would have frustrated all attempts at a solution, considering that the mathematicians of the Edo Period had a strong understanding of geometry, were adept in the use of algebra, and even had some knowledge of basic calculus. One begins to wonder whether the problem can in fact be solved.

The purpose of this article is to address the question of the existence of a closed-form expression for  $\mu(r)$ . From our analysis in Section 4 it follows that  $\mu(r)$  is a root of a polynomial whose coefficients are polynomials in  $r$ ; in light of this fact, a natural question is whether  $\mu(r)$  is expressible by radicals in terms of  $r$ . Precise terminology is defined below, but the question can be stated intuitively as follows: Is there a radical expression, such as

$$\frac{\sqrt[11]{3r - \pi} \cdot \sqrt[3]{r^5 - r + 2i} + \sqrt[4]{1 + i} + \sqrt[5]{r^6 - 7}}{3 - \sqrt{2r^3 + 9r - 5}},$$

that for every real number  $r \geq 1$  can be evaluated to yield  $\mu(r)$ ?

We provide here a negative answer to this question, thus showing that a closed-form algebraic solution does not exist in the classical sense. In order to state our results we introduce the following terminology. Recall that if  $J \subseteq \mathbb{C}$  is a nonempty set and  $f : J \rightarrow \mathbb{C}$  is a function, we say that  $f$  is an *algebraic function* if there exists a nonzero polynomial  $q \in \mathbb{C}[k, x]$  such that  $q(c, f(c)) = 0$  for every  $c \in J$ . If, moreover, this condition is satisfied by a polynomial  $q$  whose Galois group over the field  $\mathbb{C}(k)$  is solvable — or equivalently, whose splitting field is contained in a radical extension of  $\mathbb{C}(k)$  — then we say that  $f$  is a *radical function*.

We can now state our main result.

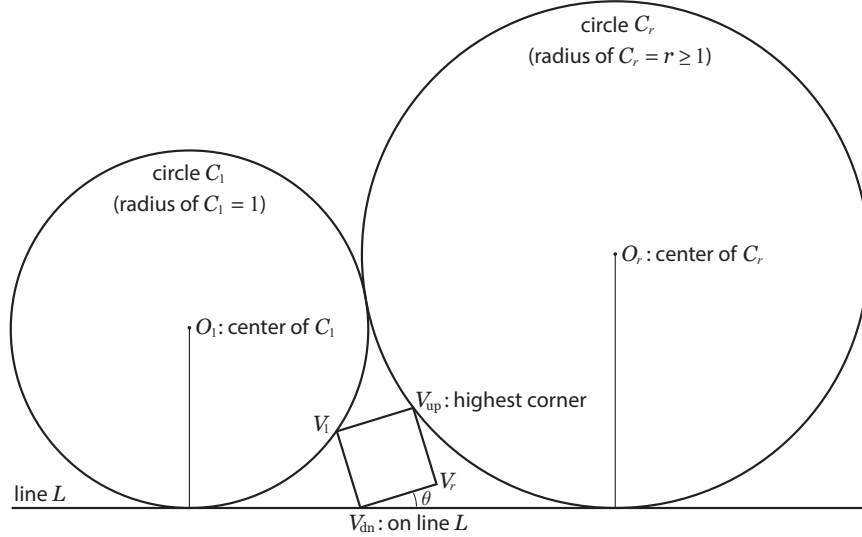
**Theorem (see Theorem 5.4).** *The function  $\mu : [1, \infty) \rightarrow \mathbb{R}$  is not radical. In fact, there is no infinite subset  $J \subseteq [1, \infty)$  such that  $\mu : J \rightarrow \mathbb{R}$  is radical.*

The proof of this theorem makes critical use of computational tools in Galois theory that have only recently become available due to work of N. Sutherland [9]. In particular, our argument relies on Sutherland’s implementation in the system MAGMA [1] of an algorithm for computing geometric Galois groups. Besides this algorithm, the proof uses elementary geometry as well as calculus and Galois theory.

This article is organized as follows. In Section 2 we provide notation and basic results about inscribed squares. In Section 3 we show that any minimal inscribed square must be positioned as in Figure 1, with a corner on each of  $C_1$ ,  $C_r$ , and  $L$ , and with no side of the square tangent to these objects. In Section 4 we derive an explicit formula for a function whose minimum value is  $\mu(r)$ ; as a byproduct we obtain a numerical method for approximating  $\mu(r)$  given the radius  $r$ . In addition, we show that  $\mu(r)$  can be expressed in terms of a root of a certain polynomial of degree 10. Finally, in Section 5 we use Galois theory to study this polynomial and thus prove the main theorem.

**2. CONFIGURATIONS FOR INSCRIBED SQUARES.** Figure 2 shows the notation that will be used throughout the article, regardless of which inscribed square is under discussion. Denoted are the circles ( $C_1, C_r$ ), centers of the circles ( $O_1, O_r$ ), line ( $L$ ), vertices of the square ( $V_1, V_r, V_{\text{up}}, V_{\text{dn}}$ ), and angle between  $L$  and the lower right side of the square ( $\theta \in [0, \pi/2)$ ). If  $\theta = 0$ , then  $V_{\text{dn}}$  is the lower left vertex. To facilitate phrasing, the line is considered horizontal as shown, with  $C_1$  on the left.

Most of the discussion in Sections 2 and 3 assumes an arbitrary but fixed value of  $r \geq 1$ . Basic geometric facts [2] are used throughout.



**Figure 2.** Notation for the line, two circles, and square.

**Lemma 2.1.** *For inscribed squares (for fixed  $r \geq 1$ ), let  $\theta \in [0, \pi/2)$  be as in Figure 2.*

- (i) *For every  $\theta \in [0, \pi/2)$ , there is a unique inscribed square.*
- (ii) *There exists a minimum side length over the set of inscribed squares. (Thus, Morikawa's problem is well-defined.)*

*Note: Throughout the rest of the article, the side length of the inscribed square at angle  $\theta \in [0, \pi/2)$  will be denoted  $s(\theta)$ .*

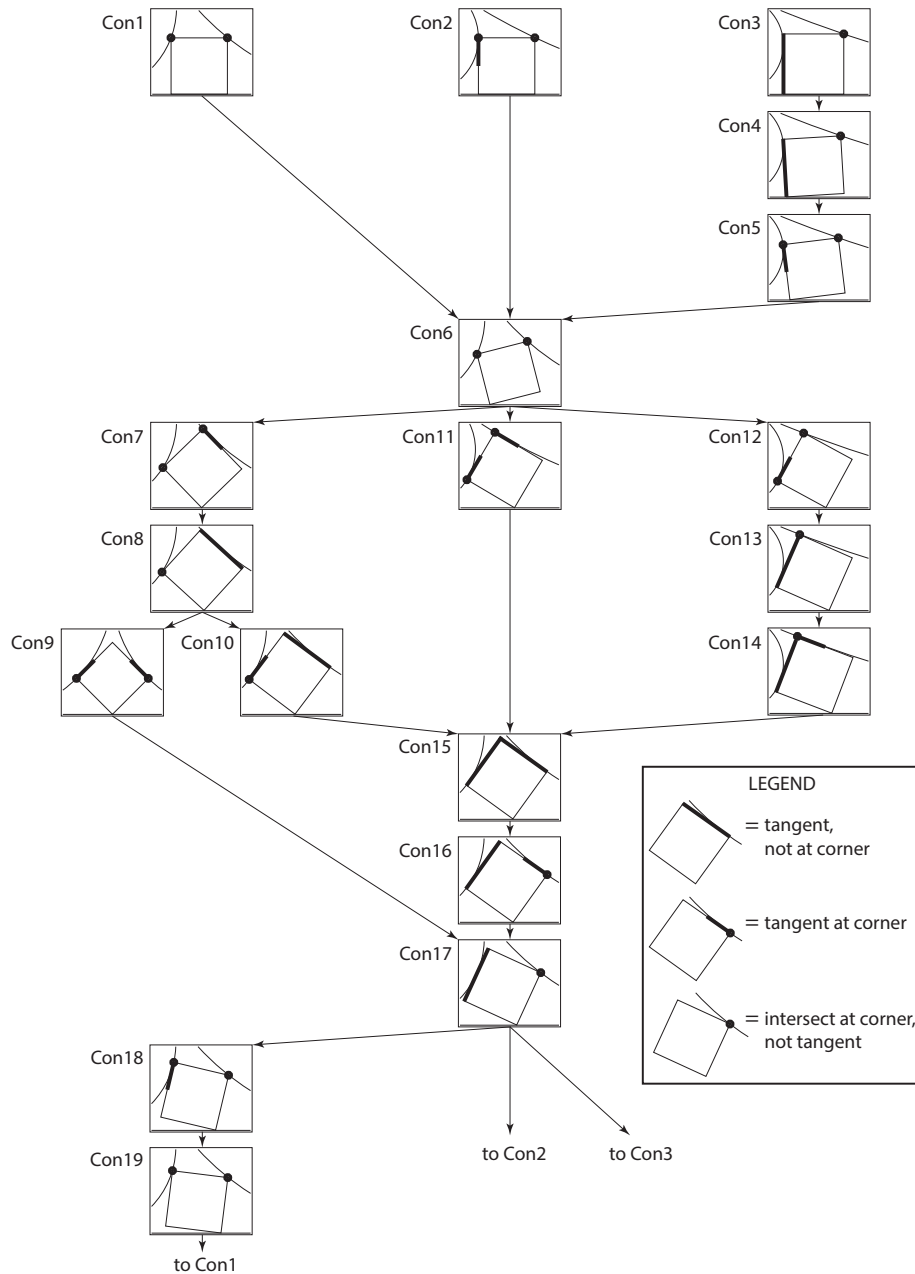
*Proof.* To prove (i), fix  $\theta \in [0, \pi/2)$ . We find the inscribed square at angle  $\theta$  as follows, stated somewhat informally to avoid excessive technicalities.

Consider all squares, inscribed or not, that make angle  $\theta$  with  $L$  as in Figure 2. Let  $s$  be a side length under consideration for being that of an inscribed square. Imagine sliding such a square — with angle  $\theta$  and side length  $s$  — along  $L$  until the square is to the right of  $C_1$  but is just touching  $C_1$ . If  $s$  is too small, then the square will not reach  $C_r$ . If  $s$  is too large, then the square will overlap  $C_r$ . By a continuity and monotonicity argument, there is a unique  $s$  such that the  $C_1$ -touching square will exactly reach  $C_r$ .

The existence and uniqueness of an inscribed square follows from that of  $s$  above, along with the fact that the inscribed square clearly cannot be moved left or right and still be inscribed.

Subsequently, (ii) follows from the facts that  $s(\theta)$  is continuous on  $[0, \pi/2)$  and  $\lim_{\theta \rightarrow \pi/2^-} s(\theta) = s(0)$ . ■

Generally in this article, “the square” will mean the inscribed square as given by the context, unless stated otherwise. Figure 3 shows the types of intersections between the square and circles as  $\theta$  increases from 0 to  $\pi/2$ , i.e., as the square rotates (and changes size as necessary). Each combination of such intersections, as shown in Figure 3, will henceforth be referred to as a *configuration*. The configurations as illustrated in Figure 3 will be denoted Con1, Con2, Con3, etc.



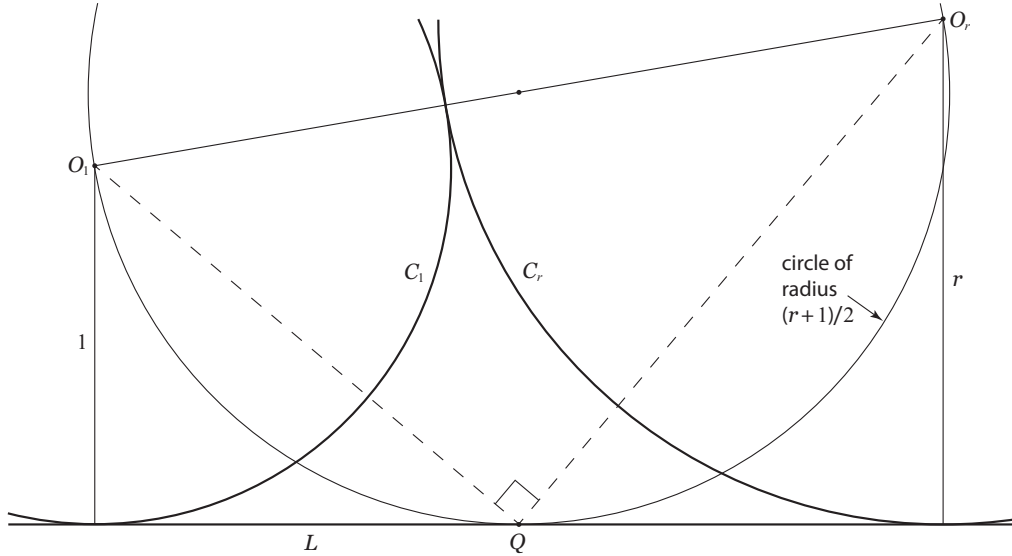
**Figure 3.** Possible configurations, in terms of the square’s types of intersections with the line and circles. In Section 3 we prove that a minimal square can occur only in Con6, or in Con19 with  $r = 1$ , by eliminating all of the others: Con1,2,3 (Lemma 3.2), Con4,5 (Lemma 3.3), Con7,10,11 (Lemma 3.4), Con12,13,14 (Lemma 3.5), Con15 (Lemma 3.6), Con16,18 (Lemma 3.7), Con19 (Lemma 3.8), Con8,9,17 (Lemma 3.16).

**Lemma 2.2.** *Figure 3 shows all possible steps through the configurations as  $\theta$  increases from 0 to  $\pi/2$ .*

*Proof.* The steps through the configurations obviously depend upon  $r$ . For example, from Con6, the value of  $r$  determines which circle first becomes tangent to a side of the square as  $\theta$  increases. Small  $r$  leads to Con7, large  $r$  leads to Con12, and a certain intermediate value of  $r$  leads to Con11. Also note that only  $r = 1$  gives Con9.

The fact that these are the steps through the configurations is generally clear, with two exceptions: from Con8 to Con9 and Con10, and from Con15 to Con16.

For Con8 to Con9 and Con10, the question is whether the upper right side of the square could instead rotate past tangency with  $C_r$  before the upper left side of the square becomes tangent to  $C_1$ . This can happen only if the lower left side of the square has steeper slope than the line through  $V_{dn}$  and  $O_1$  — i.e., the lower left side “points above”  $O_1$  — and the lower right side points above  $O_r$ . However, this is impossible because the circle of radius  $(r + 1)/2$  through  $O_1$  and  $O_r$ , as shown in Figure 4, is tangent to  $L$ . Every angle inscribed in a semicircle is a right angle, so since  $L$  is below the new circle except at the point of tangency,  $Q$ , the lower sides of the square cannot both point above their respective circles’ centers. At best, the lower sides can point exactly at the centers, but only if  $r = 1$  and  $V_{dn} = Q$ .



**Figure 4.** Circle with diameter  $O_1$ -to- $O_r$  is tangent to  $L$ .

For Con15 to Con16, the question is whether  $V_{up}$  can instead touch  $C_1$  before the lower right side of the square touches  $C_r$ . This can happen only if the square with upper right side tangent at  $V_r$  to  $C_r$  has  $V_{up}$  touching  $C_1$ . However, this is impossible. Any square with upper right side tangent at  $V_r$  to  $C_r$  must have  $V_{dn}$  at or to the left of  $Q$  in Figure 4 in order for the square to reach  $C_1$ , but then the square is angled such that  $V_{up}$  cannot touch  $C_1$ . ■

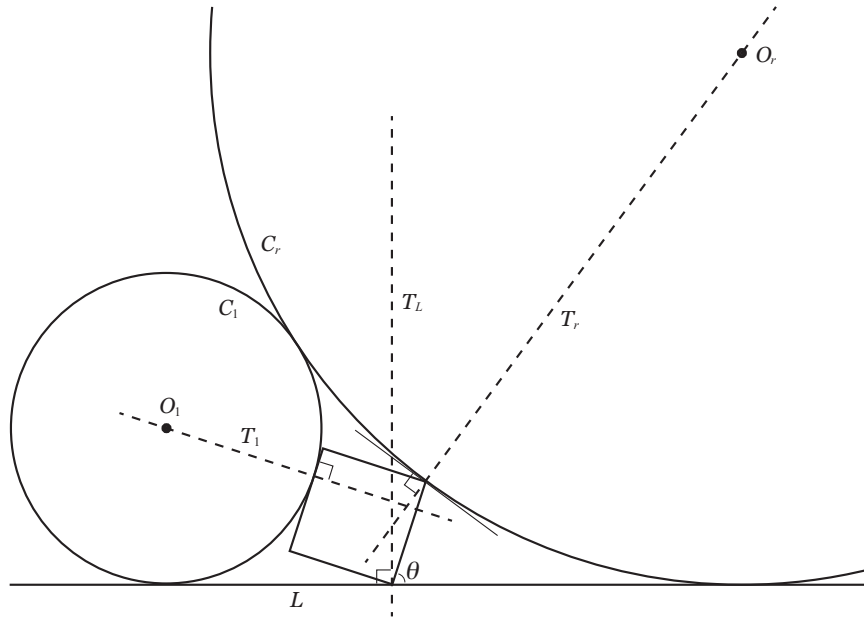
**3. CONFIGURATIONS FOR MINIMAL SQUARES.** In this section we prove that a minimal square — i.e., an inscribed square with side length that is minimal over *all* orientations — exists only in Con6, and has  $V_1$  lower than  $O_1$ . An exception occurs if  $r = 1$ , where Con19 is the reflection of Con6 and thus also has a minimal square. The

proof consists of a sequence of lemmas showing that a minimal square cannot be in any other configuration. The final result is given by Proposition 3.17.

Additional notation is used throughout this section, for the lines perpendicular to each of  $C_1$ ,  $C_r$ , and  $L$  at the points of contact with the square under consideration. As illustrated in Figure 5, these (dashed) lines are denoted  $T_1$ ,  $T_r$ , and  $T_L$ , respectively. As before, unless stated otherwise we assume an arbitrary but fixed value of  $r \geq 1$ .

**Lemma 3.1.** *For a given inscribed square, if  $T_1$  intersects  $T_L$  above  $T_r$  (respectively, below  $T_r$ ), then  $s$  is a strictly decreasing (respectively, increasing) function of  $\theta$  at that square's angle.*

*Proof.* If  $T_1$  intersects  $T_L$  above  $T_r$ , then the lines form a triangle to the right of  $T_L$ , such as in Figure 5.



**Figure 5.** Notation: Lines  $T_1$ ,  $T_r$ , and  $T_L$  through points of intersection.

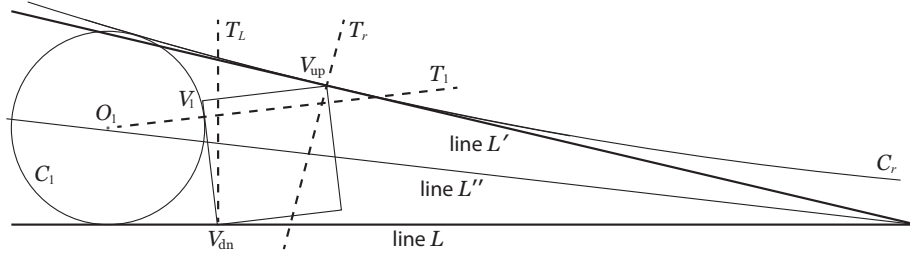
Consider fixing the size of the square, and rotating it counterclockwise about a point inside the triangle. As the square begins to rotate, it starts to overlap with each of  $C_1$ ,  $C_r$ , and  $L$ . Formally, the rate of change of the following are positive: (1) the radius of  $C_1$  minus the distance from  $O_1$  to the square, (2) the radius of  $C_r$  minus the distance from  $O_r$  to the square, and (3) the distance below  $L$  to the lowest point on the square. This means that the difference between the side lengths of an inscribed square and a fixed-size square is a strictly decreasing function of  $\theta$ ; thus  $s$  is a strictly decreasing function of  $\theta$ . Similarly, if  $T_1$  intersects  $T_L$  below  $T_r$ , then  $s$  is a strictly increasing function of  $\theta$ . ■

**Lemma 3.2.** *A minimal square cannot be in Con1, Con2, or Con3.*

*Proof.* Each of these configurations corresponds to  $\theta = 0$ . For small enough  $\epsilon > 0$ , it is easy to see that for all  $\theta \in (0, \epsilon)$ ,  $T_1$  intersects  $T_L$  above  $T_r$ . Therefore by Lemma 3.1,  $s(\theta)$  is strictly decreasing for  $\theta \in (0, \epsilon)$  and thus by continuity  $s(0)$  is not minimal. ■

**Lemma 3.3.** *A minimal square cannot be in Con4 or Con5, and a minimal square cannot be in Con6 with  $V_1$  as high or higher than  $O_1$ .*

*Proof.* Figure 6 illustrates Con4, and the same reasoning applies to the other two configurations. First, note that  $V_1$  is higher than  $O_1$  in Con4 and Con5 because some part of the left side of the square is tangent to  $C_1$ , and that left side has negative slope. Consider the line  $L'$  tangent to  $C_r$  at  $V_{up}$ , and the line  $L''$  bisecting the angle between  $L$  and  $L'$ . Line  $L'$  must intersect  $C_1$  (in order to “escape” the region between  $C_1$ ,  $C_r$ , and  $L$ ), so  $L''$  is below  $O_1$ .

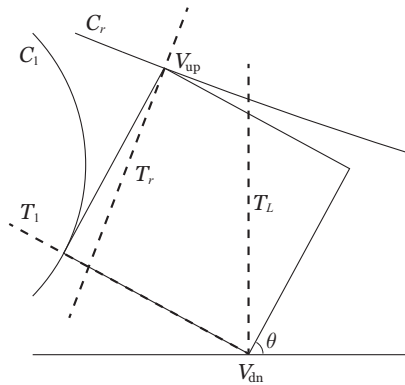


**Figure 6.** Illustration for Lemma 3.3. The line  $L''$  bisecting the angle between  $L$  and  $L'$  is below  $O_1$ , so  $V_1$  is higher than  $L''$ , leading to an application of Lemma 3.1.

Because  $V_1$  is higher than  $O_1$  and therefore higher than  $L''$ , the square is tilted in such a way that  $V_{up}$  is closer than  $V_{dn}$  to the intersection of  $L'$  and  $L$ . Therefore,  $T_r$  intersects  $L''$  to the right of where  $T_L$  intersects  $L''$ ; thus the intersection of  $T_L$  and  $T_r$  is below the intersection of  $T_L$  and  $L''$ . Meanwhile,  $T_1$  has nonnegative slope and therefore intersects  $T_L$  above the intersection of  $T_L$  and  $L''$ , thus above the intersection of  $T_L$  and  $T_r$ . Hence by Lemma 3.1,  $s(\theta)$  is a strictly decreasing function and is therefore not minimal for these values of  $\theta$ . ■

**Lemma 3.4.** *A minimal square cannot be in Con7, Con10, or Con11.*

*Proof.* In each of these configurations,  $T_1$  intersects  $T_L$  below the intersection between  $T_L$  and  $T_r$ . Therefore by Lemma 3.1,  $s(\theta)$  is strictly increasing and therefore not minimal at the corresponding  $\theta$ . ■



**Figure 7.** Illustration for Lemma 3.5. In Con12, Con13, and Con14,  $\theta > \pi/4$ , so  $T_L$  is to the right of  $V_{up}$ , leading to an application of Lemma 3.1.

The proof that a minimal square cannot be in Con8, Con9, or Con17 is substantially more complicated than the other proofs, and is saved for the end of the section.

**Lemma 3.5.** *A minimal square cannot be in Con12, Con13, or Con14.*

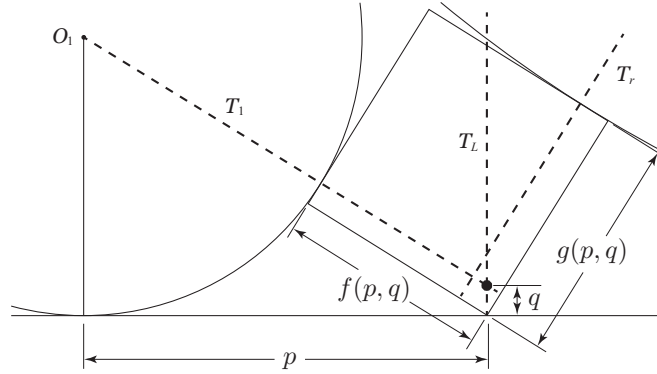
*Proof.* In each of these configurations, some part of the upper left side of the square is tangent to  $C_1$ . Therefore, in order for  $V_{\text{up}}$  to reach  $C_r$ ,  $\theta$  cannot be particularly small, and certainly  $\theta > \pi/4$ : As seen in Con9 (Figure 3) which has  $\theta = \pi/4$  and  $r = 1$ , a square with  $\theta = \pi/4$  and upper left side tangent to  $C_1$  can only just barely reach — and not even with  $V_{\text{up}}$  — the smallest possible  $C_r$ , that with  $r = 1$ . Then decreasing  $\theta$  with the square's upper left side still tangent to  $C_1$  would shrink and move the square away from  $C_r$ , so  $\theta \leq \pi/4$  is not possible here when  $r = 1$ . It is also not possible when  $r > 1$  because any larger  $C_r$  would further prevent the square from reaching it.

Therefore,  $T_L$  is to the right of  $V_{\text{up}}$ , as illustrated in Figure 7 for the case of Con12. Thus by Lemma 3.1,  $s(\theta)$  is a strictly increasing function of  $\theta$  and is therefore not minimal for these values of  $\theta$ . ■

**Lemma 3.6.** *A minimal square cannot be in Con15.*

*Proof.* We prove this by showing that as the square rotates counterclockwise through the range of Con15, the intersection between  $T_1$  and  $T_L$  moves upward strictly monotonically, and the intersection between  $T_r$  and  $T_L$  moves downward strictly monotonically. Because it is clear in Con15 that  $T_L$ 's intersection with  $T_1$  is below that with  $T_r$  for the smaller values of  $\theta$ , and above that with  $T_r$  for the larger values of  $\theta$ , Lemma 3.1 implies that  $s(\theta)$  strictly increases then strictly decreases as a function of  $\theta$ , so the result follows. (This seems to imply that there is a maximal square in Con15. However, this article does not address maximal squares.)

We give the proof that the intersection between  $T_1$  and  $T_L$  moves upward strictly monotonically; an analogous proof shows that the intersection between  $T_r$  and  $T_L$  moves downward strictly monotonically.



**Figure 8.** For Lemma 3.6, the rectangle determined by the point  $(p, q)$ .

Consider not just inscribed squares, but inscribed rectangles in general, determined by the pair  $(p, q)$  as shown in Figure 8, where  $p$  is the horizontal distance between  $O_1$  and the bottom vertex, and  $q$  is the height above  $L$  of the intersection between  $T_1$  and  $T_L$ . Let

$$f(p, q) = \text{length of lower left side of rectangle,}$$

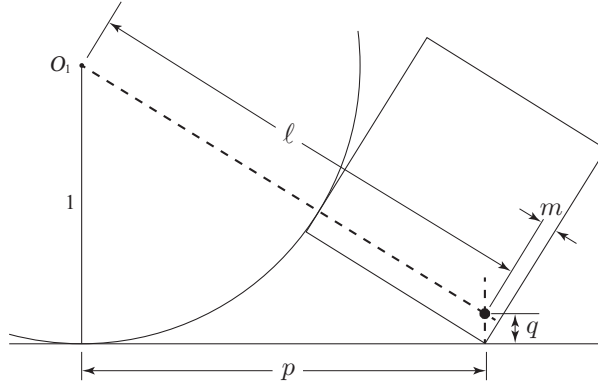
$$g(p, q) = \text{length of lower right side of rectangle.}$$

For each  $p$  there is a unique  $q$ , say  $q = h(p)$ , such that the rectangle is a square, i.e.,  $f(p, q) = g(p, q)$ . Our goal is to show that  $h(p)$  is a strictly increasing function of  $p$ . We will do this by showing that



- (1)  $f(p, q)$  is a strictly increasing function of  $p$ ,
- (2)  $g(p, q)$  is a strictly decreasing function of  $p$ ,
- (3)  $f(p, q)$  is a strictly decreasing function of  $q$ ,
- (4)  $g(p, q)$  is a strictly increasing function of  $q$ .

This will complete the proof because if point  $(p, q)$  gives a square, then (1) and (2) imply that increasing  $p$  will cause the lower left side to become larger than the lower right side, so by (3) and (4),  $q$  must be increased in order to restore equality of the side lengths.

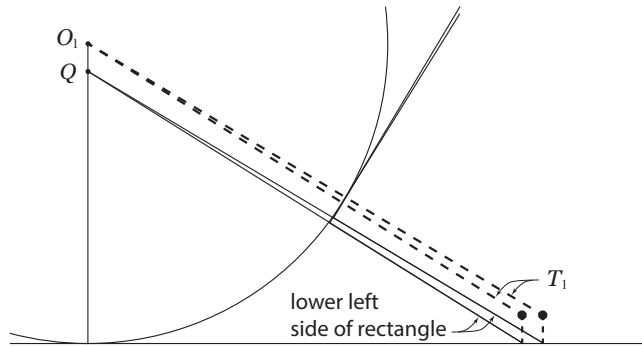


**Figure 9.** For Lemma 3.6, notation for the proof that  $f(p, q)$  is a strictly increasing function of  $p$ .

For (1), fix  $q$  and note that  $f(p, q)$  is  $\ell + m - 1$  in Figure 9. By similar triangles,  $m/q = (1 - q)/\ell$ . Define a function

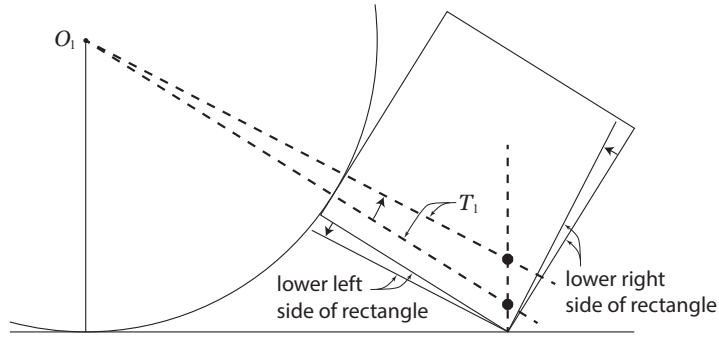
$$j(\ell) = \ell + m - 1 = \ell + \frac{q(1 - q)}{\ell} - 1,$$

which equals  $f(p, q)$ . An easy calculation shows that  $j'(\ell) > 0$  because  $\ell > 1/2$ , and since  $p$  and  $\ell$  increase together, this gives  $\frac{\partial f}{\partial p} > 0$ , thus proving (1).



**Figure 10.** For Lemma 3.6, geometry for the proof that  $g(p, q)$  is a strictly decreasing function of  $p$ .

For (2), note that  $g(p, q)$  is exactly  $r$  less than the distance from  $O_r$  to the lower left side of the rectangle. That side is on a line through a point  $Q$  below  $O_1$ . If  $p$  increases while  $q$  stays fixed, then as shown in Figure 10, the lower left side of the rectangle is



**Figure 11.** For Lemma 3.6, geometry for proofs that  $f(p, q)$  is strictly decreasing, and  $g(p, q)$  is strictly increasing, as functions of  $q$ .

still aligned with  $Q$  because that side is always parallel to  $T_1$ . Therefore, as  $p$  increases, the distance from  $O_r$  to the lower left side of the rectangle decreases, thus proving (2).

For (3) and (4), note that  $f(p, q)$  is exactly 1 less than the distance from  $O_1$  to the lower right side of the rectangle, and  $g(p, q)$  is exactly  $r$  less than the distance from  $O_r$  to the lower left side of the rectangle. If  $p$  is fixed and  $q$  increases, then as shown in Figure 11, the lower left side of the square (which is parallel to  $T_1$ ) rotates away from  $O_r$ , and the lower right side (which is perpendicular to  $T_1$ ) rotates toward  $O_1$ . This proves (3) and (4), thus completing the proof. ■

**Lemma 3.7.** *A minimal square cannot be in Con16 or Con18.*

*Proof.* In each of these configurations,  $T_1$  intersects  $T_L$  above the intersection between  $T_L$  and  $T_r$ . Therefore by Lemma 3.1,  $s(\theta)$  is strictly decreasing and therefore not minimal at the corresponding  $\theta$ . ■

If  $r = 1$ , then Con19 is simply the reflection of Con6, so for Con19 we focus on the case  $r \neq 1$ .

**Lemma 3.8.** *If  $r \neq 1$ , then a minimal square cannot be in Con19.*

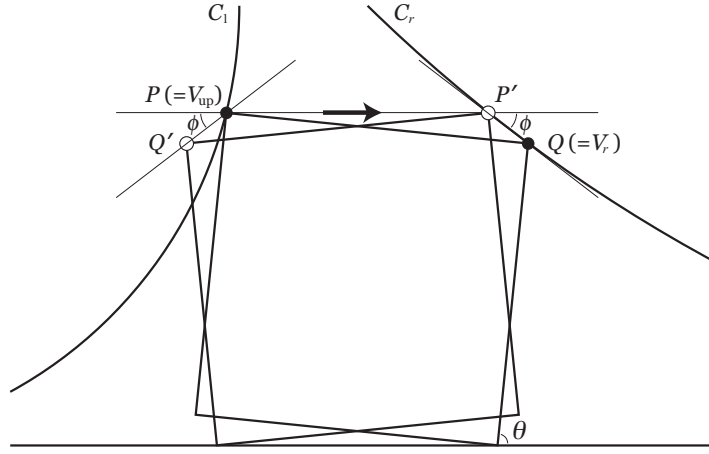
*Proof.* Given a square at angle  $\theta$  in Con19, we claim that the inscribed square in the reflected orientation, i.e., at angle  $\pi/2 - \theta$ , is smaller. This fact can be seen by mapping the original square to a horizontal reflection that is positioned so that the image  $P'$  of vertex  $P (= V_1)$  is on  $C_r$ , as shown in Figure 12. In the process, vertex  $Q (= V_r)$  is mapped to a point  $Q'$  inside the disk bounded by  $C_1$  for the following reason. The acute angles  $\phi$  made with the horizontal are the same for the line through  $P$  and  $Q'$  and the line through  $P'$  and  $Q$ , while circle  $C_1$  is steeper than  $C_r$  at every height above  $L$  between 0 and 1.

Therefore, the inscribed square at angle  $\pi/2 - \theta$  must be smaller than the reflected square and thus the original square. ■

To prove that a minimal square cannot be in Con8, Con9, or Con17, preliminary results and definitions are helpful.

**Definition 3.9.**  $Con10^+$  will refer to Con10 but will also allow  $r \leq 1$ .  $Con8^+$  will refer to the union of Con7, Con8, Con9, and Con10, and will also allow  $r \leq 1$ .

**Remark 3.10.** A proof that a minimal square cannot be in  $Con8^+$  constitutes a proof for Con8, Con9, and Con17 (as well as Con7, Con10, Con16, and Con18, for which proofs have already been given) because Con17 with  $r = r_0 \geq 1$  is equivalent to a version of Con8 with  $r = 1/r_0 \leq 1$ .



**Figure 12.** Illustration for Lemma 3.8 showing the horizontal reflection of the square that sends point  $P$  on  $C_1$  to a point  $P'$  on  $C_r$ . (Such a reflection is not necessarily about the midline of the square.)

**Lemma 3.11.** *The square in  $\text{Con}10^+$  has side length less than or equal to  $M$ , where*

$$M = \frac{2r}{r + \sqrt{8\sqrt{r} + 1}},$$

*with equality if and only if  $r = 1$ .*

*Proof.* Let  $P$  be the point on line  $L$  between  $C_1$  and  $C_r$  such that the distance from  $P$  to the closest point on  $C_1$  (along the line through  $P$  and  $O_1$ ) equals the distance from  $P$  to the closest point on  $C_r$ . The line segment from  $P$  to  $O_1$  can be viewed as the hypotenuse of a right triangle, as can the line segment from  $P$  to  $O_r$ , so the equality of distance can be written using the Pythagorean theorem as follows. Let  $p$  be the distance from  $P$  to the intersection of  $L$  and  $C_1$ . Because  $2\sqrt{r}$  is the distance from the intersection of  $L$  and  $C_1$  to the intersection of  $L$  and  $C_r$  (see Figure 15), we have

$$\sqrt{1 + p^2} - 1 = \sqrt{r^2 + (2\sqrt{r} - p)^2} - r.$$

By basic geometry, the square in  $\text{Con}10^+$  must have  $V_{\text{dn}}$  to the left of  $P$ , unless  $r = 1$ , in which case  $V_{\text{dn}} = P$ . Therefore the square has side length less than or equal to  $\sqrt{1 + p^2} - 1$ , with equality exactly if and only if  $r = 1$ .

It is straightforward to confirm, by substituting the following into the equation above, and by noting that this is the desired solution because it lies in  $(0, 2\sqrt{r})$  for  $r > 0$ , that

$$p = \frac{\sqrt{8r} + 2\sqrt{r}}{r + \sqrt{8\sqrt{r} + 1}},$$

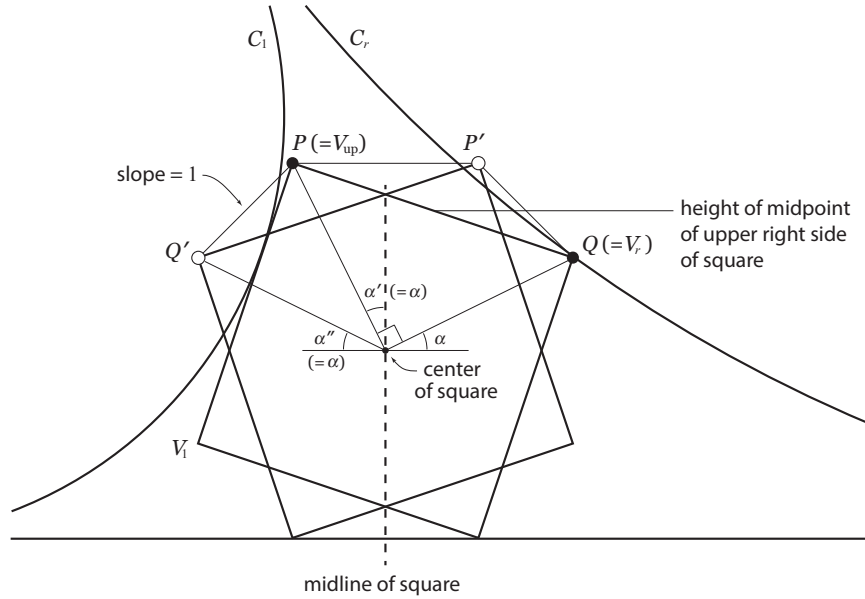
and then that

$$\sqrt{1 + p^2} - 1 = \frac{2r}{r + \sqrt{8\sqrt{r} + 1}},$$

which is  $M$  in the statement of the lemma. ■

**Lemma 3.12.** For a square in Con17, if the distance from  $L$  to the midpoint of the upper right side of the square is less than or equal to  $r - r/\sqrt{2}$ , then the square is not minimal.

*Proof.* Given such a square in Con17 as shown in Figure 13, we claim that reflecting the square about its midline — the vertical line through its center point — will cause it to overlap both  $C_1$  and  $C_r$ . This will prove the lemma because an inscribed square in the same orientation as the reflected square is thus smaller than the original square. We prove the claim by showing that reflecting the square takes vertex  $P (= V_{up})$  as shown in Figure 13 to vertex  $P'$  in the interior of the disk with boundary  $C_r$ , and takes  $Q (= V_r)$  to vertex  $Q'$  such that either  $Q'$  or part of the left side of the reflected square is in the interior of the disk with boundary  $C_1$ .



**Figure 13.** For Lemma 3.12, the horizontal reflection of the square about its midline.

First, the line through  $P$  and  $Q'$  has slope 1, and the line through  $Q$  and  $P'$  has slope  $-1$ , by the following reasoning: In Figure 13, the fact that  $\alpha' = \alpha$  is clear because of the right angle shown, and  $\alpha'' = \alpha$  because of the symmetry of the reflection. The line segment between  $P$  and  $Q'$  is seen to be the base of an isosceles triangle since  $P$  and  $Q'$  are the same distance from the center of the square. Because  $\alpha' = \alpha''$ , the isosceles triangle is symmetric about a line of slope  $-1$  through the center of the square. Therefore, the base of the triangle, and thus the line through  $P$  and  $Q'$ , has slope 1. Similar reasoning shows that the line through  $Q$  and  $P'$  has slope  $-1$ .

Next, we claim that the secant line segment intersecting  $C_r$  at  $Q$  and the point nearest horizontally to  $P'$  has slope shallower than  $-1$ . (This secant would lie almost exactly along  $C_r$  in Figure 13, and would not be distinguishable in the figure.) To prove the claim, we note that the midpoint of that secant is at the same height as the upper right side of the square, no higher than  $r - r/\sqrt{2}$  by hypothesis, because the secant's endpoints are at the heights of vertices  $P$  and  $Q$ . A secant has as a perpendicular bisector a radial line segment of  $C_r$ ; such a radial line segment for this secant must reach below a height of  $r - r/\sqrt{2}$  since it passes through the secant's midpoint. Thus, this radial line segment must have slope steeper than 1, because at slope 1 it would

reach down only to height  $r - r/\sqrt{2}$  with its length  $r$  and top endpoint at height  $r$ . The secant, which is perpendicular to it, must therefore have slope shallower than  $-1$ .

Therefore,  $P'$  is in the interior of the disk with boundary  $C_r$ .

To show the overlap with  $C_1$ , we first note that  $Q'$  is at height greater than  $1 - 1/\sqrt{2}$ . This is because  $V_r$  of a square in Con17 must be higher than that in Con9 or Con10 (whichever applies, depending on whether  $r = 1$ ) by the basic geometry of the counterclockwise rotation toward Con17. Among these three configurations, Con9 with its  $r = 1$  has the lowest  $V_r$ , at height  $1 - 1/\sqrt{2}$  since  $\theta = \pi/4$ .

Now,  $Q'$  lies on the line of slope 1 through  $P$ , a line whose lower intersection with  $C_1$  is at height less than  $1 - 1/\sqrt{2}$  because  $1 - 1/\sqrt{2}$  is where the tangent to  $C_1$  has slope 1. Therefore,  $Q'$  is on the line of slope 1 through  $P$  on either the segment from  $P$  to  $C_1$  or the segment that lies in the interior of the disk with boundary  $C_1$ . In the latter case, the proof of the lemma is complete. The former case can only occur if the original square is so large that  $Q'$  is above a portion of the top half of  $C_1$ , because  $Q'$  is straight above  $V_1$ . In that case, the left side of the reflected square must intersect the interior of the disk with boundary  $C_1$ , completing the proof of the lemma, because the reflected square is also a reflection of the original square about a *horizontal* line through its center point. Since the left side of the original square is tangent to  $C_1$ , and the horizontal line of reflection is below the center of  $C_1$ , the reflected left side must overlap  $C_1$  and thus intersect the interior of the disk with boundary  $C_1$ . ■

**Corollary 3.13.** *If  $r \geq 3$ , then a minimal square cannot be in Con17.*

*Proof.* Suppose for contradiction that  $r \geq 3$  and a minimal square is in Con17. Then the side length,  $s$ , must be less than or equal to the side lengths of all squares in all other configurations, including Con10 which is addressed in Lemma 3.11. In particular, we must have  $s \leq M$  from Lemma 3.11. Let  $h$  be the distance from  $L$  to the midpoint of the upper right side of the square. We know that  $h$  can be no greater than  $(\sqrt{5}/2)s$ , because this is the highest that a midpoint of a side can possibly be, occurring if the midpoint is straight above  $V_{\text{dn}}$ . In summary,  $h \leq (\sqrt{5}/2)s \leq (\sqrt{5}/2)M$ . Therefore,  $h \leq r - r/\sqrt{2}$  if

$$\frac{\sqrt{5}}{2} \left( \frac{2r}{r + \sqrt{8}\sqrt{r} + 1} \right) \leq r - \frac{r}{\sqrt{2}},$$

which is true if  $r \geq 3$ . However, Lemma 3.12 then implies that the square is not minimal. This gives a contradiction. ■

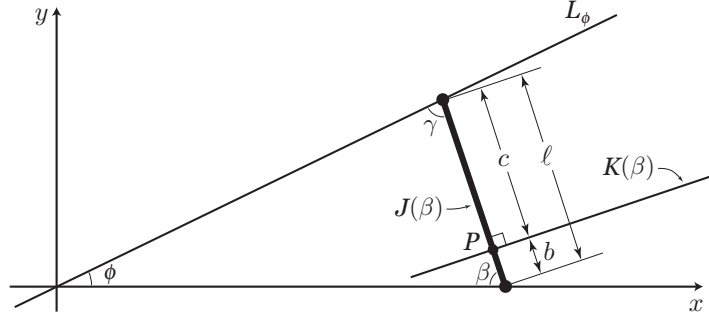
**Definition 3.14.** For a family  $\{J(t)\}_{t \in I}$  of lines or line segments, where  $I$  is an interval in  $\mathbb{R}$ , the *pivot* at a given  $t \in I$  is the point about which the line or line segment is pivoting at  $t$ , if such a point exists. In other words, for a given  $t \in I$ , if there exists  $\epsilon > 0$  such that  $u \in I \cap (t - \epsilon, t + \epsilon)$  implies that  $J(u) \cap J(t)$  consists of a single point, then the pivot,  $P$ , at  $t$  is defined by

$$P = \lim_{u \rightarrow t} J(u) \cap J(t)$$

if the limit exists. (The limit is one-sided if  $t$  is an endpoint of  $I$ .)

**Lemma 3.15.** *Let  $L_\phi$  be the line through the origin having angle  $\phi \in (0, \pi/2)$  with the positive  $x$ -axis as in Figure 14, define an interval  $I = (\pi/2 - \phi, \pi/2)$ , and fix  $\ell > 0$ . Let  $\{J(\beta)\}_{\beta \in I}$  be the family of line segments of length  $\ell$  with endpoints on the*

$x$ -axis and  $L_\phi$ , and for which  $J(\beta)$  intersects the  $x$ -axis at angle  $\beta$  as in Figure 14. Then the following hold.



**Figure 14.** Notation for Lemma 3.15.

(i) For  $\{J(\beta)\}_{\beta \in I}$ , the position of the pivot  $P$  at  $\beta$  is determined by

$$b(\cot \beta) = c(\cot \gamma),$$

where  $b$  and  $c$  are the distances along  $J(\beta)$  from  $P$  to the  $x$ -axis and  $L_\phi$ , respectively, and  $\gamma = \pi - \phi - \beta$  is the angle between  $J(\beta)$  and  $L_\phi$ .

(ii) Let  $\{K(\beta)\}_{\beta \in I}$  be the family of lines such that  $K(\beta)$  is the line perpendicular to  $J(\beta)$  through  $J(\beta)$ 's pivot. Then for  $\{K(\beta)\}_{\beta \in I}$ , the pivot at  $\beta$  has  $y$ -coordinate

$$\ell \left( \frac{\cos \gamma + \cos(\gamma - \beta) \cos \beta}{\sin(\gamma + \beta)} \right).$$

*Proof.* Given such  $L_\phi$ ,  $\ell$ ,  $\{J(\beta)\}_{\beta \in I}$ , and  $\{K(\beta)\}_{\beta \in I}$ , fix a value  $\beta_0 \in I$ .

To prove (i), let  $P$  be the pivot at  $\beta_0$  for  $\{J(\beta)\}_{\beta \in I}$ , and let  $b$  and  $c$  be the distances along  $J(\beta_0)$  from  $P$  to the  $x$ -axis and  $L_\phi$ , respectively. Define a family  $\{H(\beta)\}_{\beta \in (0, \pi/2)}$  of variable-length line segments as those for which  $H(\beta)$  is the line segment through  $P$  with endpoints on the  $x$ -axis and  $L_\phi$ , and that intersects the  $x$ -axis at angle  $\beta$ ; e.g.,  $H(\beta_0) = J(\beta_0)$ . It follows that  $d(\text{length of } H(\beta))/d\beta = 0$  at  $\beta = \beta_0$  because  $H(\beta)$  and  $J(\beta)$  have the same pivot at  $\beta_0$ , so the instantaneous rates of change of location of their respective endpoints (on the  $x$ -axis and  $L_\phi$ ) are the same. Let  $B(\beta)$  and  $C(\beta)$  be the distances along  $H(\beta)$  from  $P$  to the  $x$ -axis and  $L_\phi$ , respectively, so  $B'(\beta_0) + C'(\beta_0) = 0$ . Basic trigonometry shows that  $B(\beta) = h_B/\sin \beta$  and  $C(\beta) = h_C/\sin \gamma$ , where  $\gamma = \pi - \phi - \beta$ , and  $h_B$  and  $h_C$  are the (shortest) distances from  $P$  to the  $x$ -axis and  $L_\phi$ , respectively. Putting all of this together, (i) follows because  $b = B(\beta_0)$  and  $c = C(\beta_0)$ .

For (ii), we note that  $K(\beta)$  is the line

$$y = \cot \beta \left( x - \left( \frac{\ell \sin \gamma}{\sin \phi} - b \cos \beta \right) \right) + b \sin \beta,$$

$$\text{i.e., } y = (\cot \beta)x - \ell \left( \frac{\sin(\phi + 2\beta)}{\sin \phi \sin \beta} \right),$$

where the second form uses  $\gamma = \pi - \phi - \beta$ , as well as a substitution for  $b$  stemming from (i) which gives  $b(\cot \beta) = (\ell - b)(\cot \gamma)$ .

For all  $\beta$ , the slope of  $K(\beta)$  is greater than zero, so the pivot for  $\{K(\beta)\}_{\beta \in I}$  at  $\beta_0$  can be found as follows. For any given  $y$ , define the function  $f_y(\beta)$  giving the value of  $x$  such that  $(x, y)$  is on the line  $K(\beta)$ :

$$f_y(\beta) = (\tan \beta)y + \ell \left( \frac{\sin(\phi + 2\beta)}{\sin \phi \cos \beta} \right).$$

Then the pivot at  $\beta_0$  must have  $y$ -coordinate such that  $f'_y(\beta_0) = 0$ . Solving this equation for  $y$  gives (ii). ■

**Lemma 3.16.** *A minimal square cannot be in Con8, Con9, or Con17.*

*Proof.* Suppose for contradiction that a minimal square exists in  $\text{Con8}^+$ . Noting that the extended version of Con8 with  $r = r_0 \leq 1$  is equivalent to Con17 with  $r = 1/r_0$ , we may apply Corollary 3.13, concluding that  $r > 1/3$ .

Let  $I = [\beta_1, \beta_2]$  be the set of angles between  $L$  and the lower left sides of the squares in  $\text{Con8}^+$ , and let  $\beta_0 \in I$  be the maximum such angle of a minimal square in  $\text{Con8}^+$ . Then we know that  $\beta_0 \neq \beta_2$ , and that  $\beta_0 \neq \beta_1$  unless  $r = 1$ , because a minimal square cannot be in Con7, Con10, Con16, or Con18. Let  $\ell$  be the side length of the minimal square, and let  $\{G(\beta)\}_{\beta \in I}$  be the family of line segments of length  $\ell$  with endpoints on  $L$  and  $C_1$ , for which  $G(\beta)$  intersects  $L$  at angle  $\beta$ . Let  $m(\beta)$  be the distance from  $O_r$  to  $G(\beta)$ . Then  $m'(\beta_0) = 0$  because either  $m(\beta_0)$  is a minimal value on an open interval, or the square for  $\beta_0$  is in Con9 so the shortest distance from  $O_r$  to  $G(\beta_0)$  is the distance from  $O_r$  to  $G(\beta_0)$ 's endpoint at  $L$ , which is the pivot for  $\{G(\beta)\}_{\beta \in I}$  at  $\beta_0$ .

Let  $\{H(\beta)\}_{\beta \in I}$  be the family of lines such that  $H(\beta)$  is perpendicular to  $G(\beta)$  through  $G(\beta)$ 's pivot. Let  $h(\beta)$  be the vertical distance from  $O_r$  to  $H(\beta)$ , but considered negative if  $H(\beta)$  is below  $O_r$ . The sign of  $m'(\beta)$  is the same as the sign of  $h(\beta)$ , because  $h(\beta) > 0$  means that  $G(\beta)$  is moving away from  $O_r$  as  $\beta$  increases, and vice versa. Therefore,  $h(\beta_0) = 0$ .

Let  $L_0$  be the line tangent to  $C_1$  at  $G(\beta_0)$ 's endpoint on  $C_1$ . It follows from this tangency of  $L_0$  and  $C_1$  that the pivot at  $\beta_0$  for  $\{G(\beta)\}_{\beta \in I}$  is the same as the pivot at  $\beta_0$  for the family  $\{J(\beta)\}_{\beta \in I}$  of line segments of length  $\ell$  with endpoints on  $L$  and  $L_0$  with  $J(\beta)$  intersecting  $L$  at angle  $\beta$ . Intuitively, this match of the pivot is because  $J(\beta)$  acts like  $G(\beta)$  near  $\beta_0$ , and can be seen formally by standard  $\epsilon$ - $\delta$  reasoning. It also follows that the pivot  $P$  at  $\beta_0$  for  $\{H(\beta)\}_{\beta \in I}$  is the same as the pivot at  $\beta_0$  for the family  $\{K(\beta)\}_{\beta \in I}$  of lines such that  $K(\beta)$  is perpendicular to  $J(\beta)$  through  $J(\beta)$ 's pivot.

We now apply Lemma 3.15(ii) to  $\{K(\beta)\}_{\beta \in I}$  in order to show that the distance from  $L$  to  $P$  is less than  $r$ . This distance, the “ $y$ -coordinate” of  $P$ , is thus given by

$$y_P = \ell \left( \frac{\cos \gamma + \cos(\gamma - \beta) \cos \beta}{\sin(\gamma + \beta)} \right),$$

where  $\gamma$  is the angle that  $J(\beta)$  makes with  $L_0$ . We claim that  $y_P \leq \ell\sqrt{2}$ , and begin the proof by noting that for angle  $\phi$  between  $L$  and  $L_0$  we have  $\phi + \gamma + \beta = \pi$ , along with  $\phi, \beta \in (0, \pi/2)$ ,  $\gamma \in (0, \pi/2]$ , and  $\gamma + \beta \in (\pi/2, \pi)$ . Let  $v = (\gamma + \beta)/2$  and  $w = (\gamma - \beta)/2$ , and let

$$f(v, w) = \ell \left( \frac{\cos(v + w) + \cos(2w) \cos(v - w)}{\sin(2v)} \right)$$

so that  $y_P \leq \ell\sqrt{2}$  can be proved by showing that  $f(v, w) \leq \sqrt{2}$  on the domain defined by  $v + w \leq \pi/2$ ,  $v - w < \pi/2$ , and  $v \geq \pi/4$ . Although  $v$  cannot take the value  $\pi/4$  in the geometric interpretation, this extension of the domain of  $f$  to include  $v = \pi/4$  facilitates phrasing in the proof of a bound.

Straightforward computations show that

$$\frac{\partial f}{\partial v} = \frac{-\cos^3 \beta - 3 \cos \gamma \cos \beta}{\sin^2(\gamma + \beta)} < 0$$

on the domain, so the maximum occurs on the border  $v = \pi/4$ . Let

$$g(w) = f\left(\frac{\pi}{4}, w\right) = \cos\left(w + \frac{\pi}{4}\right) + \cos(2w) \cos\left(w - \frac{\pi}{4}\right).$$

The maximum value of  $g(w)$  occurs at  $w = 0$ , because

$$g'(w) = \sqrt{8} (\cos^3 w - \cos w - \cos^2 w \sin w)$$

and  $w \in (-\pi/4, \pi/4)$  on the domain, so  $g(w)$  increases then decreases, and moreover,  $g'(w) = 0$  only at  $w = 0$ . Therefore, we have  $f(v, w) \leq f(\pi/4, 0) = \sqrt{2}$  for all  $(v, w)$  in the domain of  $f$ , proving that  $y_P \leq \ell\sqrt{2}$ .

To complete the proof that the distance from  $L$  to  $P$  is less than  $r$ , we note that because  $\ell$  is the minimum side length of a square, Lemma 3.11 implies that

$$y_P \leq \ell\sqrt{2} \leq \left(\frac{2r}{r + \sqrt{8}\sqrt{r} + 1}\right) \sqrt{2}.$$

Since we know from Corollary 3.13 that  $r > 1/3$ , the fact that  $y_P < r$  easily follows.

Now, because  $H(\beta_0)$  has positive slope, and pivot  $P$  on  $H(\beta_0)$  is at a distance less than  $r$  from  $L$ ,  $P$  is below and to the left of  $O_r$ . Recalling that  $h(\beta)$  is the (signed) vertical distance from  $O_r$  to  $H(\beta)$ , the fact that pivot  $P$  is to the left of  $O_r$  implies that  $h'(\beta_0) < 0$ . We already know that  $h(\beta_0) = 0$ , so there exists  $\epsilon > 0$  such that  $\beta \in (\beta_0, \beta_0 + \epsilon)$  implies  $h(\beta) < 0$ . However, since  $h(\beta)$  and  $m'(\beta)$  have the same sign, this means  $m'(\beta) < 0$  on  $(\beta_0, \beta_0 + \epsilon)$ , contradicting  $m(\beta_0)$  being minimal. ■

**Proposition 3.17.** *A minimal square occurs only in Con6, and possibly in Con19, and must have  $V_1$  lower than  $O_1$ . Con19 has a minimal square if and only if  $r = 1$ .*

*Proof.* This result follows from the lemmas ruling out all other configurations, and the fact that Con19 is the reflection of Con6. ■

**4. EQUATIONS FOR MINIMUM SIDE LENGTH.** In this section we derive equations toward the pursuit of the minimum side length of an inscribed square, knowing from Proposition 3.17 that this minimum occurs in Con6 and has  $V_1$  lower than  $O_1$ . We continue using here the notation introduced at the beginnings of Sections 2 and 3.

**Lemma 4.1.** *In Con6 with  $V_1$  lower than  $O_1$ ,  $s(\theta)$  decreases then increases, both strictly monotonically, as a function of  $\theta$ . Thus, there is a unique minimal square in Con6.*

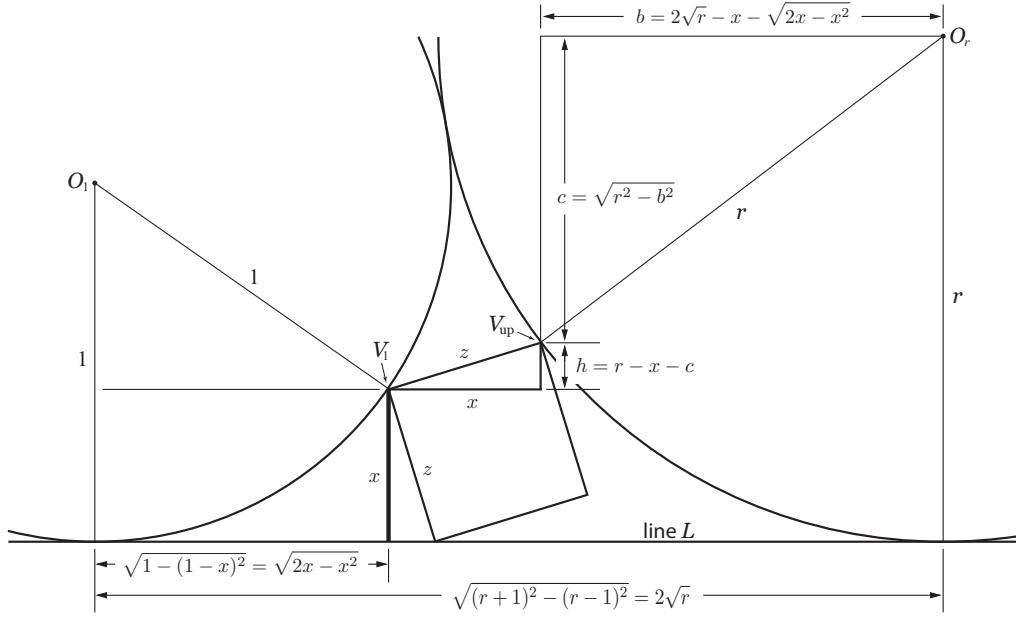
*Proof.* As  $\theta$  increases in Con6 with  $V_1$  lower than  $O_1$ ,  $T_L$  moves to the right and  $V_1$  moves down, so the intersection between  $T_1$  (which has negative slope) and  $T_L$  moves



down. Simultaneously,  $V_r$  moves up, so the intersection between  $T_r$  and  $T_L$  moves up. Therefore, as  $\theta$  increases, the triangle formed by  $T_L$ ,  $T_1$ , and  $T_r$  switches from the right side of  $T_L$  to the left side of  $T_L$  at a unique value, when  $T_1$  and  $T_r$  intersect  $T_L$  at the same point. By Lemma 3.1,  $s(\theta)$  decreases then increases, both strictly monotonically.

The uniqueness of the minimal square in Con6 follows from this and Proposition 3.17.  $\blacksquare$

For the equations used in seeking the minimum side length of an inscribed square, the parameter representing orientation will be the distance,  $x$ , from the line  $L$  to the upper left vertex  $V_1$  of the square, as shown in Figure 15. It is not difficult to see that  $x$  is related strictly monotonically to  $\theta$ .



**Figure 15.** Equations used to compute the side length of a square in Con6, given distance  $x$  from the line  $L$  to the upper left vertex  $V_1$  of the square. Here the side length is denoted  $z$ .

**Proposition 4.2.** *The minimum side length,  $\mu(r)$ , of an inscribed square is the minimum value of the function*

$$z(x) = \left[ x^2 + \left( r - x - \sqrt{r^2 - \left( 2\sqrt{r} - x - \sqrt{2x - x^2} \right)^2} \right)^2 \right]^{\frac{1}{2}} \quad (1)$$

on the interval  $(1 - 1/\sqrt{2}, 1)$ . The minimum occurs at a unique point, denoted  $x_m$ . Moreover, the function  $z$  is strictly decreasing for  $x < x_m$  and strictly increasing for  $x > x_m$ .

*Proof.* Let  $(x_1, x_2)$  be the interval of values of the distance  $x$  from  $L$  to  $V_1$  such that the associated square is in Con6 with  $V_1$  lower than  $O_1$ . We know from Proposition 3.17 that the minimum side length corresponds to some  $x \in (x_1, x_2)$ .

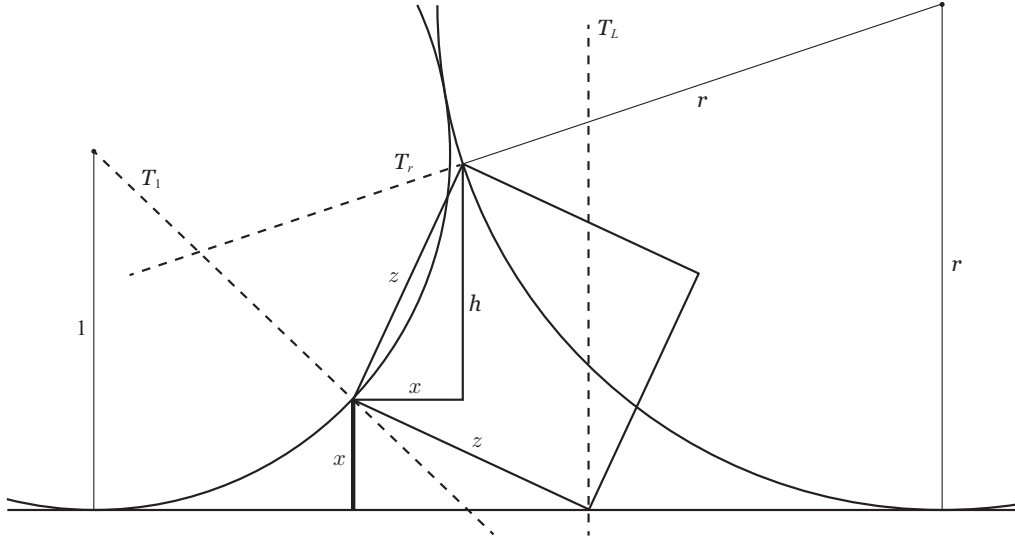
If  $x \in (x_1, x_2)$ , then as illustrated in Figure 15, congruent triangles using  $x$  and  $z$  show that  $V_{\text{up}}$  is at a horizontal distance  $x$  from  $V_1$ . Therefore, for  $x \in (x_1, x_2)$ ,

$$z = (x^2 + h^2)^{\frac{1}{2}},$$

where  $h$  is the vertical distance from  $V_1$  to  $V_{\text{up}}$ . Thus by the equations in Figure 15, the side length is given as a function of  $x \in (x_1, x_2)$  by  $z(x)$  in equation (1). In addition, it follows from Lemma 4.1 that on  $(x_1, x_2)$  the minimum occurs at a unique point,  $x_m$ , and that  $z$  is strictly decreasing for  $x < x_m$  and strictly increasing for  $x > x_m$ .

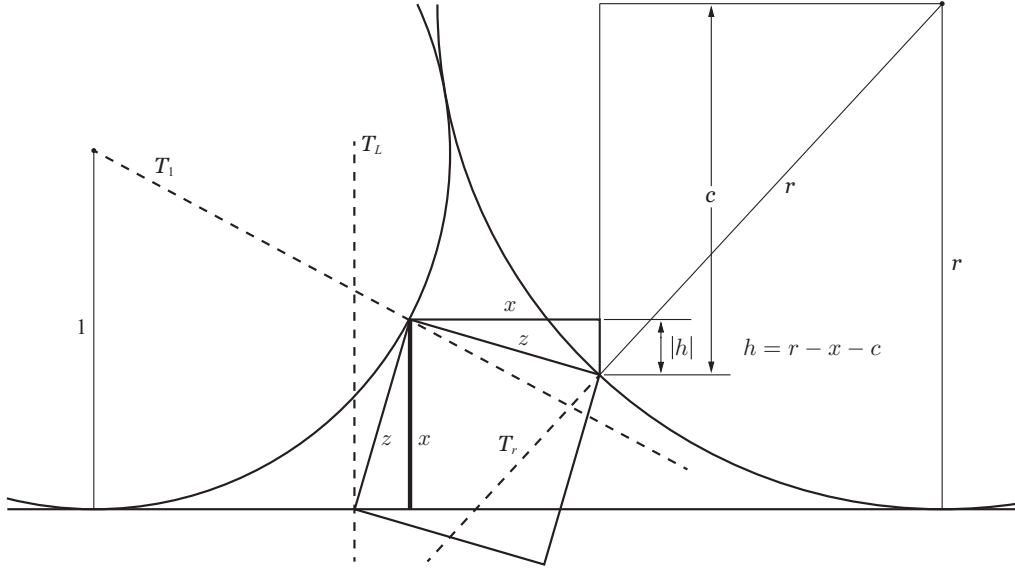
However, the stated domain  $(1 - 1/\sqrt{2}, 1)$  extends beyond  $(x_1, x_2)$ ; certainly,  $x_2 < 1$  by definition. Toward the claim that  $x_1 > 1 - 1/\sqrt{2}$ , we suppose for contradiction that  $x = 1 - 1/\sqrt{2}$  is possible. Because  $V_1$  is distance  $x$  from  $L$ ,  $x = 1 - 1/\sqrt{2}$  occurs exactly when the line through  $O_1$  and  $V_1$  has slope  $-1$ . In that case,  $1 - 1/\sqrt{2}$  is also the horizontal distance from  $V_1$  to a vertical line tangent to  $C_1$  on the right side. Because  $V_{\text{up}}$  is horizontal distance  $x (= 1 - 1/\sqrt{2})$  from  $V_1$ ,  $V_{\text{up}}$  is on that vertical line. This contradicts the fact that  $V_{\text{up}}$  is on  $C_r$  which requires  $V_{\text{up}}$  to be to the right of that vertical line. Noting that  $x < 1 - 1/\sqrt{2}$  would result in a similar contradiction, we conclude that  $x_1 > 1 - 1/\sqrt{2}$ .

Now considering  $z$  in equation (1) as a function of  $x \in (1 - 1/\sqrt{2}, 1)$ , we know on  $(x_1, x_2)$  that  $z$  represents the side length of an inscribed square, but outside  $(x_1, x_2)$  we only have equation (1). It is helpful to set up geometric interpretations for the cases  $x \in (1 - 1/\sqrt{2}, x_1)$  and  $x \in (x_2, 1)$  in order to complete the proof by showing that  $z$  strictly decreases when  $x \in (1 - 1/\sqrt{2}, x_1)$  and strictly increases when  $x \in (x_2, 1)$ .



**Figure 16.** Geometry showing that  $z$  given by (1) is not a minimum when  $x < x_1$ .

Figures 16 and 17 illustrate geometric interpretations of  $z$  as given by equation (1) when  $x$  belongs to the intervals  $(1 - 1/\sqrt{2}, x_1)$  and  $(x_2, 1)$ , respectively. Here, the three consecutive vertices still lie on the line and the two circles, although the squares are no longer inscribed. In both cases, the same equations as in Figure 15 still apply, and lead to (1). The fact that  $z$  strictly decreases when  $x \in (1 - 1/\sqrt{2}, x_1)$  and strictly



**Figure 17.** Geometry showing that  $z$  given by (1) is not a minimum when  $x > x_2$ .

increases when  $x \in (x_2, 1)$  can be seen by the same reasoning as in Lemma 4.1, this time using  $T_1$ ,  $T_r$ , and  $T_L$  defined as before except that the “point of contact” is specifically that with the relevant vertex, as shown in Figures 16 and 17. The reasoning then uses a modified form of Lemma 3.1 that applies to these squares with consecutive vertices on the line and circles, instead of to inscribed squares, and whose proof is analogous to that of Lemma 3.1. ■

**Remark 4.3 (approximating  $\mu(r)$ ).** From Proposition 4.2, one can deduce an algorithm for computing an approximation of  $\mu(r)$  given the radius  $r$ . Indeed, it suffices for this purpose to minimize the function  $z$ , which can be achieved by applying root-finding methods to  $z'$ .

**Proposition 4.4.** *In the result of Proposition 4.2, i.e., that the minimum side length is given by  $\mu(r) = z(x_m)$ , the number  $x_m$  is a root of the 10th degree polynomial*

$$f_4(f_1f_7 + f_2f_6 - 2f_3f_5)^2 - (f_3^2f_4 + f_5^2 - f_1f_4f_6 - f_2f_7)^2, \quad (2)$$

where, letting  $k = \sqrt{r}$ ,

$$f_1(x) = -2x + 4k$$

$$f_2(x) = (4k - 2)x + k^4 - 4k^2$$

$$f_3(x) = (6k - 3)x + k^4 - 2k^3 - 3k^2$$

$$f_4(x) = -x^2 + 2x$$

$$f_5(x) = 4x^3 - (2k^2 + 6k + 7)x^2 + (2k^3 + 3k^2 + 10k)x - 2k^3$$

$$f_6(x) = 8x^3 + (-4k^2 - 16)x^2 + (4k^3 - 2k^2 + 6)x - 4k^3 + 8k^2 - 4k$$

$$f_7(x) = (4k^2 - 16k)x^3 + (-k^4 + 4k^3 - 10k^2 + 40k)x^2$$

$$+ (2k^4 - 8k^3 + 4k^2 - 20k + 2)x + 4k^2.$$

*Proof.* By Proposition 4.2,  $\mu(r) = z(x_m)$ , where  $z(x)$  is given by equation (1), and  $x_m$  is the unique point in the interval  $(1 - 1/\sqrt{2}, 1)$  such that  $dz/dx = 0$ , or equivalently  $dA/dx = 0$  where  $A(x) = (z(x))^2$ .

It remains to show that  $x_m$  is a root of (2). Let

$$g(x) = (-2x + 4k)\sqrt{f_4(x)} + (4k - 2)x + k^4 - 4k^2,$$

so that expansion in (1) gives

$$A(x) = 2\left((x - k^2)\sqrt{g(x)} + (-x + 2k)\sqrt{f_4(x)} + x^2 + (-k^2 + 2k - 1)x + k^4 - 2k^2\right).$$

Note that

$$\frac{dg}{dx} = \frac{(4k - 2)\sqrt{f_4(x)} + 4x^2 - (4k + 6)x + 4k}{\sqrt{f_4(x)}},$$

which leads to

$$\begin{aligned} \frac{1}{2} \frac{dA}{dx} &= \frac{(x - k^2)\left((2k - 1)\sqrt{f_4(x)} + 2x^2 - (2k + 3)x + 2k\right)}{\sqrt{g(x)}\sqrt{f_4(x)}} + \sqrt{g(x)} \\ &\quad + \frac{x^2 - (2k + 1)x + 2k}{\sqrt{f_4(x)}} - \sqrt{f_4(x)} + 2x - k^2 + 2k - 1. \end{aligned}$$

In view of the fact that  $dA/dx = 0$  at  $x = x_m$ , set  $\frac{1}{2}dA/dx = 0$ , then multiply by  $\sqrt{g(x)}\sqrt{f_4(x)}$  and separate terms with  $\sqrt{g(x)}$  from the rest to obtain

$$\begin{aligned} &\left((2x - k^2 + 2k - 1)\sqrt{f_4(x)} + 2x^2 - (2k + 3)x + 2k\right)\sqrt{g(x)} \\ &= -f_3(x)\sqrt{f_4(x)} - f_5(x). \end{aligned}$$

Note that  $g(x) = f_1(x)\sqrt{f_4(x)} + f_2(x)$ , so that squaring both sides and omitting “(x)” for readability results in

$$(f_6\sqrt{f_4} + f_7)\left(f_1\sqrt{f_4} + f_2\right) = f_3^2 f_4 + 2f_3 f_5 \sqrt{f_4} + f_5^2.$$

Separating terms with  $\sqrt{f_4}$  from the rest gives

$$(f_1 f_7 + f_2 f_6 - 2f_3 f_5)\sqrt{f_4} = f_3^2 f_4 + f_5^2 - f_1 f_4 f_6 - f_2 f_7.$$

Squaring both sides and rearranging leads to the final polynomial equation,

$$f_4(f_1 f_7 + f_2 f_6 - 2f_3 f_5)^2 - (f_3^2 f_4 + f_5^2 - f_1 f_4 f_6 - f_2 f_7)^2 = 0.$$

This polynomial has degree 10. Terms of higher degree could arise only from  $f_5^2$  and  $f_1f_4f_6$ , each of which has degree 6, but both  $f_5^2$  and  $f_1f_4f_6$  have  $16x^6$  as their 6th degree term, so  $f_5^2 - f_1f_4f_6$  has degree 5.

In summary,  $x_m$  is a root of the polynomial in (2). ■

**5. NONEXISTENCE OF A SOLUTION BY RADICALS.** In this section we prove that  $\mu$  is not a radical function as defined in Section 1. For convenience we work instead with the function  $\lambda$  defined by  $\lambda(c) = (\mu(c^2))^2$ . It is intuitively clear that if  $\mu$  is a radical function then  $\lambda$  is also radical; a formal proof of this fact is given in a more general context in Lemma 5.1 below. The proof of our main result, Theorem 5.4, shows that  $\lambda$  is not radical on any infinite subset of  $[1, \infty)$ . As a consequence we obtain the fact that  $\mu$  cannot be radical on any such set. We refer the reader to [3, Chapters 13, 14] for the algebraic background assumed in this section.

**Lemma 5.1.** *Let  $J$  be a nonempty subset of  $[1, \infty)$ , let  $n \in \mathbb{Z}^+$ , and let  $I = \sqrt[n]{J}$  be the set of positive  $n$ th roots of elements of  $J$ . Suppose that a function  $f : J \rightarrow \mathbb{R}$  is radical, and define  $g : I \rightarrow \mathbb{R}$  and  $h : J \rightarrow \mathbb{R}$  by  $g(c) = f(c^n)$  and  $h(c) = (f(c))^n$ . Then  $g$  and  $h$  are radical.*

*Proof.* We begin by showing that  $g$  is radical. Let  $q \in \mathbb{C}[k, x]$  be a nonzero polynomial satisfying  $q(c, f(c)) = 0$  for every  $c \in J$ , and such that there is a radical extension  $R/\mathbb{C}(k)$  containing a splitting field  $S$  of  $q$ . Let  $\varphi : \mathbb{C}(k) \rightarrow \mathbb{C}(k)$  be the embedding induced by the map  $k \mapsto k^n$ , and let  $\Omega$  be an algebraic closure of  $\mathbb{C}(k)$ . By basic field theory (see [7, Chapter V, §2, Theorem 2.8]), we may extend the map  $\varphi$  to an embedding  $\varphi : R \rightarrow \Omega$ . Defining  $Q(k, x) = q(k^n, x)$  we have  $Q(c, g(c)) = 0$  for every  $c \in I$ . Moreover, since  $Q$  is the polynomial obtained by applying  $\varphi$  to the coefficients of  $q$ , the above observations imply that  $Q$  splits in the field  $\varphi(S)$ , which is contained in the radical extension  $\varphi(R)/\mathbb{C}(k)$ . Thus  $g$  is a radical function.

Next we show that  $h$  is radical. The argument will be given assuming that  $q$  is monic and has no repeated root; the general case can be proved similarly. Let  $\alpha_1, \dots, \alpha_d$  be the roots of  $q$  in  $S$ , and let

$$Q(k, x) = (x - \alpha_1^n) \cdots (x - \alpha_d^n).$$

Note that since  $q$  has coefficients in  $\mathbb{C}[k]$ , the same holds for  $Q$ . (The elementary symmetric functions of  $\alpha_1^n, \dots, \alpha_d^n$  are polynomials in the elementary symmetric functions of  $\alpha_1, \dots, \alpha_d$ .) Moreover, as explained below, one can show that  $Q(c, h(c)) = 0$  for every  $c \in J$ . Since  $Q$  splits in the field  $\mathbb{C}(k, \alpha_1^n, \dots, \alpha_d^n) \subseteq S \subseteq R$ , this implies that  $h$  is radical.

Fixing  $c \in J$ , the fact that  $Q(c, h(c)) = 0$  can be seen heuristically first: Since

$$q(k, x) = (x - \alpha_1) \cdots (x - \alpha_d)$$

and  $q(c, f(c)) = 0$ , we must have  $f(c) = \alpha_i(c)$  for some  $i$ , and thus  $Q(c, h(c)) = 0$ . This argument can be made rigorous by extending the map  $k \mapsto c$  to a ring homomorphism  $\mathbb{C}[k, \alpha_1, \dots, \alpha_d] \rightarrow \mathbb{C}$ , so that  $\alpha_i(c)$  is well-defined. We refer the interested reader to Section 3 in [7, Chapter VII] for the necessary tools. ■

Next we prove two preliminary results needed to show that  $\lambda$  is not radical.

**Lemma 5.2.** *Let  $p(k, x)$  be the polynomial defined by (2) considering  $k$  and  $x$  as indeterminates. As an element of the ring  $\mathbb{C}(k)[x]$ , the polynomial  $p$  is irreducible and has Galois group isomorphic to the symmetric group  $S_{10}$ .*

*Proof.* We rely on a computation carried out using the computer algebra system MAGMA; the code for our computation is available in the supplemental online material. Constructing  $p(k, x)$  as an element of the ring  $\mathbb{Q}(k)[x]$ , we use Sutherland's algorithm [9] to compute a permutation representation of the Galois group of  $p$  over  $\mathbb{C}(k)$ , and we obtain the group  $S_{10}$ . It follows that the Galois group acts transitively on the roots of  $p$ , so  $p$  is irreducible over  $\mathbb{C}(k)$ . ■

By Proposition 4.2 we may regard  $x_m$  as a function of  $r$  defined on the interval  $[1, \infty)$ , and moreover, we have

$$\mu(r) = z(x_m(r)) \quad \text{for all } r \geq 1. \quad (3)$$

For convenience we will make the change of variable  $k = \sqrt{r}$  and work instead with the function  $\xi : [1, \infty) \rightarrow \mathbb{R}$  be defined by  $\xi(k) = x_m(k^2)$ . From Proposition 4.4 we deduce that

$$p(k, \xi(k)) = 0 \quad \text{for all } k \in [1, \infty). \quad (4)$$

Furthermore, (3) implies that  $\lambda(k) = (z(\xi(k)))^2$ . Hence, writing  $\xi$  for  $\xi(k)$ , we have

$$\lambda(k) = \xi^2 + \left( k^2 - \xi - \sqrt{k^4 - \left( 2k - \xi - \sqrt{2\xi - \xi^2} \right)^2} \right)^2.$$

By manipulating the equation above we obtain a polynomial  $h \in \mathbb{Q}[k, x, y]$  with the property that

$$h(k, \xi(k), \lambda(k)) = 0 \quad \text{for all } k \in [1, \infty). \quad (5)$$

Explicitly,  $h$  is given by the formula

$$\begin{aligned} h(k, x, y) = & \left( (y - x^2 - c_1^2 - c_2 + c_3^2 + c_4)^2 + 4c_3^2c_4 - 4c_1^2c_2 + 4c_1^2c_3^2 + 4c_1^2c_4 \right)^2 \\ & - c_4 \left( 8c_1^2c_3 + 4c_3(y - x^2 - c_1^2 - c_2 + c_3^2 + c_4) \right)^2, \end{aligned}$$

where  $c_1 = k^2 - x$ ,  $c_2 = k^4$ ,  $c_3 = 2k - x$ , and  $c_4 = 2x - x^2$ .

**Lemma 5.3.** *Let  $\Omega$  be an algebraic closure of the field  $\mathbb{C}(k)$ . Suppose that  $\alpha, \beta \in \Omega$  satisfy  $p(k, \alpha) = h(k, \alpha, \beta) = 0$ . Then  $\mathbb{C}(k, \alpha) \subseteq \mathbb{C}(k, \beta)$ .*

*Proof.* We rely on a number of computations in MAGMA; the code used for all computations is available in the supplemental online material. Let  $F = \mathbb{Q}(k, \alpha)$ . To prove the lemma it suffices to show that  $F \subseteq \mathbb{Q}(k, \beta)$ . Regarding  $p$  as an element of the ring  $\mathbb{Q}(k)[x]$ , note that  $p$  is irreducible by Lemma 5.2, and  $\alpha$  is a root of  $p$  by hypothesis, so we may identify  $F$  with the field  $\mathbb{Q}(k)[x]/(p(k, x))$ . Constructing  $F$  in MAGMA and factoring<sup>1</sup> the polynomial  $h(k, \alpha, y)$  over  $F$ , we find that this polynomial has two roots in  $F$  and two roots that are quadratic over  $F$ . Note that  $\beta$  must be one of these four roots since  $h(k, \alpha, \beta) = 0$ .

Suppose that  $\beta \in F$ . Computing the minimal polynomial of  $\beta$  over  $\mathbb{Q}(k)$  we obtain a polynomial of degree 10; thus  $[\mathbb{Q}(k, \beta) : \mathbb{Q}(k)] = 10$ . Since  $\beta \in F$  and

$$[F : \mathbb{Q}(k)] = \deg(p) = 10,$$

<sup>1</sup>The algorithm used by MAGMA to factor polynomials over algebraic function fields is discussed in [10].

this implies that  $F = \mathbb{Q}(k, \beta)$ . In particular,  $F \subseteq \mathbb{Q}(k, \beta)$  as desired.

Now suppose that  $\beta$  is quadratic over  $F$ . Then a minimal polynomial computation shows that  $[\mathbb{Q}(k, \beta) : \mathbb{Q}(k)] = 20$ . Since  $[F(\beta) : F] = 2$  and  $[F : \mathbb{Q}(k)] = 10$ , this implies that  $F(\beta) = \mathbb{Q}(k, \beta)$ , so again  $F \subseteq \mathbb{Q}(k, \beta)$ . ■

We can now prove the main theorem of this article.

**Theorem 5.4.** *There is no infinite subset  $J \subseteq [1, \infty)$  such that  $\mu : J \rightarrow \mathbb{R}$  is radical.*

*Proof.* As above, let  $\Omega$  denote an algebraic closure of the field  $\mathbb{C}(k)$ . By Lemma 5.1, in order to prove the theorem it suffices to show that  $\lambda$  is not radical on any infinite subset of  $[1, \infty)$ . Suppose for contradiction that  $I \subseteq [1, \infty)$  is an infinite set such that  $\lambda : I \rightarrow \mathbb{R}$  is radical. Then there is a nonzero polynomial  $q(k, y) \in \mathbb{C}[k, y]$  whose Galois group over  $\mathbb{C}(k)$  is solvable, and such that

$$q(k, \lambda(k)) = 0 \quad \text{for all } k \in I. \quad (6)$$

Regarding  $q$  and  $h$  as elements of the ring  $\mathbb{C}[k, x, y]$ , let

$$f(k, x) = \text{Res}_y (h(k, x, y), q(k, x, y)),$$

where  $\text{Res}_y$  denotes the resultant as polynomials in  $y$ . By (5) and (6), for every  $k \in I$  the polynomials  $h(k, \xi(k), y)$  and  $q(k, \xi(k), y)$  have a common root, namely  $\lambda(k)$ ; hence

$$f(k, \xi(k)) = 0 \quad \text{for all } k \in I. \quad (7)$$

Similarly, letting

$$g(k) = \text{Res}_x (f(k, x), p(k, x)) \in \mathbb{C}[k],$$

equations (4) and (7) imply that  $g(k) = 0$  for every  $k \in I$ . Since  $I$  is an infinite set, we must have  $g = 0$ . Therefore,  $f$  and  $p$  have a common root  $\alpha \in \Omega$ . Given that  $f(k, \alpha) = 0$ , the definition of  $f$  implies that  $q(k, y)$  and  $h(k, \alpha, y)$  have a common root  $\beta \in \Omega$ . Note that the assumptions in Lemma 5.3 are satisfied.

Let  $N \subset \Omega$  be the splitting field of  $q(k, y)$  over  $\mathbb{C}(k)$  and let  $F = \mathbb{C}(k, \alpha)$ . By Lemma 5.3 we have  $F \subseteq \mathbb{C}(k, \beta)$  and therefore  $F \subseteq N$ . Letting  $L \subset \Omega$  be the splitting field of  $p(k, x)$  over  $\mathbb{C}(k)$ , we have  $L \subseteq N$  since  $F \subseteq N$  and  $L$  is the Galois closure of the extension  $F/\mathbb{C}(k)$ . Since the extension  $L/\mathbb{C}(k)$  is Galois, the group  $\text{Gal}(L/\mathbb{C}(k))$  is a quotient of  $\text{Gal}(N/\mathbb{C}(k))$ . The definition of  $q(k, y)$  implies that the latter group is solvable, so the former is, too. This contradicts Lemma 5.2 (since the group  $S_{10}$  is not solvable), and thus completes the proof of the theorem. ■

**ACKNOWLEDGMENTS.** JEH received support from the Colby College Research Grant Program. The authors thank Gerardo Lafferriere of the Fariborz Maseeh Department of Mathematics and Statistics at Portland State University for welcoming JEH as a Visiting Scholar during the final stages of this project.

## REFERENCES

1. Bosma, W., Cannon, J., Playoust, C. (1997). The Magma algebra system. I. The user language. *J. Symbolic Comput.* 24(3–4): 235–265.
2. Coxeter, H. S. M. (1989). *Introduction to Geometry*, 2nd ed. New York: John Wiley & Sons, Inc.

3. Dummit, David S., Foote, Richard M. (2004). *Abstract Algebra*, 3rd ed. Hoboken, NJ: John Wiley & Sons, Inc.
4. Fukagawa, H., Pedoe, D. (1989). *Japanese Temple Geometry Problems*. Winnipeg, Canada: The Charles Babbage Research Centre.
5. Fukagawa, H., Rothman, T. (2008). *Sacred Mathematics: Japanese Temple Geometry*. Princeton, NJ: Princeton Univ. Press.
6. Kinoshita, H. (2018). An unsolved problem in the Yamaguchi's travel diary. *Sangaku Journal of Mathematics*. 2: 43–53. [www.sangaku-journal.eu](http://www.sangaku-journal.eu)
7. Lang, S. (2002). *Algebra*, 3rd ed. New York, NY: Springer-Verlag.
8. Rothman, T. (1998). Japanese temple geometry. *Scientific American*. 278(5): 84–91.
9. Sutherland, N. (2018). Computations with Galois groups in Magma. *Computeralgebra Rundbrief*. 62: 16–21. [www.fachgruppe-computeralgebra.de/data/CA-Rundbrief/car62.pdf](http://www.fachgruppe-computeralgebra.de/data/CA-Rundbrief/car62.pdf)
10. Trager, B. M. (1976). Algebraic factoring and rational function integration. In: Jenks, R. D., ed. *SYMSAC '76: Proceedings of the Third ACM Symposium on Symbolic and Algebraic Computation*. New York: ACM, pp. 219–226.

**JAN E. HOLLY** received a Ph.D. in mathematics from the University of Illinois, and is a mathematical neuroscientist or a pure mathematician depending on the day.

*Department of Mathematics and Statistics, Colby College, Waterville, ME 04901*

*Jan.Holly@colby.edu*

**DAVID KRUMM** received a Ph.D. in mathematics from the University of Georgia, and is currently in a visiting position at Reed College. He first became aware of the existence of *sangaku* while visiting the Todai-ji temple in Nara, Japan.

*Mathematics Department, Reed College, Portland, OR 97202*

*dkrumm@reed.edu*