A Diophantine Property Associated with Prime Twins

P. Shiu

CONTENTS

2000 AMS Subject Classification: Primary 11A41, 11Y11

Keywords: Prime twins, Diophantine difference equation

We show that there is initial order rather than chaos for the solution to a Diophantine difference equation when, and only when, the associated parameter takes the smaller values of prime twins.

1. INTRODUCTION

Let $1 \leq \ell \leq k$, and define the sequence (a_n) by the Diophantine difference equation

$$a_1 = \ell, \quad a_n = a_{n-1} + \left\lceil \frac{ka_{n-1}}{n} \right\rceil, \quad n > 1, \qquad (1\text{–}1)$$

where the ceiling function $\lceil x \rceil$ denotes the least integer not less than the real number x. The equation was first considered by J. H. Conway, H. T. Croft, P. Erdős, and M. J. T. Guy [Conway et al. 79] in their investigation of the distribution of values of angles determined by coplanar points, wherein they needed and found the solution corresponding to $k = 3$ and $\ell = 1$. The problem appears intractable when $k > 3$, but from computational experiments, I managed to find the solution

$$a_n = \left\lceil \frac{(n+1)\cdots(n+p) - (p-1)!(n+p)}{p\,p!} \right\rceil, \quad n \geq 1, \qquad (1\text{–}2)$$

for the case when $k = p$ is a prime and $\ell = 1$; see [Shiu 96], where the solutions for the cases $k = \ell = 2$ and $k = 3$, $\ell = 2, 3$ are given.

The Diophantine condition in (1–1) dictates that one has to determine the fractional parts

$$\theta_n = \left\lceil \frac{ka_{n-1}}{n} \right\rceil - \frac{ka_{n-1}}{n}, \quad n > 1, \qquad (1\text{–}3)$$

which amounts to the identification of the residue class (mod n) to which ka_{n-1} belongs. The direct proof of (1–2) given in [Shiu 96] requires the fact that $\theta_n = 0$ when $p|n$, and that θ_n has the asymptotic average value $1/p - 1/p^2$, but the explicit formula for θ_n given in (1–8) was not stated there.

For other values of the parameters k and ℓ, computational experiments indicate that θ_n seems to be chaotic,

1058-6458/2005$ 0.50 per page

leading to the speculation that if θ_n has an asymptotic average value, then the value should be $\frac{1}{2}$. However, nothing concrete has been established and the problem of whether θ_n can be zero infinitely often appears to be difficult. Nevertheless, the results in [Shiu 96] show that the primes are closely related to Equation (1–1), and recent computational experiments have led me to discover some further interesting properties related to the primes. I found that, when $\ell = 2, 3, 4$, there is order in the initial values of θ_n if, and only if, $k = p$, where p and $p+2$ both are primes. More specifically, for such ℓ and k, there is an explicit formula for θ_n which is valid up to $n \le p(p+2)$, and that, soon after this point, the values of θ_n appear to be chaotic. Moreover, for other values of ℓ and k, the values of θ_n appear to be chaotic from the beginning.

For odd k, it follows at once from (1–1) and (1–3) that $\theta_2 = 0$ or $\frac{1}{2}$, depending on whether ℓ is even or odd. We write

$$L = \left\lceil \frac{\ell}{2} \right\rceil, \qquad (1\text{–}4)$$

so that $L = 1, 2, 2$ correspond to $\ell = 2, 3, 4$, and we prove the following two theorems.

Theorem 1.1. *Let $\ell = 2, 3,$ or 4, and $k > 3$. Then a necessary and sufficient condition for*

$$\theta_n = \frac{L}{n} \qquad (1\text{–}5)$$

to hold for each n in $3 \le n < k$ is that k and $k+2$ are both primes.

Theorem 1.2. *Let $k = p > 3$, with p and $p+2$ prime. Then, for $p \le n < 2p$ and $\ell = 2, 4$, we have*

$$\theta_n = 0,\ \frac{2L}{p+1},\ \frac{L}{p+2},\ \frac{5L}{p+3},\ 0,\ \frac{2L}{p+5},\ \frac{2L}{p+6}, \dots, \frac{2L}{2p-1}, \qquad (1\text{–}6)$$

and for $\ell = 3$, we have

$$\theta_n = 0,\ \frac{4}{p+1},\ \frac{3}{p+2},\ \frac{7}{p+3},\ \frac{2}{p+4},\ \frac{4}{p+5},\ \frac{4}{p+6}, \dots, \frac{4}{2p-1}. \qquad (1\text{–}7)$$

It is easy to check that, when $\ell = 5$ and 6, we have $3|a_2$ so that $\theta_3 = 0$. In any case, (1–5) cannot hold when $L \ge n$, so we can only have $\ell = 2, 3,$ and 4. Note that $2p - 1 < p + 5$ when $p = 5$, so only the first five terms for θ_n in (1–6) and (1–7) are relevant. The extended formulae for θ_n in $2p \le n \le p(p+2)$ are given in Section 4; they can be proved in the same way, but there are more exceptional cases.

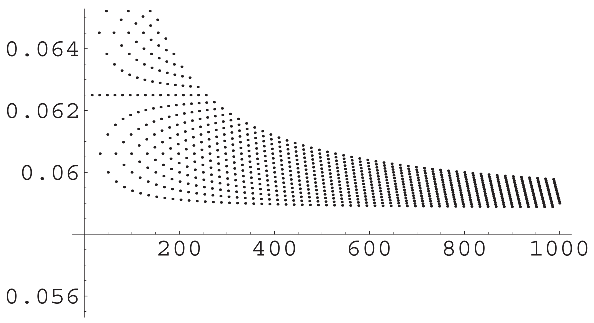

FIGURE 1. $k = 17$, $l = 1$ $(\theta_n : 2 \le n \le 1000)$.

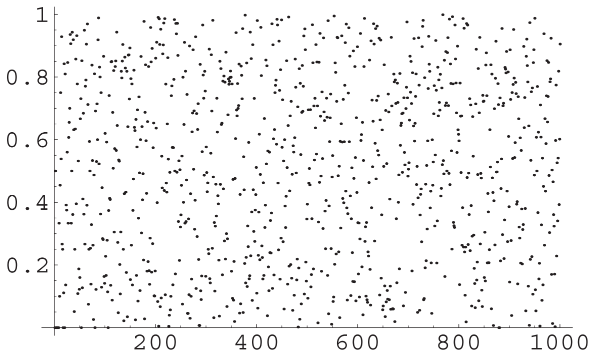

FIGURE 2. $k = 17$, $l = 5$ $(\theta_n : 2 \le n \le 1000)$.

The argument used here differs substantially from the direct proof of (1–2) given in [Shiu 96], and it is easy to adapt it to show that, for the case $k = p$ is a prime and $\ell = 1$,

$$\theta_{(r+1)p} = 0, \quad \theta_n = \frac{r+1}{n}, \qquad (1\text{–}8)$$

where $\max(1, rp) < n < (r+1)p$, for $r = 0, 1, 2, \dots$.

These results do not seem to have any practical value, such as using them as primality tests or prime twins tests, nevertheless they do give new criteria for the primes and prime twins, and it is by no means obvious that (1–1) should possess such properties.

We use Roman letters to denote integers, with p and q being reserved for primes. We write $\bar{n}$ for the integer reciprocal of n, that is $\bar{n}n \equiv 1$, with respect to the modulus being considered. Figures 1–6 are computer generated plots of θ_n, illustrating its different behaviour associated with the chosen parameters k and ℓ; in Figure 6 the values on the horizontal axis have been reduced by $10,000$. There seems little doubt from these experimental results

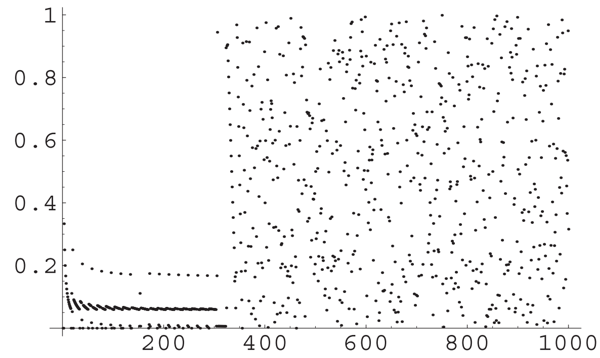

FIGURE 3. $k = 17$, $l = 2$ $(\theta_n : 2 \leq n \leq 1000)$.

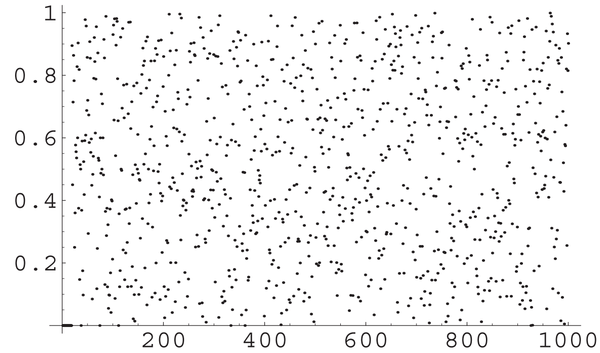

FIGURE 4. $k = 37$, $l = 2$ $(\theta_n : 2 \leq n \leq 1000)$.

that there is chaos in the system, unless the parameters concerned are as specified previously.

2. PROOF OF THEOREM 1.1

We use arithmetic (mod n) to prove the theorem, so we replace (1–5) by the equivalent formulation

$$ka_{n-1} \equiv -L \,(\mathrm{mod}\ n), \qquad 3 \leq n < k. \tag{2–1}$$

For the sufficiency part of the theorem, we write $k = p$, with $p, p+2$ primes. From (1–1) and (1–4) we find that

$$pa_2 = L(p+1)^2 - L, \tag{2–2}$$

for the cases $\ell = 2$ and 4, with the additional term $-p(p+1)/2$ being required for $\ell = 3$. Since p and $p+2$ are primes, we need to have $p + 1 \equiv 0 \,(\mathrm{mod}\ 3)$, so (2–1) certainly holds when $n = 3$. We proceed by induction on n in $3 < n < p$ and suppose that $pa_{m-1} + L \equiv 0 \,(\mathrm{mod}\ m)$ holds for $3 \leq m < n$. For $3 < j \leq n$ we then have, by (1–1),

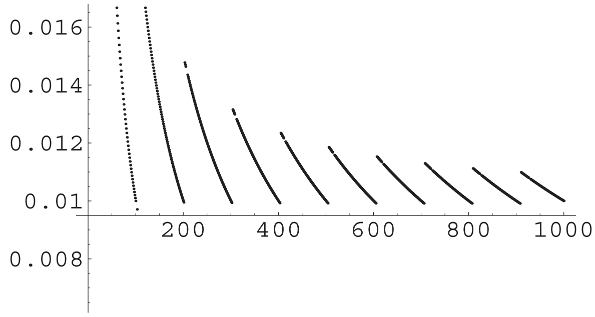

FIGURE 5. $k = 101$, $l = 2$ $(\theta_n : 2 \leq n \leq 1000)$.

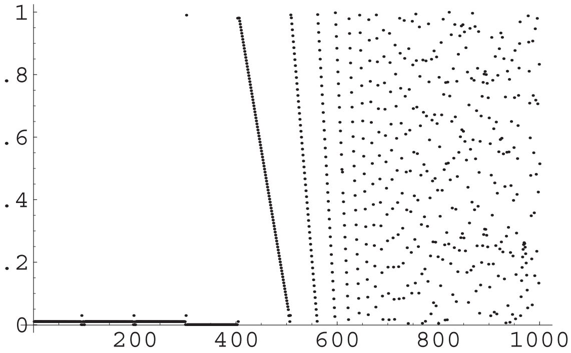

FIGURE 6. $k = 101$, $l = 2$ $(\theta_n : 10000 + 0 \leq n \leq 10000 + 1000)$.

$$\begin{aligned} pa_{j-1} + L &= p\Big(a_{j-2} + \left\lceil \frac{pa_{j-2}}{j-1} \right\rceil\Big) + L \\ &= p\Big(a_{j-2} + \frac{pa_{j-2} + L}{j-1}\Big) + L \\ &= \frac{(pa_{j-2} + L)(p + j - 1)}{j-1}, \end{aligned}$$

so that, for $3 < n < p$,

$$\begin{aligned} pa_{n-1} + L &= \frac{p+n-1}{n-1}(pa_{n-2} + L) = \cdots \\ &= \frac{p+n-1}{n-1}\frac{p+n-2}{n-2} \cdots \frac{p+3}{3}(pa_2 + L), \end{aligned}$$

that is

$$pa_{n-1} + L = n \cdot \binom{p+n}{n} \cdot \frac{2}{(p+n)(p+2)} \cdot \frac{pa_2 + L}{p+1}. \tag{2–3}$$

From (2–2) and the fact that $p + 2$ is prime, we deduce that $n | (pa_{n-1} + L)$, the required inductive step for (2–1). The sufficiency part of the theorem is established.

For the necessity part of the theorem, we note that if k has a proper prime divisor q then trivially $ka_{q-1} \equiv 0 \,(\mathrm{mod}\ q)$, so (2–1) fails for $n = q < k$. Now let $k = p$ be a prime and let q be a prime divisor of $p+2$. We proceed to prove that

$$a_i \equiv a_j \pmod{q} \qquad \text{for} \quad i+j = q+1, \tag{2–4}$$

using induction on j in $w < j < q$, where

$$w = \frac{q+1}{2}.$$

We now define

$$u_m \equiv \overline{w(w+m)} \pmod{q} \tag{2–5}$$

for $m \not\equiv w-1 \pmod{q}$. Then, for $m \not\equiv w \pm 1 \pmod{q}$, we find that

$$\begin{aligned} \bar{u}_m + \bar{u}_{1-m} &\equiv w(w+m) + w(w-m+1) \\ &\equiv 2w^2 + w \equiv 2w \equiv 1 \pmod{q}, \end{aligned}$$

and so, multiplying by $u_m u_{1-m}$,

$$u_m + u_{1-m} \pmod{q} \equiv u_m u_{1-m}. \tag{2–6}$$

Continuing with the proof of (2–4), we now consider

$$a_n \equiv a_{n-1} + \bar{n}(pa_{n-1} + L) \pmod{q}, \qquad 3 \le n < q,$$

since we may assume that (2–1) holds for $n < q$. Upon applying this congruence twice with $n = w+1$ and $n = w$, we find, using $p \equiv -2 \equiv -\bar{w} \pmod{q}$, that $a_{w+1} \equiv A_1 a_{w-1} + B_1 L \pmod{q}$, where

$$\begin{aligned} A_r &\equiv (1-u_r)(1-u_{1-r}) \pmod{q}, \\ B_r &\equiv w(u_{1-r} + u_r - u_{1-r}u_r) \pmod{q}, \end{aligned}$$

with u_r being given by (2–5). Thus $A_1 \equiv 1 \pmod{q}$ and $B_1 \equiv 0 \pmod{q}$ according to (2–6), so $a_{w+1} \equiv a_{w-1} \pmod{q}$. Taking as the induction hypothesis that $a_{w+i} \equiv a_{w-i} \pmod{q}$ holds for $1 \le i < r$ and making use of the iterative congruence $2r$ times, we find that $a_{w+r} \equiv A_r a_{w-r} + B_r L \pmod{q}$. The inductive step is completed using (2–6), so (2–4) is proved. In particular, we have $a_{q-1} \equiv a_2 \pmod{q}$.

Now if $\ell = 2$ and 4 then $a_2 = L(p+2) \equiv 0 \pmod{q}$, and if $\ell = 3$ then $a_2 = (3p+7)/2 \equiv \bar{2} \pmod{q}$, so $pa_2 \equiv -1 \not\equiv -L \pmod{q}$. Thus, $pa_{q-1} \not\equiv -L \pmod{q}$ for $\ell = 2, 3, 4$. Therefore, if q is a proper divisor of $p+2$ the congruence in (2–1) fails at $n = q < p$. Theorem 1.1 is proved.

3. PROOF OF THEOREM 1.2

First, it follows from (1–1) that $a_p = 2a_{p-1}$, so $\theta_p = 0$ for all ℓ.

Since $p+1$ is even, the next result $\theta_{p+1} = 2L/(p+1)$ is equivalent to $pa_{p-1} + L \equiv 0 \pmod{(p+1)/2}$, and it can be established similarly to the proof for the sufficiency part of Theorem 1.1. Thus, we find that

$$\begin{aligned} pa_{p-1} + L &= p\Big(a_{p-2} + \Big\lceil \frac{pa_{p-2}}{p-1} \Big\rceil\Big) + L \\ &= \frac{(pa_{p-2} + L)(2p-1)}{p-1} = \cdots \\ &= \frac{2p-1}{p-1}\,\frac{2p-2}{p-2} \cdots \frac{p+3}{3} \cdot (pa_2 + L) \\ &= \binom{2p}{p-1} \cdot \frac{pa_2 + L}{p(p+2)}, \end{aligned}$$

which shows that $p+1$ divides $pa_{p-1} + L$, and hence also $pa_p + 2L$. Therefore we have

$$\begin{aligned} a_{p+1} &= a_p + \frac{pa_p + 2L}{p+1}, \\ \theta_{p+1} &= \frac{2L}{p+1}, \\ \ell &= 2, 3, 4. \end{aligned} \tag{3–1}$$

The next modulus $p+2$ is a prime, and the above argument does not apply. However, we may take $q = p+2$ in (2–4) giving $a_{p-1} \equiv a_4 \pmod{p+2}$, and hence $pa_p \equiv -4a_4 \pmod{p+2}$. For even and odd ℓ, this residue has different formulae involving L. We deal with the cases $\ell = 2$ and 4 and find that $a_2 \equiv 0$, $a_3 \equiv \bar{3}L$, $a_4 \equiv \bar{4}(2 - \bar{3})L$, and hence $pa_p \equiv (\bar{3} - 2)L$. From (3–1) we now have $pa_{p+1} \equiv (\bar{3} - 2)L + 2 \times \bar{3}L \equiv -L$ and we have established that

$$\begin{aligned} a_{p+2} &= a_{p+1} + \frac{pa_{p+1} + L}{p+2}, \\ \theta_{p+2} &= \frac{L}{p+2}, \\ \ell &= 2, 4. \end{aligned} \tag{3–2}$$

The next two moduli, namely $p+3$ and $p+4$, are both composite, the former being even and the latter being a multiple of 3. From the established results for a_p, a_{p+1}, a_{p+2}, we find that

$$\begin{aligned} a_{p+3} &= a_{p+2} + \frac{pa_{p+2} + 5L}{p+3}, & a_{p+4} &= \frac{2(p+2)a_{p+3}}{p+4}, \\ \theta_{p+3} &= \frac{5L}{p+3}, & \theta_{p+4} &= 0, \end{aligned} \tag{3–3}$$

for $\ell = 2, 4$. Thus, we have

$$\begin{aligned} pa_{p+2} + 5L &= p\Big(\frac{(2p+2)a_{p+1} + L}{p+2}\Big) + 5L \\ &\equiv -12a_{p+1} + 8L \\ &\equiv -30a_p + 20L \\ &\equiv 20(pa_{p-1} + L) \pmod{p+3}, \end{aligned}$$

and our earlier evaluation of $pa_{p-1} + L$ gives

$$\begin{aligned} &pa_{p+2} + 5L \\ &\quad\equiv 20 \cdot \frac{(2p-1)(2p-2)\cdots(p+3)}{(p-1)(p-2)\cdots 3} \cdot (pa_2 + L) \\ &\quad\equiv \binom{2p+2}{p-1} \cdot \frac{10(p+3)(pa_2+L)}{p(p+1)(2p+1)} \pmod{p+3}. \end{aligned}$$

Therefore $(p+3) | (pa_{p+2} + 5L)$ and a similar argument shows that $(p+4) | a_{p+3}$, so Equation (3–3) holds.

Now let $5 \le n < p$. First, Formula (2–2) yields

$$\begin{aligned} pa_n &= pa_{n-1} + \frac{p}{n}(pa_{n-1} + L) \\ &\equiv pa_{n-1} - (pa_{n-1} + L) \\ &\equiv -L \pmod{p+n}. \end{aligned} \tag{3–4}$$

The argument used for the proof of (2–1) shows that $pa_{p-1} \equiv -L \pmod{p+n}$, and using $\theta_p = 0$ and (3–1), we find that $pa_p \equiv pa_{p+1} \equiv -2L$. Because of the "exceptional" values for θ_{p+2} and θ_{p+3}, the residues pa_{p+2} and $pa_{p+3} \pmod{p+n}$ will depend on n. However, with the next exceptional value for θ_{p+4}, it turns out that the residue pa_{p+4} will not depend on n, and in fact $pa_{p+4} \equiv -2L \pmod{p+n}$ again. Thus, from (3–2) and (3–3) together with $p \equiv -n \pmod{p+n}$, we find that

$$\begin{aligned} pa_{p+2} &\equiv -2L + \tfrac{n}{n-2}(-L) \\ &\equiv \tfrac{3n-4}{n-2}(-L), \\ pa_{p+3} &= \tfrac{(2p+3)pa_{p+2}}{p+3} + \tfrac{5Lp}{p+3} \\ &\equiv \tfrac{2n-3}{n-3} \cdot \tfrac{3n-4}{n-2}(-L) + \tfrac{5nL}{n-3}, \\ pa_{p+4} &\equiv \tfrac{2(n-2)}{n-4} pa_{p+3}, \end{aligned}$$

where the notation $\frac{\cdots}{n-i}$ for $i = 2, 3, 4$ means the reciprocal of $n - i$ modulo $p + n$. The claim that $pa_{p+4} \equiv -2L \pmod{p+n}$ amounts to verifying

$$\frac{n-2}{n-4}\Big(\frac{2n-3}{n-3} \cdot \frac{3n-4}{n-2} - \frac{5n}{n-3}\Big) \equiv 1 \pmod{p+n},$$

which is indeed the case. An inductive argument then delivers $pa_{p+n-1} \equiv -2L \pmod{p+n}$, so $\theta_{p+n} = 2L/(p+n)$. Thus, Formula (1–6) is established.

The derivation of (1–7) differs from that of (1–6) only in the calculations for the exceptional values θ_{p+2}, θ_{p+3}, and θ_{p+4}. Theorem 1.2 is proved.

4. THE VALUES FOR θ_n IN $2p \le n \le p(p+2)$

We give the values of θ_n in the intervals $rp \le n < (r+1)p$, with $r = 2, 3, \ldots, p+1$. In each of these intervals, we may expect from (1–6), (1–7), and (1–8) that

$$\theta_n = \frac{(r+1)L}{n}, \tag{4–1}$$

apart from a few exceptional values. This is indeed the case, and we take $\ell = 2, 4$ first. Then, for $2 \le r \le (p-3)/2$ the four exceptional values are

$$\begin{aligned} \theta_{rp} &= 0, \\ \theta_{r(p+2)} &= \frac{L}{r(p+2)}, \\ \theta_{r(p+2)+1} &= \frac{(3r+2)L}{r(p+2)+1}, \\ \theta_{r(p+2)+2} &= 0; \end{aligned} \tag{4–2}$$

in particular, θ_n takes the value 0 twice in such an interval. The same result still holds when $r = (p-1)/2$, but with only the first two exceptional values in (4–2) being relevant, because $r(p+2) + 1 = (r+1)p$, so the indices for the last two values fall outside the interval.

When $r = (p+1)/2$ we find that there are also only two exceptional values, given by

$$\theta_{rp} = \frac{L}{r}, \qquad \theta_{rp+1} = \frac{L}{rp+1}, \tag{4–3}$$

so $\theta_n > 0$ in this interval. When $(p+1)/2 < r \le p$, there are four exceptional values again, and they are given by

$$\begin{aligned} \theta_{rp} &= 0, & \theta_R &= \frac{2L}{R}, \\ \theta_{R+1} &= \frac{3rL}{R+1}, & \theta_{R+2} &= \frac{L}{R+2}, \end{aligned} \tag{4–4}$$

with $R = (r-1)(p+1)$. Note that θ_n takes the value 0 only once in such intervals.

For the interval $p(p+1) \le n \le p(p+2)$, Formula (4–1) no longer applies, and we find that, for $\ell = 2$,

$$\theta_{p(p+1)} = \frac{p}{p+1}, \quad \theta_{p(p+2)} = \frac{p}{p+2}, \quad \theta_n = \frac{2}{n},$$

for $p(p+1) < n < p(p+2)$, while for $\ell = 4$,

$$\theta_{p(p+1)} = \frac{p-1}{p+1}, \quad \theta_{p(p+2)} = \frac{p-1}{p+2}, \quad \theta_n = \frac{p+4}{n}, \tag{4–5}$$

for $p(p+1) < n < p(p+2)$.

When $\ell = 3$, some of the numerators in these formulae have to be changed. The last three numerators in (4–2) and (4–4) have to be changed to $r+2, 4r+3, r+1$, and the two in (4–3) have to be changed to 1 and $(p+5)/2$, respectively. Also, in (4–5), we need to change $p-1$ to p. Note that θ_n is positive in the last given interval, and that $\theta_{p(p+1)}$ and $\theta_{p(p+2)}$ are near 1 when p is large; in fact they are the first values to exceed $\frac{1}{2}$.

The same method used in establishing (1–6) can be applied to these extended formulae, but we omit the rather tedious argument. From Figures 3 and 6 we see that although it is possible to keep track of θ_n for a few more such intervals, its values seem to become chaotic before long. In particular, we do not know any useful necessary or sufficient condition for $\theta_n = 0$ when $n > p(p+2)$.

REFERENCES

[Conway et al. 79] J. H. Conway, H. T. Croft, P. Erdős, and M. J. T. Guy. “On the Distribution of Values of Angles Determined by Coplanar Points.” *J. London Math. Soc.* 2:19 (1979), 137–143.

[Shiu 96] P. Shiu. “On a Family of Diophantine Difference Equations.” *Bull. London Math. Soc.* 28 (1996), 343–350.

Peter Shiu, Department of Mathematical Sciences, Loughborough University, Loughborough, Leicestershire LE11 3TU, United Kingdom (P.Shiu@lboro.ac.uk)

Received March 2, 2004; accepted October 6, 2004.