J. Austral. Math. Soc. (*Series A*) **57** (1994), 341–356

ON RECURSIVE SOLUTIONS OF A UNIT FRACTION EQUATION

LAWRENCE BRENTON and ROBERT R. BRUNER

(Received 24 October 1990; revised 10 March 1992)

Communicated by J. H. Loxton

Abstract

We consider the Egyptian fraction equation $1 = \sum_{i=1}^{N} 1/n_i + 1/(\prod_{i=1}^{N} n_i)$ and discuss techniques for generating solutions. By examining a quadratic recurrence relation modulo a family of primes we have found some 500 new infinite sequences of solutions. We also initiate an investigation of the randomness of the distribution of solutions, and show that there are infinitely many solutions not generated by the aforementioned technique.

1991 *Mathematics subject classification* (*Amer. Math. Soc.*): 11D68, 11N45, 11B50, 11Y50.

In [2] the first author, in joint work with Richard Hill, presented a natural correspondence between certain homeomorphism classes of complex surface singularities with perfect local fundamental group, and solutions in positive integers n_i to the Diophantine equation

$$1 = \sum_{i=1}^{N} \frac{1}{n_i} + \frac{1}{\prod_{i=1}^{N} n_i}. \tag{1}$$

Equation (1), moreover, is of interest in its own right as a special case of the problem of expressing 1 as the sum of distinct unit fractions (cf. [1, 3], for example). The best-known solutions to (1) are the initial terms of the sequence

$$2, 3, 7, 43, 1807, 3263443, \ldots, \tag{2}$$

first investigated rigorously by Sylvester [10], whose $(n+1)^{\text{st}}$ term is defined recursively by

$$x_{n+1} = x_n^2 - x_n + 1, \tag{3}$$

or, equivalently, by

$$x_{n+1} = \left(\prod_{i=1}^{n} x_i\right) + 1. \tag{4}$$

Indeed, for each N the first N terms of (2) are known to give the smallest positive value of $1 - \sum_{i=1}^{N}(1/n_i)$ among all choices of positive integers $n_1, \ldots, n_N$ ([5, 7]). In this paper we will exploit properties of sequences defined by (3) or (4) modulo certain primes P to obtain new solutions to (1), and thus to generate new families of isolated singular points of complex surfaces with interesting topological properties. The reader is referred to [2] for a discussion of the relevant complex geometry and for the list of all solutions to (1) of length $N \leq 7$. (We take this opportunity to correct a misprint in the list on page 65 of [2]: in the last line 6020772531 should be 6020372531.) We thank the referee for suggesting several corrections and improvements to the original manuscript.

Minimal Solutions

We seek all solutions to the equation (1), which hereafter we will write as $1 = \sum_{i=1}^{N}(1/n_i) + \prod_{i=1}^{N}(1/n_i)$, a form which is more symmetric in appearance and easier to set in type. As a first reduction we note that if $(n_1, \ldots, n_N)$ is a solution of length N, then we immediately obtain a new solution $(n_1, \ldots, n_N, n_{N+1})$ of length $N + 1$ by putting $n_{N+1} = (\prod_{i=1}^{N} n_i) + 1$. Thus every solution to (1) leads to an infinite sequence of solutions by iterations of this recursive formula. The sequence $2, 3, 7, 43, 1807, \ldots$ is generated in this way, starting with the trivial solution $1 =$ *empty sum* $+ (1/$*empty product*$)$. To find all solutions, then, it is enough to find all solutions which are *minimal* in the following sense.

DEFINITION 1. A solution $(n_1, \ldots, n_N)$, $1 < n_1 < \cdots < n_N$, to equation (1) is called *minimal* if $(n_1, \ldots, n_{N-1})$ is *not* a solution to (1).

Minimal solutions of equation (1) are of independent interest in number theory because they provide solutions to Znám's problem: find positive integers $n_1, \ldots, n_N$ such that each n_i is a proper divisor of $n_1 \cdots n_{i-1} n_{i+1} \cdots n_N + 1$ (see [4], for example). The following lemma gives a construction for a large number of minimal solutions.

LEMMA 2. *Let* $(n_1, \ldots, n_N)$ *be a solution to* (1) *and put* $\Pi = \prod_{i=1}^{N} n_i$. *Then* $(n_1, \ldots, n_N, n_{N+1}, n_{N+2})$ *is a solution to* (1) *if and only if* $n_{N+1} = \Pi + F$, $n_{N+2} = \Pi + G$, *and* $\Pi^2 + 1 = FG$. *The new solution is minimal if and only if the factorization of* $\Pi^2 + 1$ *is proper.*

The proof is a straightforward calculation (cf. [2, Proposition 12]).

https://doi.org/10.1017/S1446788700037745 Published online by Cambridge University Press

COROLLARY 3. *Let* $(n_1, \ldots, n_N)$ *be a solution to* (1). *Then the number of distinct solutions of the form* $(n_1, \ldots, n_N, n_{N+1}, n_{N+2})$, $n_{N+1} < n_{N+2}$, *is equal to half the number of divisors of* $(\Pi_{i=1}^{N} n_i)^2 + 1$.

EXAMPLE. Let us apply this result to the sequence $2, 3, 7, \ldots, x_{n-1}^2 - x_{n-1} + 1, \ldots$. For $1 \le N \le 5$ the numbers $(\Pi_{i=1}^{N} x_i)^2 + 1$ are small and easily factored. For $N = 6$ and $N = 7$ we have

$$\begin{aligned}(\Pi_{i=1}^{6} x_i)^2 + 1 &= 11342371305541119430404963 7\\ &= 841349 \cdot 2721250733 \cdot 49540355461, \qquad \text{and}\\ (\Pi_{i=1}^{7} x_i)^2 + 1 &= 12864938683278671740537145884937248491231415124195365\\ &= 5 \cdot 223681 \cdot 227693 \cdot 457822213\\ &\qquad \cdot 11034739397607023042427262095993 7.\end{aligned}$$

By applying Lemma 2 to all non-trivial factorizations of these numbers we obtain three new minimal solutions of length 8 and 15 of length 9. Taking the factorization $F = 223681 \cdot 227693 \cdot 457822213$, $G = 5 \cdot 11034739397607023042427262095993 7$ of $(\Pi_{i=1}^{7} x_i)^2 + 1$, for instance, gives the solution

$$2, 3, 7, 43, 1807, 3263443, 10650056950807,$$
$$113423736372580899460286171, 55173708330406420754320746580012 7$$

of length 9, where the last two numbers are $(\Pi_{i=1}^{7} x_i) + F$ and $(\Pi_{i=1}^{7} x_i) + G$.

For $N > 7$ the task of finding the divisors of $(\Pi_{i=1}^{N} x_i)^2 + 1$ brings us quickly to the limits of computability. For $N = 8$ we must factor a 105-digit composite. For $N = 9, 10, \ldots$, the orders of magnitude of the numbers to be factored are 10^{208}, 10^{416}, etc., which put them beyond present computational reach. Thus we need a different approach to obtain additional minimal solutions by Lemma 2. The following idea belongs to the class of 'ρ methods' (terminology due to Pollard).

For $n = 1, 2, \ldots$, define $A_n = x_{n+1} - 1 = \Pi_{i=1}^{n} x_i$, where $\{x_n\}$ is the sequence $2, 3, 7, \ldots, x_{n-1}^2 - x_{n-1} + 1, \ldots$. Notice, then, that we have the recursive formula $A_{n+1} = A_n^2 + A_n$, with $A_0 = 1$. Fix an odd prime P, and consider the sequence $\{A_n\} \bmod P$. Since A_{n+1} is determined by a function of A_n alone, there exist unique smallest indices $n_0 < n_0 + \lambda \le P$ such that $A_{n_0+\lambda} \equiv A_{n_0} \bmod P$, and then $A_{n+\lambda} \equiv A_n \bmod P$ for all $n \ge n_0$. The positive integer λ is the *period* of $\{A_n\} \bmod P$.

PROPOSITION 4. *Let* P *be an odd prime. Then* $A_N^2 + 1 \equiv 0 \bmod P$ *for some* N *if and only if the sequence* $\{A_n\}$ *has period* $2 \bmod P$. *This can occur only if* $P \equiv 1 \bmod 4$. *Except for* $P = 5$, *which divides* $A_n^2 + 1$ *for all odd* n, $A_n^2 + 1 \equiv 0 \bmod P$ *for at most one* n.

https://doi.org/10.1017/S1446788700037745 Published online by Cambridge University Press

PROOF. If $A_N^2 + 1 \equiv 0 \bmod P$, then -1 is a square mod P, so $P \equiv 1 \bmod 4$. If $P \equiv 1 \bmod 4$, let i denote a square root of $-1 \bmod P$. Then $A_N \equiv \pm i$ implies that $A_{N+1} = A_N^2 + A_N \equiv -1 \pm i$, $A_{N+2} = A_{N+1}^2 + A_{N+1} \equiv -1 \mp i$, and $A_{N+3} = A_{N+2}^2 + A_{N+2} \equiv -1 \pm i \equiv A_{N+1}$. Thus $\{A_n\}$ has period $2 \bmod P$ (for clearly $A_{N+3} \not\equiv A_{N+2}$ in this case).

Conversely, if $\{A_n\}$ has period 2 we have two cases. If $6 = A_2 \equiv A_0 = 1 \bmod P$, then $P = 5$. It is easy to check by induction that for all N, $A_N \equiv 2 \bmod 5$ if N is odd and $A_N \equiv 1 \bmod 5$ if N is even. Hence 5 divides $A_N^2 + 1$ for all odd N. If $P \neq 5$, let $N \geq 0$ be the smallest of the indices n for which $A_{n+3} \equiv A_{n+1} \bmod P$. Then A_N satisfies

$$[(A_N^2 + A_N)^2 + A_N^2 + A_N]^2 + (A_N^2 + A_N)^2 + A_N^2 + A_N \equiv A_N^2 + A_N,$$

or $(A_N^2 + A_N)^2[(A_N^2 + A_N + 1)^2 + 1] \equiv 0$. Since $A_N^2 + A_N = A_{N+1} \not\equiv 0$ (lest $A_{N+2} = A_{N+1}^2 + A_{N+1} \equiv 0$ and $\{A_n\}$ have period $1 \bmod P$), we have

$$0 \equiv (A_N^2 + A_N + 1)^2 + 1 \equiv (A_N - i)(A_N + i)(A_N + 1 - i)(A_N + 1 + i),$$

where i is a square root of -1. Since Z_P is a field we conclude that $A_N \equiv \pm i$ or $A_N \equiv -1 \pm i$. But A_N cannot be congruent to $-1 \pm i$ lest $A_{N+2} \equiv A_N$, contradicting minimality of N. Thus $A_N^2 \equiv -1 \bmod P$ and P divides $A_N^2 + 1$.

For uniqueness, suppose again that $P \neq 5$, that P divides $A_N^2 + 1$, and that $N \geq 0$ is the smallest index with this property. Then for all $m > N$, $A_{m+2} \equiv A_m$ and so as above $A_m \equiv -1 \pm i$, which are not congruent to $\pm i$ when $P \neq 5$. Thus $A_m^2 + 1 \not\equiv 0 \bmod P$, so A_N is the unique term of the sequence $\{A_n\}$ for which P divides $A_n^2 + 1$.

We now have an easy method for checking whether or not a given prime P is a factor of $A_N^2 + 1$ for any N, and hence for finding many minimal solutions to our equation (1) by applying Lemma 2. That is, we compute the sequence $\{A_n\} \bmod P$ by the recursion formula $A_{n+1} = A_n^2 + A_n$, $n = 1, 2, \ldots$, until we get a repetition $A_{N+1+\lambda} \equiv A_{N+1}$. If $\lambda = 2$ then P divides $A_N^2 + 1$ and we obtain the solution

$$(5) \qquad (x_1, \ldots, x_N,\ \Pi_{i=1}^N x_i + P,\ \Pi_{i=1}^N x_i + (\Pi_{i=1}^N x_i^2 + 1)/P).$$

This solution is minimal except when P actually equals $(\Pi_{i=1}^N x_i)^2 + 1$, as is the case for $P = 5 = x_1^2 + 1$ and $P = 37 = (x_1 x_2)^2 + 1$, for instance. On the other hand, for the special prime $P = 5$, (5) provides a solution for every odd index N, and the solution is minimal except for $N = 1$. This result alone guarantees that equation (1) has infinitely many minimal solutions, a fact exploited by Sun [9] in connection with Znám's problem.

We have carried out the required recursion calculations for all primes $P \leq 46340$. (The upper bound $46340 \approx \sqrt{2^{31}}$ was chosen because it was a convenient word size for the hardware and software available when we began our investigations. We have

https://doi.org/10.1017/S1446788700037745 Published online by Cambridge University Press

since run extended precision versions of our programs for some larger primes, with similar results.) In this range there are 53 primes P giving rise to minimal solutions of this type. The lengths $\tilde{N} = N+2$ of the solutions range from $\tilde{N} = 5$ for $P = 353$ (that is, 353 divides $(\Pi_{i=1}^{3} x_i)^2 + 1 = 1765$, giving the solution $(2, 3, 7, 42 + 353, 42 + 5)$) to $\tilde{N} = 476$ for the prime 45737. Since the numbers X_N are between $2^{2^{N-2}}$ and $2^{2^{N-1}}$, this last mentioned minimal solution involves numbers which are very large indeed. The list of primes P for which $\{A_n\}$ has period 2 mod P are available from the authors on request, as are printouts of all of the calculations reported in this paper.

REMARK. The numbers $x_n = (\Pi_{i=1}^{n-1} x_i) + 1$ of the sequence $2, 3, 7, \ldots$, are candidates for large primes, since they are of the form 'the product of many small numbers, plus 1'. The same techniques as in Proposition 4 can be used to determined whether a given prime P is a factor of any x_n. Indeed, these methods are in the spirit of certain standard primality tests, such as the Lucas-Lehmer test for primality of the Mersenne numbers $2^P - 1$.

PROPOSITION 5. (Sylvester) *A prime P divides x_N for some N if and only if the sequence $\{A_n\}$ has period 1. Except for $P = 2$ and $P = 3$, this can occur only for $P \equiv 1 \bmod 3$. Each P divides at most one x_N.*

PROOF. If P divides x_N, then surely P divides $A_N = \Pi_{i=1}^{N} x_i$. Then $A_{N+1} = A_N^2 + A_N \equiv 0 \equiv A_N \bmod P$, and $\{A_n\}$ has period 1. Conversely, if $\{A_n\}$ has period 1 mod P, suppose that N is the smallest of the indices n for which $A_{n+1} \equiv A_n \bmod P$. That is, $A_N^2 + A_N \equiv A_N$, whence $A_N^2 = (\Pi_{i=1}^{N} x_i)^2 \equiv 0$, and P divides some x_n.

Now $A_0 = 1$ and $A_1 = 2$, so if $P > 3$ and N is as above, then N is not 0 or 1. Thus A_{N-1} and A_{N-2} are well-defined and P does not divide either of them. But then

$$0 \equiv A_N = A_{N-1}(A_{N-1} + 1) = A_{N-1}(A_{N-2}^2 + A_{N-2} + 1).$$

Since $A_{N-1} \not\equiv 0 \bmod P$ we have $A_{N-2}^2 + A_{N-2} + 1 \equiv 0$. This equation has solutions mod P if and only if $P \equiv 1 \bmod 3$. (The solutions are $(-1 \pm \sqrt{-3})/2$, and -3 is a square mod P exactly when $P \equiv 1 \bmod 3$ by quadratic reciprocity.)

Finally, it is clear from the relation $x_{n+1} = (\Pi_{i=1}^{n} x_i) + 1$ that P can divide at most one x_n.

Again, we have compiled the list of all primes $P \leq 46340$ which properly divide some member of the sequence $\{x_n\}$ (there are 47 of them). As a curiosity we discovered that *all* of the factors of $x_7 = 10650056950807$ happen to be smaller than this bound, and thus we obtained the complete factorization $x_7 = 547 \cdot 607 \cdot 1033 \cdot 31051$.

https://doi.org/10.1017/S1446788700037745 Published online by Cambridge University Press

Generalizations

From the proofs of the foregoing theorems it is clear that if instead of the sequence $2, 3, 7, 43, \ldots$ we choose any sequence $n_1, \ldots, n_K, n_{K+1}, \ldots$ with $1 = \sum_{i=1}^{K}(1/n_i) + \prod_{i=1}^{K}(1/n_i)$ and with $n_m = (\prod_{i=1}^{m-1} n_i) + 1$ for all $m > K$, then the same analysis goes through, and further minimal solutions to (1) are obtained. To make this precise, consider again the recursive formula $X_{n+1} = X_n^2 + X_n$ with given initial value X_0. For any positive integer P we have a relation 'succeeds' which is defined on the set Z_P of congruence classes mod P by 'y succeeds x if $y = x^2 + x$'. This relation induces the structure of a directed graph, denoted L_P, on the ring Z_P. Each connected component of L_P terminates in a cycle of some period λ, which we will call a λ-*loop*. For fixed λ we use the symbol $L_P(\lambda)$ to denote the union of the connected components of L_P whose terminal loops have length λ. For instance, for $P = 13$ the directed graph L_P is as follows:

$$\begin{array}{ccccccccccccccc} & & 11 & & 10 & & & & 9 & & & & 5 & & 8 \\ & & \downarrow & & \downarrow & & & & \downarrow & & & & \downarrow & & \downarrow \\ 1 & \to & 2 & \to & 6 & \Rightarrow & 3 & \to & 12 & \to & 0 & \hookleftarrow & 4 & \leftrightarrow & 7 \end{array}$$

Thus L_{13} has only 2 components, $L_{13}(1)$ and $L_{13}(2)$. Note that exactly $(P+1)/2$ of the elements y of Z_P are successors, and that if y succeeds x then y also succeeds $\hat{x} = -(x+1)$ (and y succeeds only x and $\hat{x}$). This reflects the fact that for exactly $(P+1)/2$ elements y the equation $X^2 + X = y$ admits a solution in Z_P. If a solution x exists, then there are exactly two solutions, x and $\hat{x}$. (The double arrow $x \Rightarrow y$ means that x is a double root of the polynomial $X^2 + X - y$; that is, that $x = \hat{x}$ in Z_P.) If P is not prime, then of course the directed graph L_P may be more complicated.

Some of the elementary properties of L_P for general primes P are as follows.

REMARK 1. For every prime P, L_P contains a unique 1-loop, namely $0 \hookleftarrow$. If $P \equiv 1 \bmod 3$, then the 1-component contains at least 4 elements, namely

$$\begin{array}{ccccc} & & \omega^2 & & \\ & & \downarrow & & \\ \omega & \to & -1 & \to 0 & \hookleftarrow \end{array}$$

where ω is a primitive cube root of 1. If $P \equiv 2 \bmod 3$ then $L_P(1)$ contains only the two elements $-1 \to 0 \hookleftarrow$.

REMARK 2. If $P \equiv 3 \bmod 4$, then L_P has no 2-loop. If $P \equiv 1 \bmod 4$ (but $P \neq 5$), then the 2-component contains at least the 4 elements

$$i \to 1 + i \leftrightarrow -1 - i \leftarrow -i$$

https://doi.org/10.1017/S1446788700037745 Published online by Cambridge University Press

where i is a square root of -1. This subset comprises the entire 2-component if and only if neither $1+4i$ nor $1-4i$ is a square mod P. If $P \equiv \pm 3, \pm 5, \pm 6$, or $\pm 7 \bmod 17$, then exactly one of these two numbers is a square, and in this case the 2-component has at least 6 elements.

The proofs of Remark 1 and the first two assertions of Remark 2 are identical to the proofs of Propositions 5 and 4 above. As for the last two claims of Remark 2, if the 2-component contains more than the 4 elements pictured, then (and only then) there is an element x in Z_P for which $x^2+x = \pm i$. (Note that the equations $X^2+X = -1 \pm i$ have only the solutions $x = \pm i$, $x = -1 \mp i$ already accounted for.) In any field of characteristic $\neq 2$ the solutions to this quadratic equation are

$$x = (-1 \pm \sqrt{1 \pm 4i})/2, \tag{6}$$

provided this makes sense, and there are no other solutions. Thus the 2-loop contains a fifth element x if and only if either $1+4i$ or $1-4i$ is a square mod P.

Let $P > 5$ be a prime $\equiv 1 \bmod 4$, and fix a square root i of $-1 \bmod P$. Let $(\ )$ denote the Legendre symbol. We have by quadratic reciprocity

$$\left(\frac{1+4i}{P}\right)\left(\frac{1-4i}{P}\right) = \left(\frac{17}{P}\right) = \left(\frac{P}{17}\right),$$

which is equal to 1 if $P \equiv \pm 1, \pm 2, \pm 4$, or $\pm 8 \bmod 17$, and to -1 if $P \equiv \pm 3, \pm 5, \pm 6$, or $\pm 7 \bmod 17$. In the second case we must have that exactly one of $1+4i$, $1-4i$ is a square mod P, and so the 4-component has at least two more elements of the form (6). (The two roots are distinct for all $P \neq 17$).

If $\left(\frac{P}{17}\right) = 1$ then we can conclude only that either both or neither of $1+4i$ and $1-4i$ is a square mod P. We do not know a nice criterion for distinguishing the two cases, that is, for deciding when the fourth degree polynomial $(X^2+X)^2+1$ has 0 roots or 4 roots mod P. For instance, both 13 and 149 are congruent to 1 mod 4 and to $-4 \bmod 17$, but this polynomial has 0 roots in Z_{13} and 4 roots (18, 23, 125, and 130) in Z_{149}.

REMARK 3. In general, for given λ the λ-loops consist precisely of those elements x in Z_P for which $f^\lambda(x) - x = 0$, where $f(X) = X^2+X$, but for which $f^\mu(x) - x \neq 0$ for all μ dividing λ. For fixed λ the polynomial $f^\lambda(X)$ can be calculated explicitly. For instance, taking $\lambda = 3$ we have that x is in the 3-loop mod P if and only if

$$g_3(x) = (f^3(x) - x)/x^2 = x^6 + 4x^5 + 8x^4 + 10x^3 + 9x^2 + 6x + 3 = 0$$

mod P. For given P, $g_3(X)$ will have either 0, 3, or 6 roots. Consequently, $L_P(3)$ will either be empty, have 1 component, or have 2 components.

https://doi.org/10.1017/S1446788700037745 Published online by Cambridge University Press

For example, take $P = 41$. Then $g_3(X)$ admits the irreducible factors $X - 1$, $X - 2$, $X - 6$, and $X^3 + 13X^2 - 18X + 10$. Thus L_{41} has only one 3-loop, namely $1 \to 2 \to 6 \to 1$. Indeed, the entire 3-component $L_{41}(3)$ is

$$\begin{array}{ccccccc} 34 & \to & 1 & & & & \\ & & \downarrow & \nwarrow & & & \\ 39 & \to & 2 & \to & 6 & \leftarrow & 38 \end{array}$$

and every other element of Z_{41} is either in $L_{41}(1)$, $L_{41}(2)$, or $L_{41}(5)$. By Remark 1 (since $41 \equiv 2 \bmod 3$), $L_{41}(1)$ has only two elements, $40 \to 0 \hookleftarrow$. By Remark 2 (since $41 \equiv 1 \bmod 4$ and $\equiv 7 \bmod 17$), $L_{41}(2)$ contains at least 6 elements, including $i \to 1 + i \leftrightarrow -1 - i \leftarrow -i$. In fact, the entire 2-component is

$$\begin{array}{ccccccccccc} & & 23 & & 21 & & 29 & & & & 32 \\ & & \downarrow & & \downarrow & & \downarrow & & & & \downarrow \\ 17 & \to & 19 & \to & 11 & \to & 9 & \to & 8 & \leftrightarrow & 31 \end{array}$$

and all other elements of Z_{41} are in the (single) 5-component.

REMARK 4. In general, for any integer x and any period λ we can in principle determine the primes P for which x is in the λ-loop mod P. That is, we compute the integers $f^{\lambda}(x) - x$ and $f^{\mu}(x) - x$ for all μ dividing λ. Then x is in the λ-loop mod P exactly when P divides $f^{\lambda}(x) - x$ but P does not divide $f^{\mu}(x) - x$ for any μ.

For example, for $g_3(X) = X^6 + 4X^5 + 8X^4 + 10X^3 + 9X^2 + 6X + 3$ as above, $g_3(2) = 451 = 11 \cdot 41$, so 2 is in the 3-loop mod 11 and mod 41, and for no other primes P. Therefore also 6, the unique successor of 2 mod any prime $P > 6$, must also be in the 3-loop mod 11 and mod 41. But $g_3(6) = 90651 = 11 \cdot 41 \cdot 3 \cdot 67$, so 6 is in the 3-loop mod 67 as well.

We have now the following generalizations of the results of the first section.

PROPOSITION 6. *Let P be an odd prime, let X_0 be a positive integer, and define a sequence $\{X_n\}$ of integers by $X_{n+1} = X_n^2 + X_n$, $n = 0, 1, \ldots$. Then P divides X_N for some N if and only if $X_0 \in L_P(1)$. If $P \neq 5$ then P divides $(X_N)^2 + 1$ for some N if and only if $X_0 \in L_P(2)$, but X_0 is not in the 2-loop* mod P. *(For $P = 5$, P divides some $(X_N)^2 + 1$ if and only if $X_0 \neq 0$ or -1* mod 5.*)*

The proof is the same as for Propositions 5 and 4 above.

COROLLARY 7. *Let $n_1, \ldots, n_K$ satisfy $1 = \sum_{i=1}^{K}(1/n_i) + \prod_{i=1}^{K}(1/n_i)$. Put $X_0 = \prod_{i=1}^{K} n_i$, $X_{n+1} = X_n^2 + X_n$. Let P be an odd prime such that X_0 is in $L_P(2)$ but not in the 2-loop* mod P, *and let $N \geq 0$ be the smallest integer for which $X_{N+1} \equiv X_{N+3}$* mod P. *Then $\{n_1, \ldots, n_K, X_0 + 1, \ldots, X_{N-1} + 1, X_N + P, X_N + (X_N^2 + 1)/P\}$ is a solution to* (1) *of length $K + N + 2$, minimal except when $P = X_N^2 + 1$.*

https://doi.org/10.1017/S1446788700037745 Published online by Cambridge University Press

We have examined, by computer, all primes $P < 46340$ for all choices of initial solution $(n_1, \ldots, n_K)$ with $K \leq 6$. The lists of the primes for which $\Pi_{i=1}^{K} n_i \in L_P(2)$ are available from the authors. To pick an example at random to illustrate the technique, take $P = 10289$, $X_0 = 2 \cdot 3 \cdot 11 \cdot 23 \cdot 31 = 47058$. Then the sequence $X_0 \to X_1 \to \cdots \bmod P$ is

$$47058 \equiv 5902 \to 952 \to 1824 \to 5453 \to 5452 \leftrightarrow 4835.$$

Thus P divides $X_3^2 + 1$ and we obtain a minimal solution of length $5 + 3 + 2 = 10$, namely

2, 3, 11, 23, 31, 47059, 2214502423, 4904020979258368507,
2404942176500620759344455001215105083 1,
562129154661438714443373800466448749170113437215080075864044184347734 27,

where the last two numbers are $X_3 + 10289$ and $X_3 + (X_3^2 + 1)/10289$. In all, we achieved a total of 522 new minimal solutions to (1) of length ≥ 8 for primes in this range. To each such solution there corresponds a unique (up to homeomorphism) complex surface singularity whose dual intersection graph is a star and which is locally the cone on a homology 3-sphere, as described in [2, part 1]. For the sake of completeness of this article we have adjoined (see Appendix below) a brief outline of the construction.

On the randomness of the sequence $X_{n+1} = X_n^2 + X_n$

Lacking a parameterization of the solutions to the equation $1 = \sum_{i=1}^{N}(1/n_i) + \Pi_{i=1}^{N}(1/n_i)$, we can at least seek an asymptotic formula for the number $S(N)$ of solutions for fixed N. For instance, the results of [2] show that $S(1) = S(2) = S(3) = S(4) = 1$, $S(5) = 3$, $S(6) = 8$, and $S(7) = 26$. The techniques of the present paper give the recursive lower bound

$$S(N+2) \geq \sum_{(n_1,\ldots,n_N)\in\Gamma_N} \frac{1}{2}\tau(\Pi_{i=1}^{N} n_i^2 + 1),$$

where Γ_N is the set of all solutions of length N and where τ counts the number of divisors. By Corollary 7 this is related to the behavior of the sequence $X_{n+1} = X_n^2 + X_n \bmod$ various primes P. If for many primes the sequence behaves, in an appropriate sense, 'randomly', we might thus be able to apply probabilistic methods to predict the number of solutions for large N. In this section we offer some empirical results bearing on this question.

Apart from these considerations, the question of recursive generation of random numbers is an important theme in the application of computer technology to problems

of interest in number theory. Since the phrase 'computer generated random numbers' seems almost oxymoronic, much effort has gone into the development of satisfactory statistical tests for apparent randomness, the most elementary being the χ^2 test for relative frequencies. A standard 'random number' generating device is a recursive relation $X_{n+1} = f(X_n)$ mod a large prime P. For instance, if P is taken to be the largest prime which a computer word can hold, then the linear relation $X_{n+1} = AX_n + B$, for appropriate choice of A near $\sqrt{P}$ and for suitable B, is often quite successful in satisfying various randomness criteria.

The next most obvious class of functions to try are the quadratic functions. In particular, the relation $X_{n+1} = X_n^2 + X_n$ was first investigated for randomness by Coveyou, using, among other methods, the technique of the finite Fourier transform. (This sequence is also related to the 'middle square' method of von Neumann. See [8, Chapter 3] for an extended discussion of these ideas.) A very simple randomness test that is appropriate to the problems under view in this paper is as follows. Choose an initial value X_0 and define X_{n+1} recursively as above. For each prime P in some fixed range M_1 to M_2, calculate the length $l(P)$ of the sequence $X_0, X_1, \ldots$ mod P before a repetition occurs. Then $l(P)$ is compared to the median of the probability distribution $f_P(k) = kP!/(P-k)!P^{k+1}$, which is the probability that $l(P)$ would equal k if the sequence were chosen randomly. Finally, a non-parametric one sample sign test based on the Gaussian statistic $z = (2Q - R)/\sqrt{R}$ is applied, where R is the number of primes P, $M_1 < P < M_2$, Q is the number of times that $l(P)$ exceeds the median, and z is the parameter of the standard normal approximation to the binomial distribution with probability (1/2). (For large P the median $m = m(P)$ of the distribution f_P can be estimated by the formula $m \approx \sqrt{2P \log P}$, but for these tests we actually calculated the median separately for each P by summing values of $f_P(k)$.)

Omitting the first few primes, which might have peculiar features, we have made the calculations required to carry out this test for all primes between 1000 and 46340, starting with various initial values X_0. There are 4624 such primes, so we expect that $l(P)$ will be greater than $m(P)$ about 2312 times, plus or minus about $\sqrt{4624}/2 = 34$. For the initial values X_0 we chose the 8 values of $\prod_{i=1}^{N} n_i$ for the minimal solutions of (1) of length ≤ 6, these being the initial values that we are especially interested in. Since this choice compromises the randomness of the results, we also ran 200 values of X_0 chosen randomly from 1 to 100,000. The data also lend themselves to randomness tests based on the statistic U = 'number of runs above or below the median', which for large randomly ordered samples of size N is approximately normally distributed with mean $(N/2) + 1$ and with standard deviation approximately $\sqrt{N}/2$.

The results of these tests are summarized in Table 2 below. First we note that for the 8 special values of X_0, 13 of the 16 values of z are within a 68% confidence interval and 15 are within a 95% confidence interval for random normal variation. For the larger sample of 200 randomly chosen values of X_0, the distribution shows

a clear bias toward negative values of z. On the average, then, over many values of (P, X_0), the sequence X_n shows a tendency to repeat somewhat sooner than expected, in comparison with random sequences. Presumably, this reflects the fact that in the directed graphs L_P each element with antecedents has precisely two antecedents, thus shortening the probable lengths of non-repeating chains. (We have run these tests several times, with different sets of 200 initial values, and achieved a similar distribution each time.)

Randomness tests of greater sophistication are possible, but because of Proposition 4 above the most interesting feature of the sequence $X_{n+1} = X_n^2 + X_n$ for our purposes is the frequency of primes P for which it has period 2 mod P. For a randomly chosen sequence modulo a fixed prime P the probability that the first repetition gives a loop of size λ is

$$h_p(\lambda) = \sum_{k=\lambda}^{P} \frac{P!}{(P-k)!P^{k+1}}.$$

Taking $\lambda = 2$ and summing over a large set $\mathscr{P}$ of primes, we expect to find a loop of length 2 in about

$$\tilde{T} = \sum_{P \in \mathscr{P}} \sum_{k=2}^{P} \frac{P!}{(P-k)!P^{k+1}}$$

cases. For $\mathscr{P}$ the collection of primes between 1000 and 46340 this sum is approximately $\tilde{T} \approx 48.7$. In practice, for the sequence under view and the initial values X_0 given by minimal solutions to the equation (1) for $N \leq 6$ the numbers T of 2-loops for primes in this range are 43, 63, 58, 62, 55, 51, 58, and 56 (listed in the order in which the numbers X_0 appear in Table 2 below). We see that except for $X_0 = 1$, $T = 43$, these numbers tend in general to be somewhat higher than the expected value $\tilde{T}$, possibly reflecting the fact that many primes $P \equiv 1 \bmod 4$ have fairly large 2-components. These results, the tendency for our sequence to repeat sooner than expected, and the tendency for 2-loops to appear more frequently than expected, give some empirical support for our hope that it may be possible to use probabilistic techniques to achieve lower bounds for the growth of the number of solutions to (1) for large N. The next section shows, however, that much work remains to be done before we can give a good estimate of the total number of solutions.

Further minimal solutions

So far we have discovered many new solutions to our equation (1) by finding factors of $(\Pi n_i)^2 + 1$ for given initial solutions $(n_1, \ldots, n_N)$. In this section we show that there are also infinitely many solutions that are *not* derived in this way.

https://doi.org/10.1017/S1446788700037745 Published online by Cambridge University Press

PROPOSITION 8. *Let P, Q be integers greater than 2 such that $(-Q)$ is a square* mod P, $(-P)$ *is a square* mod Q, *and for some choice of* $\sqrt{-P}$ mod Q, $P^2 - 4\sqrt{-P}$ *is also a square* mod Q. *Let* $(n_1, \ldots, n_N)$ *be a solution to* $1 = \sum(1/n_i) + \Pi(1/n_i)$. *Put* $Y_0 = \Pi_{i=1}^{N} n_i$, $Y_1 = Y_0 + 1$, *and for* $n > 1$, *put* $Y_n = Y_{n-1}^2 - Y_{n-1} + 1$. *Suppose that*

(a) $Y_0 \equiv \sqrt{-Q} \bmod P$, *and*

(b) $2Y_0 \equiv -P + (P^2 - 4\sqrt{-P})^{1/2} \bmod Q$ *for some choice of roots. Suppose also that for the directed graphs L_P and L_Q generated by the formula $X \to X^2 + X$,*

(c) Y_0 *is in the λ_1-loop* mod P *and the λ_2-loop* mod Q, *for some λ_1, λ_2.*

Then for each non-negative integer k, we have the solution $(n_1, \ldots, n_N, Y_1, \ldots, Y_{k\mu}, Z_1, Z_2, Z_3)$ to equation (1), *where μ is the least common multiple of λ_1 and λ_2, and where $Z_1 = (\Pi_{i=0}^{k\mu} Y_i) + P$, $Z_2 = (\Pi_{i=0}^{k\mu} Y_i) + ((\Pi_{i=0}^{k\mu} Y_i)^2 + Q)/P$, $Z_3 = (\Pi_{i=0}^{k\mu} Y_i) + ((\Pi_{i=0}^{k\mu} Y_i)^2 + R)/P$, with $R = [((\Pi_{i=0}^{k\mu} Y_i)^2 + (\Pi_{i=0}^{k\mu} Y_i)P)^2 + P]/Q$. This solution will be minimal except possibly for $k = 0$, and will not be equivalent to any solution of the type discussed in Corollary 7 above, except possibly for $k = 0$.*

PROOF. It is an elementary exercise in algebra to confirm that in any case the *rational numbers* $(n_1, \ldots, n_N, Y_1, \ldots, Y_{k\mu}, Z_1, Z_2, Z_3)$ so defined satisfy $\sum_{i=1}^{N}(1/n_i) + \sum_{j=1}^{k\mu}(1/Y_j) + \sum_{\ell=1}^{3}(1/Z_\ell) + 1/\Pi n_i \Pi Y_j \Pi Z_\ell = 1$. The various conditions on roots and congruences guarantee that these rational numbers are in fact positive integers.

EXAMPLE. Take $P = 5$, $Q = 89$. Then (a) $\sqrt{-89} = 1 \bmod 5$, and 1 is in the 2-loop $1 \to 2 \to 1$ of the sequence $X_{n+1} = X_n^2 + X_n \bmod 5$. (b) $\sqrt{-5} = 23 \bmod 89$, $\sqrt{25 - 4 \cdot 23} = 17 \bmod 89$, $(-5+17)/2 = 6 \bmod 89$, and 6 is in the 6-loop $6 \to 42 \to 26 \to 79 \to 1 \to 2 \to 6 \bmod 89$. (c) For the initial solution $1 = (1/2)+(1/3)+(1/2\cdot 3)$, $X_0 = \Pi_{i=1}^{2} n_i = 6$ is congruent to 1 mod 5 and to 6 mod 89. Thus Proposition 8 says that for any integer $N \equiv 2 \bmod 6$ we get a new solution $(n_1, \ldots, n_{N+3})$, where $(n_1, \ldots, n_N)$ are the first N terms of the sequence 2, 3, 7, 43, 1807, ..., and where $n_{N+i} = Z_i$, $i = 1, 2, 3$, are as defined in the proposition. The first such example $(N = 2)$ is $n_1 = 2$, $n_2 = 3$, $n_3 = 6 + 5 = 11$, $n_4 = 6 + (6^2 + 89)/5 = 31$, $n_5 = 6 + (6^2 + 49)/5 = 23$, where $R = 49$ is computed by $R = [(6^2 + 6 \cdot 5)^2 + 5]/89$. The second such example $(N = 8)$ is

$$2, 3, 7, 43, 1807, 3263443, 10650056950807, 113423713055421844361000443,$$
$$1286493868327867174053714599836096154665325948519581 1,$$
$$3310132946490399283969363908888783600350263054127550-$$
$$9860321224376244566157631280050551634948955117193751,$$
$$6155606810922029533350614447567345231729070644400209-$$
$$3622045671868376697641681034122987466926832652377344-$$
$$3691384989226339714126570196838605889872443504398382-$$
$$3795042059528249824555392708594985233726877457310 3.$$

https://doi.org/10.1017/S1446788700037745 Published online by Cambridge University Press

2, 3, 7, 43, 1807, 3263443,	10650057792155, 134811739261383753719
	10652778201539, 41691378583707695
	10699597306267, 2300171639909623
2, 3, 7, 43, 1823, 193667,	637617223459, 31273517203328870463055
2, 3, 7, 47, 395, 779731,	60797965264, 21743485766025360000683
	6079796526837, 6974325623477705424647
	60797953531, 410254449012081168631
	607979655287, 139119028839856004123
	607979697799, 8183472856913555659
	607979793451, 2624887933109395111
	607982046587, 154405744751990423
2, 3, 7, 47, 403, 19403,	15435513395, 8215692183434294399
	15435513463, 2456237880094942747
	15435516179, 84697872837562655
2, 3, 7, 47, 415, 8111,	6644612339, 1522443894582665279
	6644613463, 38292177286592827
	6644645747, 1320426321921983
2, 3, 7, 47, 583, 1223,	1407479807, 48317057302587443
	1468268915, 33995520959
	2202310039, 3899834875
2, 3, 7, 55, 179, 24323,	10057317287, 5949978284730273323
	10057317311, 2467064172726591731
	10057317467, 513449911932648503
	10057317967, 145121431390804003
	10057320619, 30202945461748519
	10057325347, 12523178395739983
	10057454579, 736667018400959
2, 3, 11, 23, 31, 47059,	2214502427, 980804197623275639
	2214502475, 92528699894575367
	2214502687, 18505741750517011
	2214502831, 11990273552017987
	2214504467, 2398056482005535
	2214524099, 226233749172527
	2214610807, 45248521436443
	2215070383, 8636647107907
	2217342227, 1729101023519
	2244604355, 165128325167
	2294166883, 63772955407
	2365012087, 34797266971
	2446798471, 23325584587
	2612824727, 14526193019
	3375982667, 6436718855

TABLE 1. A partial list of minimal solutions to $1 = \sum_{i=1}^{N}(1/n_i) + \Pi_{i=1}^{N}(1/n_i)$ for $N = 8$.

https://doi.org/10.1017/S1446788700037745 Published online by Cambridge University Press

X_0	$l(P)$ above median	% above median	z	Number of runs	z
1	2303	49.8%	-0.26	2304	-0.26
$2 \cdot 3 \cdot 7 \cdot 47 \cdot 395$	2324	50.3	0.35	2346	0.97
$2 \cdot 3 \cdot 11 \cdot 23 \cdot 31$	2294	49.6	-0.53	2228	-2.50
$2 \cdot 3 \cdot 7 \cdot 43 \cdot 1823 \cdot 193667$	2370	51.3	1.71	2333	0.59
$2 \cdot 3 \cdot 7 \cdot 47 \cdot 403 \cdot 19403$	2318	50.1	0.18	2326	0.38
$2 \cdot 3 \cdot 7 \cdot 47 \cdot 415 \cdot 8111$	2299	49.7	-0.38	2248	-1.91
$2 \cdot 3 \cdot 7 \cdot 47 \cdot 583 \cdot 1223$	2326	50.3	0.41	2330	0.50
$2 \cdot 3 \cdot 7 \cdot 55 \cdot 179 \cdot 24323$	2292	49.6	-0.59	2338	0.74

TABLE 2. Randomness tests for the sequence $X_{n+1} = X_n^2 + X_n \bmod P$, $1000 < P < 46340$ (4624 primes)

The third in this infinite sequence of minimal solutions involves numbers up to the order 10^{13337}. Other values of P, Q and X_0 for which this construction can be carried out are as follows (note that P, Q need not be prime): $25, 89, 6(\mu = 6)$; $41, 5, 6(\mu = 6)$; $5, 41, 42(\mu = 6)$; $361, 41, 42(\mu = 12)$; and $1457, 9329, 1806(\mu = 36)$. We do not know whether there are such triples P, Q, X_0 for infinitely many distinct P and Q.

Finally, it must be remarked that there are many solutions to (1) which still do not fall within the compass of our present techniques. For instance the solution (2, 5, 7, 11, 17, 157, 961, 4398619) of length 8 does not seem to be related in an obvious way to any solution of smaller length.

The list in Table 1 is obtained by applying Lemma 2 to all factorizations of $(\Pi_{i=1}^6 n_i)^2 + 1$, for all solutions $(n_1, \ldots, n_6)$ of length 6. In each block of the table the initial solution $n_1, \ldots, n_6$ is listed first, followed by the possible choices for n_7 and n_8. As of this writing, only one other minimal solution of length 8 is known: (2, 5, 7, 11, 17, 157, 961, 4398619). There are 26 solutions of length 8 which are not minimal, namely those of the form $(n_1, \ldots, n_7, 1 + \Pi_{i=1}^7 n_i)$ for each of the 26 solutions of length 7 recorded in [2].

Appendix

Construction of complex surface singularities from solutions to the equation $1 = \sum_{i=1}^N 1/n_i + \Pi_{i=1}^N (1/n_i)$.

Let $(n_1, \ldots, n_N)$ be a solution to this Diophantine equation, and for $i = 0, 1, \ldots, N$, let $C_i \subset L_i$ denote the Riemann sphere contained as the zero section of the unique complex line bundle L_i on $P^1(C)$ with Chern class $-n_i$. Choose points $P_1, \ldots, P_N$ on C_0, and a point Q_i on C_i for all $i \geq 1$. Let $D_1, \ldots, D_N$ be disks on C_0, centered

https://doi.org/10.1017/S1446788700037745 Published online by Cambridge University Press

z	number	percent
$-\infty$ to -3	0	.000
-3 to -2	2	.010
-2 to -1	27	.135
-1 to 0	86	.430
0 to 1	65	.325
1 to 2	19	.095
2 to 3	1	.005
3 to ∞	0	.000

TABLE 3. Distribution of z's for $l(P)$ above median for 200 random X_0's

at $P_1, \ldots, P_n$ respectively and with mutually disjoint closures, and let $\tilde{D}_1, \ldots, \tilde{D}_n$ be disks on $C_1, \ldots, C_n$ respectively centered at $Q_1, \ldots, Q_n$. Plumb these spaces together by identifying the cross sections of L_0 over D_i with the fibers of the unit disk bundle $B_i \subset L_i$ over $\tilde{D}_i$. The result is a two-dimensional complex manifold $\tilde{X}$ with the curve $C = \cup_{i=0}^{N} C_i$ embedded as the support of a negative Cartier divisor. By a theorem of Grauert [6], the topological space X obtained by collapsing C to a point $x \in X$ admits (uniquely) the structure of a normal complex space whose only singular point is x and is such that the quotient map $\pi : \tilde{X} \to X$ is a biholomorphism of $X - C$ onto $X - x$. The singular point x admits a neighborhood U whose boundary is a compact connected 3-manifold M^3 and is such that $\bar{U}$ is homeomorphic to the cone on M^3.

For this construction, the equation (1) implies that M^3 is a homology sphere — that is, that $H_1(M^3, Z) = H_2(M^3, Z) = 0$ (equivalently, that the fundamental group π_1 of M^3 is a perfect group). Indeed, in general, if $n_1, \ldots, n_N$ are positive integers satisfying $1 = \sum_{i=1}^{N}(1/n_i) + D/\Pi_{i=1}^{N} n_i$ for some positive integer D, then this construction produces a singular point for which $H_1(M, Z)$ is a finite Abelian group of order D. The complex space X is unique up to homeomorphism, but admits an $(N-3)$-parameter family of complex structures, depending on choice of the points $P_1, \ldots, P_N$.

ADDED IN PROOF. Since the writing of this paper we have discovered an additional 20 minimal solutions of length 8, making a total of 89 known solutions of this length.

References

[1] E. Barbeau, 'Expressing one as a sum of distinct reciprocals: comments and a bibliography', *Eureka (Crux Math.)* **3** (1977), 178–181.

[2] L. Brenton and R. Hill, 'On the diophantine equation $1 = \sum \frac{1}{n_i} + \frac{1}{\Pi n_i}$ and a class of homologically trivial complex surface singularities', *Pacific J. Math.* **133** (1988), 41–67.

https://doi.org/10.1017/S1446788700037745 Published online by Cambridge University Press

[3] N. Burshtein, 'On distinct unit fractions whose sum equals 1', *Discrete Math.* **5** (1973), 201–206.
[4] Z. Cao, R. Liu and L. Zhang, 'On the equation $\sum_{j=1}^{s}(1/x_j) + (1/(x_1 \cdots x_s)) = 1$ and Znám's problem', *J. Number Theory* **27** (1987), 206–211.
[5] D. Curtiss, 'On Kellogg's diophantine problem', *Amer. Math. Monthly* **29** (1922), 380–387.
[6] H. Grauert, 'Über modifikationen und exzeptionelle analytische Mengen', *Math. Ann.* **146** (1962), 331–368.
[7] D. M. Johannessen and T. V. Søhus, 'On stambrøker', *Nord. Mat. Tidskr.* **22** (1974), 103–107.
[8] D. Knuth, *The art of computer programming*, volume 2 (Addison-Wesley, Reading, 1968).
[9] Q. Sun, 'On a problem of Znám', *Sichuan Daxue Xuebao* **4** (1983), 9–11.
[10] J. Sylvester, 'On a point in the theory of vulgar fractions', *Amer. J. Math.* **3** (1880), 332–335, 388–389.

Wayne State University
Detroit
Michigan 48202
U.S.A.
brenton@math.wayne.edu

Wayne State University
Detroit
Michigan 48202
U.S.A.
rrb@math.wayne.edu