On Prime Reciprocals in the Cantor Set

Christian Salas
Open University
c/o PO Box 15236
Birmingham
United Kingdom B31 9DT
cphs2@tutor.open.ac.uk

Abstract

The middle-third Cantor set $\mathcal{C}_3$ is a fractal consisting of all the points in $[0, 1]$ which have non-terminating base-3 representations involving only the digits 0 and 2. We prove that all prime numbers $p > 3$ whose reciprocals belong to $\mathcal{C}_3$ must satisfy an equation of the form $2p+1 = 3^q$ where q is also prime. Such prime numbers have base-3 representations consisting of a contiguous sequence of 1's and are known as base-3 *repunit* primes. We also show that the reciprocals of all base-3 repunit primes must belong to $\mathcal{C}_3$. We conjecture that this characterisation is unique to the base-3 case.

1 Introduction

A prime number p is called a base-N repunit prime if it satisfies an equation of the form

$$(N-1)p + 1 = N^q \tag{1}$$

where $N \in \mathbb{N} - \{1\}$ and where q is also prime. Such primes have the property that

$$p = \frac{N^q - 1}{N - 1} = \sum_{k=1}^{q} N^{q-k} \tag{2}$$

so they can be expressed as a contiguous sequence of 1's in base N. For example, $p = 31$ satisfies (1) for $N = 2$ and $q = 5$ and can be expressed as 11111 in base 2. The term *repunit* was coined by A. H. Beiler [1] to indicate that numbers like these consist of repeated units.

More importantly for what follows, the reciprocal of any such prime is an infinite series of the form

$$\frac{1}{p} = \frac{N-1}{N^q - 1} = \sum_{k=1}^{\infty} \frac{N-1}{N^{qk}} \tag{3}$$

as can easily be verified using the usual methods for finding sums of series. Equation (3) shows that $\frac{1}{p}$ can be expressed in base N using only zeros and the digit $N-1$. This single non-zero digit will appear periodically in the base-N representation of $\frac{1}{p}$ at positions which are multiples of q.

The case $N = 2$ corresponds to the famous Mersenne primes for which there are numerous important unsolved problems and a vast literature [2]. They are sequence number A000668 in The Online Encyclopedia of Integer Sequences [3]. The literature on base-N repunit primes for $N \geq 3$ is principally concerned with computing and tabulating them for ever larger values of N and q. An example is Dubner's [4] tabulation for $2 \leq N \leq 99$ with large values of q. Relatively little is known about any peculiar mathematical properties that repunit primes in these other bases may possess.

In this paper we discuss one such property pertaining to base-3 repunit primes, i.e., those which satisfy an equation of the form $2p + 1 = 3^q$ with q prime. They are sequence number A076481 in OEIS. We show that any prime number $p > 3$ whose reciprocal is in the middle-third Cantor set $\mathcal{C}_3$ must satisfy an equation of the form $2p+1 = 3^q$. Conversely, the reciprocals of all base-3 repunit primes belong to $\mathcal{C}_3$. We conjecture that this characterisation is peculiar to the case $N = 3$ and discuss this at the end of the paper.

For easy reference in the discussion below, it is convenient to give a name to prime numbers whose reciprocals belong to $\mathcal{C}_3$. A logical one is the following:

Definition 1 (Cantor prime)**.** A Cantor prime is a prime number p such that $\frac{1}{p} \in \mathcal{C}_3$.

The following is then a succinct statement of the theorem we wish to prove.

Theorem 1. *A prime number p is a Cantor prime if and only if it satisfies an equation of the form $2p + 1 = 3^q$ where q is also prime.*

2 Proof of Theorem

In order to prove Theorem 1 it is necessary to consider the nature of $\mathcal{C}_3$ briefly. It is constructed recursively by first removing the open middle-third interval $(\frac{1}{3}, \frac{2}{3})$ from the closed unit interval $[0, 1]$. The remaining set is a union of two closed intervals $[0, \frac{1}{3}]$ and $[\frac{2}{3}, 1]$ from which we then remove the two open middle thirds $(\frac{1}{9}, \frac{2}{9})$ and $(\frac{7}{9}, \frac{8}{9})$. This leaves behind a set which is a union of four closed intervals from which we now remove the four open middle thirds, and so on. The set $\mathcal{C}_3$ consists of those points in $[0, 1]$ which are never removed when this process is continued indefinitely.

Each $x \in \mathcal{C}_3$ can be expressed in ternary form as

$$x = \sum_{k=1}^{\infty} \frac{a_k}{3^k} = 0.a_1 a_2 \ldots \tag{4}$$

where all the a_k are equal to 0 or 2. The construction of $\mathcal{C}_3$ amounts to systematically removing all the points in $[0, 1]$ which cannot be expressed in ternary form with only 0's and 2's, i.e., the removed points all have $a_k = 1$ for one or more $k \in \mathbb{N}$ [5].

The construction of the Cantor set suggests some simple conditions which a prime number must satisfy in order to be a Cantor prime. If a prime number $p > 3$ is to be a Cantor prime, the first non-zero digit a_{k_1} in the ternary expansion of $\frac{1}{p}$ must be 2. This means that for some $k_1 \in \mathbb{N}$, p must satisfy

$$\frac{2}{3^{k_1}} < \frac{1}{p} < \frac{1}{3^{k_1-1}} \tag{5}$$

or equivalently

$$3^{k_1} \in (2p, 3p) \tag{6}$$

Prime numbers for which there is no power of 3 in the interval $(2p, 3p)$, e.g., 5, 7, 17, 19, 23, 41, 43, 47, ..., can therefore be excluded immediately from further consideration. If the next non-zero digit after a_{k_1} is to be another 2 rather than a 1, it must be the case for some $k_2 \in \mathbb{N}$ that

$$\frac{2}{3^{k_1+k_2}} < \frac{1}{p} - \frac{2}{3^{k_1}} < \frac{1}{3^{k_1+k_2-1}} \tag{7}$$

or equivalently

$$3^{k_2} \in \left(\frac{2p}{3^{k_1}-2p}, \frac{3p}{3^{k_1}-2p}\right) \tag{8}$$

Thus, any prime numbers which satisfy (6) but for which there is no power of 3 in the interval $(\frac{2p}{3^{k_1}-2p}, \frac{3p}{3^{k_1}-2p})$ can again be excluded, e.g., 37, 113, 331, 337, 353, 991, 997, 1009.

Continuing in this way, the condition for the third non-zero digit to be a 2 is

$$3^{k_3} \in \left(\frac{2p}{3^{k_2}(3^{k_1}-2p)-2p}, \frac{3p}{3^{k_2}(3^{k_1}-2p)-2p}\right) \tag{9}$$

and the condition for the nth non-zero digit to be a 2 is

$$3^{k_n} \in \left(\frac{2p}{3^{k_{n-1}}(\cdots(3^{k_2}(3^{k_1}-2p)-2p)\cdots)-2p}, \frac{3p}{3^{k_{n-1}}(\cdots(3^{k_2}(3^{k_1}-2p)-2p)\cdots)-2p}\right) \tag{10}$$

The ternary expansions under consideration are all non-terminating, so at first sight it seems as if an endless sequence of tests like these would have to be applied to ensure that $a_k \neq 1$ for any $k \in \mathbb{N}$. However, this is not the case: (6) and (10) capture all the information that is required. To see this, let p be a Cantor prime and let 3^{k_1} be the smallest power of 3 that exceeds $2p$. Since p is a Cantor prime, both (6) and (10) must be satisfied for all n. Multiplying (10) through by $3^{k_1-k_n}$ we get

$$3^{k_1} \in \left(\frac{3^{k_1-k_n}\cdot 2p}{3^{k_{n-1}}(\cdots(3^{k_2}(3^{k_1}-2p)-2p)\cdots)-2p}, \frac{3^{k_1-k_n}\cdot 3p}{3^{k_{n-1}}(\cdots(3^{k_2}(3^{k_1}-2p)-2p)\cdots)-2p}\right) \tag{11}$$

Given $3^{k_1} \in (2p, 3p)$, and the fact that (11) must be consistent with this for all values of n, we must have $3^{k_1} - 2p = 1$ in (11). Since there can only be one power of 3 in $(2p, 3p)$, (8) then implies that $3^{k_1} = 3^{k_2}$, (9) implies $3^{k_1} = 3^{k_3}$ and so on, so we must have $k_1 = k_n$ for all n. Otherwise, if $3^{k_1} - 2p > 1$, it is easily seen that $3^{k_{n-1}}(\cdots(3^{k_2}(3^{k_1}-2p)-2p)\cdots)$ goes to infinity as $n \to \infty$. This is because (8) implies $3^{k_2} \geq \frac{2p}{3^{k_1}-2p} + 1$, so $3^{k_2}(3^{k_1}-2p) \geq 3^{k_1}$ with equality only if $3^{k_1} - 2p = 1$. Therefore with $3^{k_1} - 2p > 1$, (8) implies $3^{k_2}(3^{k_1} - 2p) > 3^{k_1}$, then (9) implies $3^{k_3}(3^{k_2}(3^{k_1} - 2p) - 2p) > 3^{k_2}(3^{k_1} - 2p)$, and so on. Since the numerators in (11) are bounded above by $3^{k_1} \cdot 3p$, there must be a value of n for which the interval in (11) will lie entirely to the left of $(2p, 3p)$, thus producing a contradiction between (6) and (11). It follows that we cannot have $3^{k_1} - 2p > 1$, so if p is a Cantor prime we must have $2p + 1 = 3^{k_1}$ as claimed.

We note that the primality of p plays a role in the above in that it prevents $2p$ having 3 as a factor, which would make $3^q - 2p = 1$ impossible. Since $3^q - 2p \equiv p \pmod 3$, we deduce that only prime numbers of the form $p \equiv 1 \pmod 3$ can be Cantor primes. Primality in itself is not necessary, however. The theorem also encompasses non-prime numbers of the form $x \equiv 1 \pmod 3$. An example is 4, which satisfies the equation $2x + 1 = 3^y$ with $y = 2$, and we find $\frac{1}{4} \in \mathcal{C}_3$.

Next we use a standard approach to show that q in $2p + 1 = 3^q$ must be prime if p is prime [6]. To see this, note that if $q = rs$ were composite we could obtain an algebraic factorisation of $3^q - 1$ as

$$3^q - 1 = (3^r)^s - (1)^s = (3^r - 1)(3^{(s-1)r} + 3^{(s-2)r} + \cdots + 1) \tag{12}$$

We would then have

$$p = \frac{3^q - 1}{2} = \frac{(3^r - 1)}{2}(3^{(s-1)r} + 3^{(s-2)r} + \cdots + 1) \tag{13}$$

Since $2|(3^r - 1)$, this would imply that p is composite which is a contradiction. Therefore q must be prime.

Finally we prove that if p satisfies an equation of the form $2p + 1 = 3^q$ then it must be a Cantor prime. This can be done by simply putting $N = 3$ in (1) and (3). This shows that $\frac{1}{p}$ can be expressed in base 3 using only zeros and the digit 2, which will appear periodically in the base-3 representation at positions which are multiples of q. Since only zeros and the digit 2 appear in the ternary representation of $\frac{1}{p}$, $\frac{1}{p}$ is never removed in the construction of $\mathcal{C}_3$, so p must be a Cantor prime as claimed.

3 Uniqueness of the Base-3 Case

In the case of the Mersenne primes, corresponding to $N = 2$, the first four stages in the construction of the middle-half Cantor set $\mathcal{C}_2$ involve removing the open middle-half intervals $(\frac{1}{4}, \frac{3}{4})$, $(\frac{1}{16}, \frac{3}{16})$, $(\frac{1}{64}, \frac{3}{64})$ and $(\frac{1}{256}, \frac{3}{256})$ among others. These contain the first four Mersenne-prime reciprocals $\frac{1}{3}$, $\frac{1}{7}$, $\frac{1}{31}$ and $\frac{1}{127}$ respectively, so it is clear that Mersenne primes cannot be characterised in the same way as Cantor primes. Is the characterisation likely to hold for $N > 3$? We conjecture that the answer is no.

Conjecture 1. *It is possible to characterise base-N repunit primes as primes whose reciprocals belong to $\mathcal{C}_N$ (and vice versa) only in the case $N = 3$. This characterisation does not hold for $N \neq 3$.*

Although we cannot provide a formal proof of this conjecture, we offer some counterexamples for $N > 3$ and some intuitive arguments. It is relatively easy to find counterexamples showing that there are base-N repunit primes p for $N > 3$ such that $\frac{1}{p}$ does not belong to $\mathcal{C}_N$. One such counterexample for $N = 4$ is the prime number $p = 5$. This satisfies (1) with $N = 4$ and $q = 2$, so it is a base-4 repunit prime. However, its reciprocal $\frac{1}{5}$ does not belong to the middle-quarter Cantor set. The first stage in the construction of $\mathcal{C}_4$ involves the removal of the middle-quarter interval $\left(\frac{3}{8}, \frac{5}{8}\right)$ from $[0, 1]$. The second stage involves the removal of the middle-quarter interval $\left(\frac{9}{64}, \frac{15}{64}\right)$ from $[0, \frac{3}{8}]$. The removed interval $\left(\frac{9}{64}, \frac{15}{64}\right)$ contains $\frac{1}{5}$. A counterexample for $N = 5$ is the number $p = 31$ which satisfies (1) with $N = 5$ and $q = 3$. It is therefore a base-5 repunit prime. However, its reciprocal $\frac{1}{31}$ does not belong to the middle-fifth Cantor set. It can be shown straightforwardly that the fourth stage in the construction of $\mathcal{C}_5$ involves the removal of the open interval $\left(\frac{16}{625}, \frac{24}{625}\right)$ which contains $\frac{1}{31}$.

Going in the other direction, we can argue intuitively that any primes whose reciprocals belong to $\mathcal{C}_N$ for $N > 3$ are unlikely to be base-N repunit primes. This is because, for $N > 3$, the middle-Nth Cantor set $\mathcal{C}_N$ is *not* characterised by the fact that all its elements can be represented in base N in a way that involves only zeros and one non-zero digit, $N-1$. Thus, for an arbitrary prime number p whose reciprocal is contained in $\mathcal{C}_N$, it is possible for the base-N representation of $\frac{1}{p} \in \mathcal{C}_N$ to involve non-zero digits other than $N - 1$. Therefore p will not generally satisfy (1) because (1) implies a base-N representation involving *only* zeros and the digit $N - 1$ as shown in (3).

4 Acknowledgements

I would like to thank the anonymous referees who reviewed the manuscript, and my colleagues Jim Naylor and Thomas Hibbits at the Open University who made a number of useful suggestions.

References

[1] A. Beiler, Recreations in the theory of numbers, Dover, 1964.

[2] R. K. Guy, Unsolved problems in number theory, Springer, 2004.

[3] N. J. A. Sloane, The On-Line Encyclopedia of Integer Sequences, 2009.

[4] H. Dubner, Generalized repunit primes, *Math. Comp.* **61** (1993), 927–930.

[5] B. R. Gelbaum and J. M. H. Olmsted, Counterexamples in analysis, Holden-Day, 1964.

[6] O. Ore, Number theory and its history, Dover, 1988.

2000 *Mathematics Subject Classification*: Primary 11A41.
Keywords: repunit, Cantor set.

(Concerned with sequence A076481.)