Three Cousins of Recamán's Sequence

Joseph Samuel Myers
c/o Trinity College
Cambridge
CB2 1TQ, UK
jsm@polyomino.org.uk

Richard Schroeppel
500 South Maple Dr.
Woodland Hills,
UT 84653, USA
rcs@xmission.com

Scott R. Shannon
P.O. Box 2260,
Rowville, Victoria 3178,
AUSTRALIA
scott_r_shannon@hotmail.com

N. J. A. Sloane*
The OEIS Foundation Inc.
11 South Adelaide Ave.
Highland Park, NJ 08904, USA
njasloane@gmail.com

Paul Zimmermann
INRIA Nancy – Grand Est, LORIA,
F-54600 Villers-lès-Nancy
FRANCE
Paul.Zimmermann@inria.fr

April 28, 2020

Abstract

Although 10^{230} terms of Recamán's sequence have been computed, it remains a mystery. Here three distant cousins of that sequence are described, one of which is also mysterious. (i) $\{A(n), n \geq 3\}$ is defined as follows. Start with n, and add $n+1$, $n+2$, $n+3, \ldots$, stopping after adding $n+k$ if the sum $n+(n+1)+\ldots+(n+k)$ is divisible by $n+k+1$. Then $A(n) = k$. We determine $A(n)$ and show that $A(n) \leq n^2 - 2n - 1$. (ii) $\{B(n), n \geq 1\}$ is a multiplicative analog of $\{A(n)\}$. Start with n, and successively multiply by $n+1$, $n+2, \ldots$, stopping after multiplying by $n+k$ if the product $n(n+1)\cdots(n+k)$ is divisible by $n+k+1$. Then $B(n) = k$. We conjecture that $\log^2 B(n) = (\frac{1}{2} + o(1)) \log n \log\log n$. (iii) The third sequence, $\{C(n), n \geq 1\}$, is the most interesting. Start with n, and successively concatenate the decimal digits of $n, n+1, n+2, \ldots$ until the concatenation $n\|n+1\|\ldots,\|n+k$ is divisible by $n+k+1$. Then $C(n) = k$. We have found $C(n)$ for all $n \leq 1000$ except for two cases. Some of the numbers involved are quite large. For example, $C(92) = 218128159460$, and the concatenation $92\|93\|\ldots\|(92+C(92))$ is a number with about $2 \cdot 10^{12}$ digits. We have only a probabilistic argument that $C(n)$ exists for all n.

1 Introduction

Recamán's sequence $\{R(n), n \geq 0\}$ is defined by $R(0) = 0$ and, for $n \geq 0$, $R(n) = R(n-1) - n$ if that number is positive and not already in the sequence, and otherwise $R(n) = R(n-1) + n$ (in the latter

*To whom correspondence should be addressed.

case repeated terms are permitted). Terms $R(0)$ through $R(11)$ are $0, 1, 3, 6, 2, 7, 13, 20, 12, 21, 11, 22$. The sequence was contributed by Bernardo Recamán Santos in 1991 to what is now the *On-line Encyclopedia of Integer Sequences* (or *OEIS*) [12]. The most basic question about this sequence is still unanswered: does every nonnegative integer appear? The fourth author (NJAS) and several Bell Labs colleagues developed a method for speeding up the computation of the sequence, and in 2001 Allan Wilks used it to compute the first 10^{15} terms. At that point every number below 852655 had appeared, but $852655 = 5 \cdot 31 \cdot 5501$ itself was missing. Benjamin Chaffin has continued this work, and in 2018 reached 10^{230} terms [2]. However, 852655 is still missing.

Thirty years ago it seemed like a very plausible conjecture that every number would eventually appear. Today, it is not so clear. For much more about Recamán's sequence, see entry A005132[1] in [12].

A somewhat similar situation arose in connection with the third of our new sequences, $\{C(n)\}$, discussed in Sect. 4. We have no proof that $C(n)$ exists for all n, and after reaching 10^{11} in our search for $C(44)$, we were beginning to have doubts. However, after considerably more computation using a different algorithm (described in §4.1) we were able to show that $C(44) = 2783191412912$. Similar results for other hard-to-find values of $C(n)$ have convinced us that for this problem, $C(n)$ should always exist.

In Recamán's sequence we always start by trying to subtract n from the previous term. In the three sequences discussed here, to compute $A(n)$, $B(n)$, or $C(n)$ we define an intermediate sequence which starts with n and is extended by either *adding* ($A(n)$, Section 2), *multiplying by* ($B(n)$, Section 3), or *concatenating* ($C(n)$, Section 4) $n + i$ to the ith term to get the next term.

Notation. A centered dot ($\cdot$) indicates multiplication. In Section 2, T_n denotes the triangular number $n(n+1)/2$, in Section 3 a vertical bar ($|$) means "divides" and $\nu_2(n)$ is the maximal k such that 2^k divides n, and in Section 4, $\|$ denotes decimal concatenation. Also in Section 4 we distinguish between the *number* $\alpha \bmod \gamma$ and the *congruence* $\alpha \equiv \beta \pmod{\gamma}$.

2 The additive version, $\{A(n)\}$.

To find $A(n)$, $n \geq 3$, we define an intermediate sequence $\{a_n(i), i \geq 0\}$ by starting with $a_n(0) = n$, and, for $i \geq 1$, letting $a_n(i) = a_n(i-1) + n + i$. We stop when we reach a term $a_n(k)$ which is divisible by $d = n + k + 1$, and set $A(n) = k$. In other words, if the number d that we are about to add to $a_n(k)$ actually divides $a_n(k)$, then instead of adding it we stop.

An equivalent definition is that $A(n)$ is the smallest positive integer $k = k(n)$ such that $d(n) = n + k + 1$ divides

$$a_n(k) \;=\; (k+1)n + \frac{k(k+1)}{2}\,. \tag{1}$$

If $n = 3$, for example, the sequence $\{a_3(i)\}$ is $a_3(0) = 3$, $a_3(1) = 7$, $a_3(2) = 12$, and we stop with $k = 2 = A(3)$ since 12 is divisible by $d = 3 + 2 + 1$. For $n = 4$ the sequence $\{a_4(i)\}$

[1] Six-digit numbers prefixed by A refer to entries in [12].

is $4, 9, 15, 22, 30, 39, 49, 60$, where we stop with $k = 7 = A(4)$ since $a_4(7) = 60$ is divisible by $d = 4 + 7 + 1$.

Table 1 gives the values of $A(n) = k(n)$, $d(n) = n + k(n) + 1$, $p(n) = a_n(k)$, and $q_n = p(n)/d(n)$ for $n = 3, 4, \ldots, 17$. The last column gives the values of a parameter m that will arise when we relate this problem to triples of triangular numbers. We start the table at $n = 3$, because although we can certainly define the sequence $\{a_2(i)\}$, it turns out that $a_2(i) = T_{i+2} - 1$, and it is easy to show that $T_{i+2} - 1$ is never divisible by $i + 3$. So $A(2)$ does not exist.

On the other hand, $A(n)$ exists for all $n \geq 3$. The "high-water marks" for $A(n)$ in the table at $n = 3, 4, 5, 8, 17$ suggest that $A(n) \leq (n-1)^2 - 2$, and if we take $k(n) = (n-1)^2 - 2$ for $n \geq 3$ we find from (1) that $a_n(k) = (n+1)n(n-1)(n-2)/2$, which is indeed divisible by $n + k(n) + 1 = n(n-1)$.

Table 1:

n	$A(n) = k(n)$	$d(n)$	$p(n) = a_n(k)$	$q_n = p(n)/d(n)$	m
3	2	6	12	2	3
4	7	12	60	5	6
5	14	20	180	9	10
6	3	10	30	3	6
7	6	14	70	5	8
8	47	56	1512	27	28
9	14	24	240	10	13
10	4	15	60	4	10
11	10	22	176	8	13
12	20	33	462	14	18
13	25	39	663	17	21
14	11	26	234	9	16
15	5	21	105	5	15
16	31	48	1008	21	26
17	254	272	36720	135	136

The sequences $\{A(n)\}$, $\{d(n)\}$, and $\{p(n)\}$ have now been added to [12]: $\{A(n)\}$ is A332542. However, to our surprise, the $\{q_n\}$ sequence appeared to match an existing sequence, although with a shift in subscripts. For $n \geq 2$, let $\tau(n)$ denote the smallest $k > 0$ such that

$$T_n + T_k = T_m \tag{2}$$

for some integer m. The initial values are $\tau(2) = 2$, $\tau(3) = 5$, $\tau(4) = 9, \ldots$ (A082183) and apparently agree with q_{n+1}. We will show in Theorem 4 that this is true.

The representation of numbers as sums or differences of triangular numbers is a classical subject, going back to Fermat and Gauss, and has been studied in many recent papers [1, 6, 8, 11, 13, 14, 15, 16]. However, we were unable to find Theorems 3 and 4 in the literature.

Following [8] we define a *triangular triple* to be an ordered triple of nonnegative integers $[n, k, m]$ satisfying (2). We say that a triple is *trivial* if any of n, k, m are zero.

It is easy to see that $\tau(n)$ exists, since it is straightforward to check that $[n, T_n - 1, T_n]$ is a triangular triple for $n \geq 1$. So $\tau(n) \leq T_n - 1$.

We will say exactly what all the triangular triples $[n, k, m]$ are for a given $n \geq 1$ (this is a consequence of Theorem 1), and then use this to determine $\tau(n)$ (Theorem 3). The next theorem is essentially due to Nyblom [11]. We give a proof since we will use the argument in the proof of Theorem 3.

Theorem 1. *For a given integer $S \geq 1$, all pairs of nonnegative integers m, k such that*

$$S = T_m - T_k \tag{3}$$

are obtained in a unique way by factorizing $2S$ as a product $d \cdot e$ where d is odd, and taking

$$k = \frac{\max(d, e) - \min(d, e) - 1}{2}, \tag{4}$$

$$m = \frac{\max(d, e) + \min(d, e) - 1}{2}. \tag{5}$$

Proof. From (3) we have

$$2S = m(m+1) - k(k+1) = (m-k)(m+k+1).$$

Since their sum is odd, $m - k$ and $m + k + 1$ are of opposite parity, and also $m - k < m + k + 1$. Let d be the odd integer among $m - k$ and $m + k + 1$, and e the even one. Then $m - k = \min(d, e)$, $m + k + 1 = \max(d, e)$, and solving for k and m we get (4), (5). The uniqueness follows since conversely k and m determine d and e. □

In particular, as Nyblom [11] showed, the number of pairs (m, k) such that (3) holds is equal to the number of odd divisors of $2S$.

We now take $S = T_n$. Theorem 1 gives all triangular triples $[n, k, m]$ containing n. There are always two obvious factorizations, $2T_n = 1 \cdot n(n+1)$ with $d = 1$ and $e = n(n+1)$, and $2T_n = n \cdot (n+1)$, with $\{d, e\} = \{n, n+1\}$. The first case leads to the triple $[n, T_n - 1, T_n]$ already mentioned, and the second leads to the trivial solution $[n, 0, n]$. It follows that the number of nontrivial triangular triples for a given n (see A309507) is equal to the number of odd divisors $d > 1$ of $2T_n$. This result is reminiscent of the fact that the number of primitive Pythagorean triples with an even leg $2uv$ is equal to the number of odd divisors of $2uv$ (cf. [14], A024361).

The nontrivial triangular triples $[n, k, m]$ sorted into lexicographic order are given by

$$[n, A333530(n), A333531(n)],$$

or by $[n, A198455(n), A198456(n)]$ if we impose the restriction that $k \geq n$. (Lee and Zafrullah [8] also give some tables of triangular triples.) The numbers n such that there is a triple $[n, n, m]$ are listed in A053141.

The following property and its elegant proof are due to Bradley Klee (personal communication).

Theorem 2. *If $[n, k, m]$ is a triangular triple, then*

$$n + k \ \geq \ m \, . \tag{6}$$

Equality holds if and only if $n = 0$ or $k = 0$.

Proof. If we set $x = 2n + 1$, $y = 2k + 1$, $z = 2m + 1$ then (2) becomes

$$x^2 + y^2 \ = \ z^2 + 1 \, .$$

Certainly $[x, y, z]$ is not (quite) a Pythagorean triple, but this equation does suggest using the triangle inequality, which yields

$$x + y \ \geq \ \sqrt{z^2 + 1} \ \geq \ z \, ,$$

and so

$$n + k \ \geq \ m - \frac{1}{2} \, ,$$

and (6) follows since all the quantities are integers. If equality holds in (6) then $n^2 + k^2 = m^2$ (from (2)) and so $kn = 0$. □

We can now apply Theorem 1 to determine $\tau(n)$.

Theorem 3. *For $n \geq 2$, $\tau(n)$ is obtained by choosing that odd divisor d of $n(n + 1)$ which is different from n and $n + 1$, and minimizes*

$$\left| d \ - \ \frac{n(n+1)}{d} \right| . \tag{7}$$

Then $\tau(n)$ is is the value of k given by (4) with this value of d and $e = n(n + 1)/d$.

Proof. From (4) we see that the minimal k is obtained by choosing d and e so as to minimize $\max(d, e) - \min(d, e)$. But d and e are constrained by $d \cdot e = n(n + 1)$. So we must minimize (7). Since we require $k > 0$, we must avoid $d = n$ and $d = n + 1$. □

Remark. In a few cases there is no need to do any minimization. For if n is a Mersenne prime, or if $n + 1$ is a Fermat prime, then the only odd divisor of $n(n + 1)$ apart from n or $n + 1$ is $d = 1$, and we get $\tau(n) = T_n - 1$.

We now return to our study of $\{A(n)\}$, and explain the connection with triangular triples. The agreement of q_n and $\tau(n - 1)$ is no coincidence.

Theorem 4. *For $n \geq 3$, $q_n = \tau(n - 1)$.*

Proof. Note that R is a triangular number if and only if $8R + 1$ is the square of an odd integer. Indeed, $8T_n + 1 = 4n^2 + 4n + 1 = (2n + 1)^2$. The proof of the theorem is in two parts.

(i) Given $n \geq 3$, let k denote the smallest nonnegative integer such that $d = n + k + 1$ divides

$$p \;=\; (k+1)n + \frac{k(k+1)}{2}\,.$$

Then

$$q \;=\; \frac{(k+1)n + \frac{k(k+1)}{2}}{n+k+1} \tag{8}$$

is such that $R = T_{n-1} + T_q$ is a triangular number. Indeed[2], $8R + 1 = (\alpha/d)^2$, where

$$\alpha = 2n^2 + 2kn + k^2 + n + 2k + 1 = (2n + 2k + 1)d - 2p\,,$$

which is certainly divisible by d. This proves that $\tau(n-1) \leq q_n$.

(ii) Conversely, suppose $n \geq 3$ and $q = \tau(n-1)$ is such that

$$T_{n-1} + T_q \;=\; T_m \tag{9}$$

for some integer m. For given values of n and q, (8) is a quadratic equation for k, and the unique solution with $k \geq 0$ is

$$k \;=\; -n + q - \frac{1}{2} + \frac{1}{2}\sqrt{4n^2 + 4q^2 - 4n + 4q + 1}\,.$$

Using (9) we can rewrite this as

$$k \;=\; q + m - n\,,$$

from which we get

$$p \;=\; (k+1)n + \frac{k(k+1)}{2} \;=\; \frac{(q+m+n)(q+m-n+1)}{2} \;=\; q\,(n+k+1).$$

This proves that $q_n \leq \tau(n-1)$. □

In row n of Table 1, $A(n)$ corresponds to to the triangular triple $[n-1, q_n, m]$, where m is given in the final column.

3 The multiplicative version, $\{B(n)\}$.

For the multiplicative version we replace the addition of $n + i$ in the definition of $a_n(i)$ by multiplication, keeping the stopping rule. So we define $B(n)$ for $n \geq 1$ by introducing an intermediate sequence $\{b_n(i), i \geq 0\}$ which starts with $b_n(0) = n$, and, for $i \geq 1$, satisfies $b_n(i) = b_n(i-1)\cdot(n+i)$. We stop when we reach a term $b_n(k)$ which is divisible by $d = n + k + 1$, and set $B(n) = k$. In other words, if the number d that we are about to multiply $b_n(k)$ by actually divides $b_n(k)$, then instead of multiplying by it we stop.

[2]These calculations were performed in Maple, but they can easily be verified by hand.

An equivalent definition is that $B(n)$ is the smallest positive integer $k = k(n)$ such that $d(n) = n + k + 1$ divides

$$b_n(k) \;=\; \frac{(n+k)!}{(n-1)!}. \tag{10}$$

When $n = 1$, for example, the sequence $\{b_1(i)\}$ is $1, 2, 6, 24, 120$, and we stop with $k = 4 = B(1)$ since 120 is divisible by $d = 1 + 4 + 1$. For $n = 4$, the sequence $\{b_4(i)\}$ is $4, 20, 120, 840$, and we stop with $k = 3 = B(4)$ since 840 is divisible by $d = 4 + 3 + 1$.

Table 2 gives the values of $B(n) = k(n)$, $d(n) = n + k(n) + 1$, $p(n) = b_n(k)$, and $q_n = p(n)/d(n)$ for $n = 1, 2, \ldots, 12$.

Table 2:

n	$B(n) = k(n)$	$d(n)$	$p(n) = b_n(k)$	$q_n = p(n)/d(n)$
1	4	6	120	20
2	3	6	120	20
3	2	6	60	10
4	3	8	840	105
5	4	10	15120	1512
6	5	12	332640	27720
7	4	12	55440	4620
8	3	12	7920	660
9	5	15	2162160	144144
10	4	15	240240	16016
11	6	18	98017920	5445440
12	5	18	8910720	495040

The sequences $\{B(n)\}$, $\{d(n)\}$, $\{p(n)\}$, $\{q_n\}$ have now been added to [12]: $\{B(n)\}$ is A332558. Just as in the additive version, there is a close match with an existing sequence in [12]. If we add 1 to the values of $k(n)$ we get $5, 4, 3, 4, 5, 6, \ldots$, which appears to match the entry for A061836, although the definitions are different. The older sequence, which we will denote by $\{\beta(n)\}$, has a more natural definition: $\beta(n)$ for $n \geq 0$ is defined to be the smallest integer $\kappa > 0$ such that $n + \kappa$ divides $\kappa!$.

Theorem 5. *For $n \geq 1$, $\beta(n) = B(n) + 1$.*

Proof. By definition, $B(n)$ is the smallest $k > 0$ such that

$$n + k + 1 \mid n(n+1)(n+2)\cdots(n+k)\,, \tag{11}$$

whereas $\beta(n)$ is the smallest $\kappa > 0$ such that

$$n + \kappa \mid 1 \cdot 2 \cdot 3 \cdots \kappa\,,$$

or, replacing κ by $k + 1$, the smallest k such that

$$n + k + 1 \mid 1 \cdot 2 \cdot 3 \cdots (k+1). \tag{12}$$

The ratio of the right-hand sides of (11) and (12) equals $\binom{n+k}{k+1}$ which is an integer, thus the right-hand side of (12) divides the right-hand side of (11). So the value of k defined by (11) is less than or equal to the value defined by (12). To complete the proof, it is enough to show that if $n+k+1$ divides $n(n+1)(n+2)\cdots(n+k)$ then it divides $(k+1)!$. But $n+k+1$ also divides $(\sigma+n)(\sigma+n+1)(\sigma+n+2)\cdots(\sigma+n+k)$ for any σ that is a multiple of $n+k+1$. Taking $\sigma = -(n+k+1)$, that expression becomes $(-1)^{k+1}(k+1)!$. □

We do not know of any simple formula for $B(n)$ in terms of n. The following is a weak upper bound, which at least shows that $B(n)$ always exists.

Theorem 6. *For $n \geq 3$, $B(n) \leq n-1$.*

Proof. Substituting $k = n-1$ in (10), we get $b_n(n-1) = (2n-1)!/(n-1)!$, which is divisible by $n+k+1 = 2n$ for $n \geq 3$. □

3.1 Asymptotic growth of $B(n)$.

We conjecture that as n goes to infinity,

$$B(n) = \exp\left((c+o(1))(\log n)^{1/2}(\log\log n)^{1/2}\right), \tag{13}$$

with $c = 1/\sqrt{2} = 0.7071\ldots$ In the rest of this section we sketch some arguments that support the conjecture.[3]

Since $B(n) = \beta(n) - 1$ from Theorem 5, we study the asymptotic growth of $\beta(n)$ instead. Let $\beta'(n)$ be the smallest integer $k \geq 1$ such that $n+k$ is k-smooth (i.e., it has only factors less than or equal to k). Since $k!$ is k-smooth, clearly $\beta'(n) \leq \beta(n)$. The converse is not always true: $\beta(2) = 4$ but $\beta'(2) = 2$ since $2+2$ is 2-smooth. However for large n this phenomenon becomes increasingly rare. For $10^8 \leq n < 2\cdot 10^8$, only 5.7% of the values of n are such that $\beta'(n) < \beta(n)$, and for $10^9 \leq n < 2\cdot 10^9$ the proportion decreases to 4.2%. Our first unproved assumption is that $\beta(n)$ and $\beta'(n)$ have the same asymptotic behavior, so that it suffices to study the asymptotic behavior of $\beta'(n)$.

The number $\Psi(n,k)$ of k-smooth numbers $\leq n$ is given by Dickman's ρ function:

$$\frac{\Psi(n,k)}{n} \approx \rho(u),$$

where $u = \log n/\log k$ [4, 5]. As u goes to infinity, we have [5, Eq. (1.6)]:

$$\rho(u) = u^{-u+o(u)}. \tag{14}$$

However, what we want is the *local density* $\Psi'(n,k)$ around n. This is studied in Kruppa's Ph. D. thesis [7, formula (5.6)], where it is shown that

$$\frac{\Psi'(n,k)}{n} \approx \rho(u) - \gamma\frac{\rho(u-1)}{\log n}, \tag{15}$$

[3] A proof of Equation (13) might be possible using the techniques of [9].

γ being the Euler-Mascheroni constant. In our case the local density is close to the global density. For example for $n = 10^{25} + 2554$ we have $\beta'(n) = 29972$, thus $u \approx 5.584$, which yields $\rho(u) \approx 6.7 \cdot 10^{-5}$, and $\gamma\rho(u-1)/\log n \approx 1.1 \cdot 10^{-5}$. Our second unproved assumption is that the local density is $\approx \rho(u)$ asymptotically. The expected distance between two k-smooth numbers around n being $\approx 1/\rho(u)$, the expected distance between a random n and the next k-smooth number is thus $\approx 1/(2\rho(u))$.

The above arguments, together with (14), combine to suggest that $B(n)$ is approximately equal to the solution k of the equations

$$k \;\approx\; \frac{u^u}{2}, \quad u \;=\; \frac{\log n}{\log k}\,. \tag{16}$$

We have

$$\begin{aligned} \log k \;&\approx\; u \log u \;\approx\; \frac{\log n}{\log k}\,(\log\log n - \log\log k)\,, \\ (\log k)^2 \;&\approx\; \log n\,(\log\log n - \log\log k)\,, \\ 2\log\log k \;&\approx\; \log\log n\,, \end{aligned}$$

and so

$$(\log k)^2 \;\approx\; \frac{1}{2}\log n \log\log n\,,$$

which gives (13).

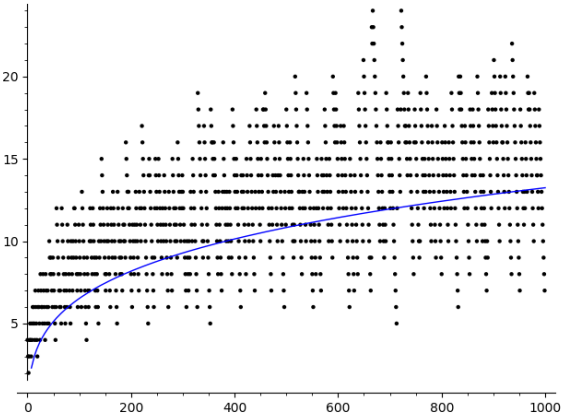

Figure 1: The first 1000 terms of $B(n)$ and (the blue line) $\overline{B(n)}$.

Let

$$\overline{B(n)} \;:=\; \exp\left(\frac{1}{\sqrt{2}}\,(\log n)^{1/2}(\log\log n)^{1/2}\right) \tag{17}$$

denote the main term on the right-hand side of (13). $\overline{B(n)}$ is a reasonably good fit to $B(n)$, even for small n. The graph in Fig. 1 shows the first 1000 terms of $B(n)$ and (the blue line) $\overline{B(n)}$. We find that $\overline{B(n)}$ is still a reasonably good fit to $B(n)$, even out to $n = 10^{30}$. Furthermore, it appears that $\overline{\lim}_{n\to\infty} B(n)$ is also given by the right-hand side of (13). It would be nice to know more about the asymptotic behavior of $B(n)$.

4 The concatenation version, $\{C(n)\}$.

For the third version we replace addition and multiplication by concatenation, but again keep the same stopping rule. We define $C(n)$ for $n \geq 1$ by introducing an intermediate sequence $\{c_n(i), i \geq 0\}$ which starts with $c_n(0) = n$, and, for $i \geq 1$, satisfies $c_n(i) = c_n(i-1)\|(n+i)$, where $r\|s$ denotes the number whose decimal expansion is the concatenation of the decimal expansions of r and s. We stop if and when we reach a term $c_n(k)$ which is divisible by $d = n + k + 1$, and set $C(n) = k$. In other words, if the number d that we are about to concatenate to $c_n(k)$ actually divides $c_n(k)$, then instead of multiplying by it we stop. In contrast to the first two versions, here we do not have a proof that such a k always exists. It is theoretically possible that the sequence $c_n(i)$ never stops, in which case we define $C(n)$ to be -1.

Table 3:

n	$C(n)$	n	$C(n)$	n	$C(n)$	n	$C(n)$
1	1	26	33172	51	2249	76	320
2	80	27	9	52	21326	77	59
3	1885	28	14	53	53	78	248
4	6838	29	317	54	98	79	31511
5	1	30	708	55	43	80	20
6	44	31	1501	56	20	81	5
7	13	32	214	57	71	82	220
8	2	33	37	58	218	83	49
9	1311	34	34	59	91	84	12
10	18	35	67	60	1282	85	25
11	197	36	270	61	277	86	22
12	20	37	19	62	56	87	105
13	53	38	20188	63	47	88	34
14	134	39	78277	64	106	89	4151
15	993	40	10738	65	1	90	1648
16	44	41	287	66	890	91	2221
17	175	42	2390	67	75	92	218128159460
18	124518	43	695	68	280	93	13
19	263	44	2783191412912	69	19619	94	376
20	26	45	3	70	148	95	23965
21	107	46	700	71	15077	96	234
22	10	47	8303	72	64	97	321
23	5	48	350	73	313	98	259110640
24	62	49	21	74	34	99	109
25	15	50	100	75	557	100	346

When $n = 1$, for example, the sequence $\{c_1(i)\}$ is $c_1(0) = 1$, $c_1(1) = 1\|2 = 12$, and we stop with $k = 1 = C(1)$ since 12 is divisible by $d = 1 + 1 + 1$.

For $n = 7$, the sequence $\{c_7(i)\}$ is

$$7, 78, 789, 78910, 7891011, \ldots, 7891011121314151617181920\,,$$

and after concatenating 20 we stop with $k = 13 = C(7)$, since the last number there, which is $c_7(13)$, is a multiple of 21.

For $n = 2$ the sequence $\{c_2(i)\}$ is $2, 23, 234, \ldots$ and stops with $k = 80 = C(2)$ at the 154-digit number

$$c_2(80) \;=\; 2345678910111213141516171819 20\ldots 69707172737475767777879808182\,,$$

which is divisible by 83.

Although a purist may be unhappy because its definition involves base 10 arithmetic,[4] we find $\{C(n)\}$ more interesting than $\{A(n)\}$ and $\{B(n)\}$ because its behavior is so erratic for such a simple rule, and there is currently no theory to explain this mixture of very small and very big numbers.

Table 3 gives the values of $C(n)$ for $n \le 100$. The values up to about $2 \cdot 10^5$ were found by straightforward direct search, but for the larger values we used the sieving algorithm described in §4.1. At the present time we have found the exact value of $C(n)$ for all $n \le 1000$ except for $n = 539$, where we only have upper and lower bounds, and $n = 158$, where we have searched up to 10^{14} without success, and it is possible that $c_{158}(i)$ does not terminate. The entry for $C(n)$ in [12], A332580, includes a table for $n \le 1000$.

Although we do not have a proof that the sequence $\{c_n(i)\}$ always terminates, the following heuristic argument suggests that it should. After k steps, we test $c_n(k)$ for divisibility by $d = n + k + 1$. There are three obvious cases when the division is impossible: (i) when $n + k + 1$ is even, since $c_n(k) \equiv n + k \pmod{10}$ is odd and cannot be divisible by an even number; (ii) when $n + k + 1$ is a multiple of 5, since then $c_n(k) \equiv 4$ or $9 \pmod{10}$ cannot be divisible by 5; (iii) when $n + k + 1$ is a multiple of 3 in the case when $n \equiv 2 \pmod 3$. Apart from this, $c_n(k)$ is essentially a very large random number.[5] The chance that $c_n(k)$ is divisible by d is roughly $1/d$, and since for a fixed n the sum $\sum_{k=1}^{\infty} 1/(n + k + 1)$ diverges, we expect one of the divisions to succeed. However, even when we try to make this argument more precise by taking into account conditions (i), (ii), (iii), the results do not convincingly explain the extremely irregular values of $C(n)$ (see Table 3), and so this argument cannot be regarded as much more than a suggestion that $C(n)$ should always exist.

4.1 A sieving algorithm for $C(n)$.

In this section we describe a sieving algorithm for $C(n)$ which we used to obtain

$$\begin{aligned} &C(44) = 2783191412912,\ C(92) = 218128159460,\ C(494) = 2314160375788,\\ &10^{14} < C(539) \le 887969738466613,\ C(761) = 615431116799,\\ &C(854) = 440578095296, \text{and } C(944) = 1032422879252. \end{aligned} \tag{18}$$

[4] There is also a base-2 version, although we will not discuss it here: see A332563.

[5] If k has j digits, $c_n(k)$ has about jk digits, and we routinely search for k up to 10^{11}.

p	p divides $n+k+1$	p divides $c_n(k)$	p divides both
3	0	2	$\emptyset$
5	0	1	$\emptyset$
7	4	$0, 2$	$\emptyset$
11	10	$0, 10$	10
17	6	$0-1, 3-16$	6

Figure 2: For $n = 44$, and certain primes p, values of $k \bmod p$ satisfying the constraints (c1) and (c2), for $100 \le n + k < 1000$ ($\emptyset$ denotes the empty set).

The idea behind the algorithm is the following. Assume $C(n) = k$, which implies that $n+k+1$ divides $c_n(k)$, and let p be a prime factor of $n + k + 1$. Then p must simultaneously satisfy two constraints: (c1) p must divide $n+k+1$, and (c2) p must divide $c_n(k)$. If we consider numbers $n+k$ that have the same number of digits, each of these constraints reduces to requiring that k is in S for some (possibly empty) set S. Consider for example $n = 44$, with $100 \le n+k < 1000$. Figure 2 shows the modular classes $k \bmod p$ that satisfy (c1), (c2), and both (c1) *and* (c2), for certain primes p. We shall see in §4.2 and §4.3 how to efficiently compute the set S given a prime p or prime power q. This leads to Algorithm 1, where the output value FAIL means that $C(n) \notin [L/10 - n, L - n - 1]$.

Algorithm 1 Sieving Algorithm for $C(n)$

Input: an integer n, and an upper bound $L = 10^\ell$
Output: either $C(n) = k$ with $L/10 \le n + k < L$, or FAIL

1: initialize an array $T[n + k] = 1$ for $n + k < L$
2: **for** each odd prime $p \le L$ **do**
3: **for** each prime power $q = p^j \le L$ **do**
4: compute the smallest k such that $n + k \ge L/10$ and q divides $n + k + 1$
5: compute $\mu := (c_n(k) \bmod q)$ [see §4.2]
6: compute a, b from L, q [see Lemma 7]
7: **while** $n + k < L$ **do**
8: **if** $\mu \equiv 0 \pmod q$ **then**
9: $T[n + k] \leftarrow p \cdot T[n + k]$
10: $k \leftarrow k + q$, $\mu \leftarrow (a\mu + b \bmod q)$
11: **for** $n + k < L$ **do**
12: **if** $T[n + k] = n + k + 1$ **then**
13: return $C(n) = k$
14: return FAIL

In line 4 of Algorithm 1, the computation of k is easy: choose $t \equiv (-1 - L/10) \pmod q$ such that $0 \le t < q$, and set $k = L/10 - n + t$. Then $n + k + 1 \equiv 0 \pmod q$.

We discuss the computation of $\mu := (c_n(k) \bmod q)$ (at line 5) in §4.2, and the computation of the constants a, b that enable us to update μ from k to $k+q$ (at line 10) in Lemma 7 (§4.3). We will use as a running example the case $n = 44$ and $L = 10^{13}$, which we used to find $C(44) = 2783191412912$.

4.2 Initial computation of $c(n)$

We consider a given 'decade' where the integers $n+k$ have exactly ℓ digits. Let $L = 10^\ell$. Consider two numbers k and $k' = k + \delta$, $\delta \geq 1$, in that decade, i.e., $L/10 \leq k < k' < L$. Then $c_n(k')$ can be computed efficiently from $c_n(k)$ as the sum of three terms x_n, y_n, and z_n (which depend on both n and k):

- $x_n(k+\delta)$ corresponds to the concatenation $n\|n+1\|\cdots\|n+k\|000\|000\|\cdots\|000$, where the string 000 stand for ℓ consecutive zeros, and there are $k' - k = \delta$ such strings. Therefore
$$x_n(k+\delta) = L^\delta c_n(k). \tag{19}$$
- The second term $y_n(k+\delta)$ corresponds to the concatenation $0\|\cdots\|0\|n+k+1\|n+k+1\|\cdots\|n+k+1$, where we have $k+1$ initial zeros, followed by δ copies of $n+k+1$, each having ℓ digits. Therefore
$$y_n(k+\delta) = (n+k+1)\frac{L^\delta - 1}{L-1}. \tag{20}$$
- The last term $z_n(k+\delta)$ corresponds to the concatenation $0\|\cdots\|0\|0\|1\|\cdots\|\delta-1$, where we have $k+2$ initial zeros, followed by $\delta - 1$ nonzero terms. Therefore
$$z_n(k+\delta) = \sum_{i=1}^{\delta-1} iL^{\delta-1-i} = \frac{L^\delta - \delta L + \delta - 1}{(L-1)^2}. \tag{21}$$

Then $c_n(k+\delta) = x_n(k+\delta) + y_n(k+\delta) + z_n(k+\delta)$. Although Equations (19), (20), (21) are valid over the integers, we will use them to compute $\mu := (c_n(k) \bmod q)$ in line 5 of Algorithm 1. It is therefore enough to evaluate (19), (20), (21) modulo q. Equation (19) reduces to exponentiation modulo q. For (20), we know that $L^\delta - 1$ is an integral multiple of $L-1$, so we may compute it modulo $q(L-1)$, divide the result by $L-1$, and then multiply by $n+k+1$. For (21) we proceed similarly, computing modulo $q(L-1)^2$.

To illustrate, suppose $n = 44$, $L = 10^{13}$, and $q = 61$. We start with $c_n(0) := 44 \bmod 61$, and use Equations (19), (20), (21) to compute $c_n(55) \bmod 61$, where $44 + 55 = 99$ is the last number of the decade $[10, 99]$. We then compute $c_n(955) \bmod 61$, where $44 + 955 = 999$ is the last number of the decade $[100, 999]$. And so on, until we reach $k = 10^{12} - 42$, for which we get $n + k + 1 \equiv 0 \pmod q$, and $c_n(k) = 54$.

4.3 Incremental computation of $c(n)$

In this section, for a prime power q, we show that we can efficiently compute $c_n(k+q) \bmod q$ from $c_n(k) \bmod q$ in the special case where $n+k+1 \equiv 0 \pmod q$, which is the one we are interested in.

Lemma 7. *Assume $L/10 \leq k < k+q < L$ for $L = 10^\ell$, where q is a prime power dividing $n+k+1$. Then $c_n(k+q) = ac_n(k) + b \bmod q$, where $a = L^q \bmod q$ and*
$$b = \frac{L^q - qL + q - 1}{(L-1)^2} \bmod q.$$

Proof. Since q divides $n+k+1$, it follows from Eq. (20) that $y_n(k+\delta) \equiv 0 \pmod q$, so $c_n(k+\delta) \equiv x_n(k+\delta) + z_n(k+\delta) \pmod q$. The rest follows by replacing δ by q in Eq. (19) and (21). □

Depending on the value of q, we may be able to further simplify the expressions for a and b. The most common case is when q is prime and coprime to $L-1$. In that case we have $a \equiv L \pmod q$, $b \equiv 1/(L-1) \pmod q$, and the values of μ at line 8 of Algorithm 1 satisfy the affine recurrence $\mu_{i+1} \equiv a\mu_i + b \pmod q$, with $\mu_0 \equiv c_n(k)$. Let $\mu'_i := \mu_i + \lambda \bmod q$, where $\lambda = b/(a-1) \equiv 1/(L-1)^2 \pmod q$. Then the sequence $\{\mu'_i\}$ satisfies the linear recurrence $\mu'_{i+1} = a\mu'_i \bmod q$, with $\mu'_0 \equiv c_n(k) + \lambda$, and we can replace the test $\mu \equiv 0 \pmod q$ by $\mu' \equiv \lambda \pmod q$. If k_0 is the value computed at line 4, and $k = k_0 + iq$ in the inner loop, the test in line 8 then becomes $a^i\mu'_0 \equiv \lambda \pmod q$, which is a discrete logarithm equation for i. We can find a first solution i_0 —if one exists—using Shanks's baby-step giant-step algorithm [3, §5.4.1], and then the other solutions are $i_0 + g, i_0 + 2g, \ldots$ where g is the multiplicative order of $a \bmod q$, which can also be efficiently computed.

It thus follows from Lemma 7 that the set S—outlined in the beginning of this section—of values of k such that p divides both $n+k+1$ and $c_n(k)$ is either empty, or is an arithmetic progression.

The correctness of Algorithm 1 follows from the fact that when we enter line 11, we have $T[n+k] = \gcd(c_n(k), n+k+1)$. Indeed, if a prime p divides both $c_n(k)$ and $n+k+1$, then for each prime power $q = p^j \le n+k+1$ dividing both $c_n(k)$ and $n+k+1$, $T[n+k]$ is multiplied by a factor p.

For our running example, with $n = 44$ and $q = 61$, we had found $k_0 := k = 10^{12} - 42$ in line 4 of Algorithm 1, and $\mu = c_n(k) = 54$ in line 5. Lemma 7 gives $a = 31$ and $b = 59$. This yields $\lambda = 4$, and the recurrence $\mu'_{i+1} \equiv 31\mu'_i \pmod q$ for $k = k_0 + iq$. The initial value is $\mu'_0 \equiv 54 + 4 \equiv 58 \pmod q$, and we are looking for $\mu'_i \equiv 4 \bmod q$. We easily find the first solution $i_0 = 34$, and since the multiplicative order of 31 modulo 61 is $g = 60$, the solutions are exactly $i = 34 + 60j$ for $j \ge 0$, which correspond to $k = k_0 + (34 + 60j)q$. Algorithm 1 will thus multiply $T[n+k]$ by q for $k = k_0 + 34q$, and then for $k_0 + 94q$, $k_0 + 154q$, ...

We conclude with a few further remarks about Algorithm 1:

- We store in $T[n+k]$ the product of prime powers dividing both $c_n(k)$ and $n+k+1$. Those products can be as large as $n+k+1$. For the record values we have obtained, this requires a 64-bit type per entry, i.e., 8 bytes. To save memory, we rely on a classical trick used in integer factorization. Let $g = \gcd(c_n(k), n+k+1)$. Instead of storing g in $T[n+k]$, we store an approximation of $\log g / \log r$ in some well-chosen radix r such that the obtained values do not exceed 255, and the array T may be implemented as a byte array.
- Instead of bounding p and q by L in lines 2 and 3, we can use a smaller bound. In that case, if we find a solution k, this will prove only that $C(n) \le k$. This is what we did for $n = 539$, taking the bound to be $5 \cdot 10^9$.
- Algorithm 1 can be trivially parallelized on t cores, with each core dealing with different primes p. However a lock is required for the accesses to the array T.

- If memory is a bottleneck, Algorithm 1 can be called several times on sub-intervals of $n+k \in [L/10, L-1]$, since the main memory usage (the sieve array T) is proportional to the sub-interval length. We used this for $n = 44,494$, and 539, with sub-intervals of length $2 \cdot 10^{12}$. And by storing only even values of $T[n+k]$ we get a factor of 2 gain.

Algorithm 1 was implemented in the C language, using the GMP library for the large integer operations, and OpenMP for the parallel code. The locks for accesses to T use the OpenMP `atomic update` instruction. On a processor with 112 hyperthreaded cores, this implementation takes about 100 seconds (of wall-clock time) to find $C(98) = 259110640$ with $L = 10^9$, and 13 hours 20 minutes to find $C(44) \le 2783191412912$ with sub-interval $[10^{12}, 2783191412958]$.

Acknowledgments

We thank Michael J. Collins, Bradley Klee, Victor S. Miller, Kerry Mitchell, and Allan C. Wechsler for helpful comments during our work on $A(n)$, David A. Corneth, Rémy Sigrist, and Jinyuan Wang for computing further terms in certain sequences arising in our study of $B(n)$, and Pierrick Gaudry for his help with the asymptotics of $B(n)$ and for independently checking some of the calculations for $C(n)$.

References

[1] D. W. Ballew and R. C. Weber, Pythagorean triples and triangular numbers, *Fibonacci Quart.*, **17:2** (1979), 168–172.

[2] B. Chaffin, E. M. Rains, N. J. A. Sloane, and A. R. Wilks, Numerical investigations of Recamán's sequence, in preparation, 2020.

[3] H. Cohen, *A Course in Computational Algebraic Number Theory*, Springer, Berlin, 1993.

[4] K. Dickman, On the frequency of numbers containing prime factors of a certain relative magnitude, *Ark. Mat. Astr. Fys.*, **22:10** (1930), 1–14.

[5] A. Granville, Smooth numbers: computational number theory and beyond, in *Algorithmic Number Theory: Lattices, Number Fields, Curves and Cryptography*, Math. Sci. Res. Inst. Publ., **44**, Cambridge Univ. Press, Cambridge, 2008, 267–323.

[6] P. W. Haggard, Pythagorean triples and sums of triangular numbers, *Internat. J. Mathematical Education in Science and Technology*, **28.1** (1997), 109–116.

[7] A. Kruppa, Speeding up Integer Multiplication and Factorization, PhD thesis, University Henri Poincaré Nancy 1, 2010, http://docnum.univ-lorraine.fr/public/SCD_T_2010_0054_KRUPPA.pdf.

[8] H. Lee and M. Zafrullah, A Note on triangular numbers, *Punjab University J. Math.*, **26** (1993), 75–83.

[9] H. W. Lenstra Jr, J. Pila, and C. Pomerance, A hyperelliptic smoothness test. I, *Phil. Trans. Roy. Soc. London, Series A*, **345:1676** (1993), 397–408.

[10] I. Niven, H. S. Zuckerman, and H. L. Montgomery, *An Introduction to the Theory of Numbers*, John Wiley & Sons, New York, Fifth Ed., 1991.

[11] M. A. Nyblom, On the representation of the integers as a difference of nonconsecutive triangular numbers, *Fibonacci Quart.*, **39:3** (2001), 256–263.

[12] The OEIS Foundation Inc., *The On-Line Encyclopedia of Integer Sequences*, https://oeis.org.

[13] W. Sierpiński, On triangular numbers which are sums of two smaller triangular numbers [Polish], *Wiadom. Mat.*, **(2) 7** (1963): 27–28; MR0182602.

[14] A. Tripathi, On Pythagorean triples containing a fixed integer, *Fibonacci Quart.*, **46/47:4** (2008/09), 331–340.

[15] M. Ulas, A note on Sierpiński's problem related to triangular numbers, *Colloq. Math.*, **117:2** (2009): 165–173.

[16] A. M. Vaidya, On representing an integer as a sum of two triangular numbers, *Vidya*, **B 15:2** (1972), 104–105.

2010 *Mathematics Subject Classification*: 11B83 (11D72, 11D85)

Keywords: Number sequences, Recamán sequence, recurrences, triangular number triples, smooth numbers.
