WHICH NUMBERS ARE NOT THE SUM PLUS THE PRODUCT OF THREE POSITIVE INTEGERS?

BRIAN CONREY & NEIL SHAH

ABSTRACT. We investigate the number $R_3(n)$ of representations as the sum plus the product of three positive integers. On average, $R_3(n)$ is $\frac{1}{2}\log^2 n$. We give an upper bound for $R_3(n)$ and an upper bound for the number of $n \leq N$ such that $R_3(n) = 0$. We conjecture that $R_3(n) = 0$ infinitely often.

1. INTRODUCTION

It is well-known that the number of representations of an integer as a product of k positive integers, denoted $\tau_k(n)$, satisfies $\tau_k(n) \ll_{\epsilon} n^{\epsilon}$ for any $\epsilon > 0$ as n goes to infinity (see Section 1.1 for notation). The proof of this standard result relies crucially on the fact that τ_k is a multiplicative function. For k = 3 this amounts to counting integral solutions of

$$n = xyz$$

The slightest perturbation of this problem turns it from an exercise to a seriously hard research question. One such example is the problem of trying to estimate the number of representations of a number n as the sum plus the product of three positive integers, i.e.

$$(1) n = xyz + x + y + z.$$

Let $R_3(n)$ denote the number of solutions to (1) in positive integers x, y, z. Here are the values of $R_3(n)$ for $1 \le n \le 12$:

0, 0, 0, 1, 0, 3, 0, 3, 3, 3, 0, 9

Our motivating problem was the following conjecture:

Conjecture 1. For any $\epsilon > 0$ there is a $C(\epsilon) > 0$ such that

$$R_3(n) \le C(\epsilon)n^{\epsilon}$$

for all positive integers n.

While we were unable to prove this conjecture, we were able to find and prove a number of other interesting results as well as formulate some new conjectures.

Date: February 2, 2022.

This work is partially supported by an NSF grant.

1.1. Notation. We use some notation that is standard in analytic number theory but perhaps not standard to everybody. In this section, we define this notation for the reader.

Suppose f(n) and g(n) are functions defined on the positive integers. Then, f(n) = O(g(n)) means that there exists some M > 0 and N_0 (which could depend on M) such that $|f(n)| \leq Mg(n)$ for all $n \geq N_0$. Also, $f(n) = \Omega(g(n))$ means that there exists some M > 0 and a sequence $n_1 < n_2 < \ldots$ such that $f(n_j) \geq Mg(n_j)$ for all j. Additionally, $f(n) \ll g(n)$ is equivalent to f(n) = O(g(n)) and $f(n) \gg g(n)$ is equivalent to $g(n) \ll f(n)$. Note that \ll_a, \gg_a , and Ω_a indicate that the implicit constant M is dependent on a.

Lastly, suppose that f(n) and g(n) tend to ∞ with n. Then, $f(n) \sim g(n)$ means that $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1$.

2. Statement of Results

We can prove Conjecture 1 on average.

Theorem 1. As N goes to infinity,

$$\frac{1}{N}\sum_{n\leq N}R_3(n)\sim \frac{1}{2}\log^2 N.$$

For an upper bound, we can show the following.

Theorem 2. For any $\epsilon > 0$,

$$R_3(n) \ll_{\epsilon} n^{1/3} \log n (\log \log n)^4.$$

In the course of proving an upper bound for $R_3(n)$, we discovered some interesting properties of $R_3(n)$ which we state here.

Proposition 1. If $R_3(n) = 0$ for n > 2, then n is prime and 2n - 3 is prime.

Proof. Let

$$f_3(x, y, z) = xyz + x + y + z.$$

Observe that $f_3(x, y, 1) = (x+1)(y+1)$. Thus, if n is composite, there exists a triple (x, y, 1) such that $f_3(x, y, 1) = n$. Thus, n is prime. The second assortion is obviously true for n = 3. If n > 3 and 2n - 3 is composite, then

$$2n - 3 = (2x + 1)(2y + 1)$$

for some positive x, y. This implies

$$n = 2xy + x + y + 2,$$

which is a solution with z = 2, which means that $f_3(x, y, 2) = n$. Thus, 2n - 3 is prime. \Box

WHICH NUMBERS ARE NOT THE SUM PLUS THE PRODUCT OF THREE POSITIVE INTEGERS? 3

This result led us to wonder how many primes $p \leq x$ there are for which $R_3(p) = 0$. Define

$$U_3(N) = \#\{n \le N : R_3(n) = 0\}.$$

Using the large sieve we can show the following.

Theorem 3. There exists some constant c > 0 such that

$$U_3(N) \ll \frac{N}{e^{c\sqrt{\log N}}}$$

as N goes to infinity.

We expect that this estimate is far from optimal. We did calculations for N up to the 250 millionth prime (5336500537) and found 2014 prime numbers which are not the sum plus the product of three positive integers.

Based on numerical evidence, we do believe that $U_3(N) \to \infty$ as $N \to \infty$ but very slowly, maybe like a power of log N.

Conjecture 2. There exist positive constants A < B such that

$$\log^A N \ll U_3(N) \ll \log^B N.$$

It is interesting to compare this problem with the analogous problem in four variables. Let $R_4(n)$ denote the number of positive integral solutions to

$$n = xyzw + x + y + z + w$$

and let

$$U_4(N) = \#\{n \le N : R_4(n) = 0\}.$$

Conjecture 3. For any $\epsilon > 0$,

 $R_4(n) \ll_{\epsilon} n^{\epsilon}.$

Again, we can prove this conjecture on average.

Theorem 4. As N goes to infinity,

$$\frac{1}{N}\sum_{n\leq x}R_4(n)\sim \frac{1}{6}\log^3 N.$$

We can also obtain an upper bound of $R_4(n)$.

Theorem 5. For any $\epsilon > 0$,

$$R_4(n) \ll_{\epsilon} n^{\frac{1}{2}+\epsilon}.$$

By the large sieve inequality we can show the following.

Theorem 6. For any C > 0 we have

$$U_4(N) \ll \frac{N}{e^{C\sqrt{\log N}}}$$

as N goes to infinity.

Unlike the situation for $R_3(n)$, we believe that only a finite number of n have $R_4(n) = 0$. Conjecture 4. If n > 45752, then

$$R_4(n) > 0.$$

2.1. **Omega results.** If $n \ge 5$, by consideration of solutions where at least one of x, y, z is equal to 1, it is easy to see that

$$R_3(n) \ge 6\tau(n) - 6$$

where $\tau(n) = \tau_2(n)$ is the number of divisors of n. Using what we know about how large $\tau(n)$ can get, it follows that for every $\epsilon > 0$,

(2)
$$R_3(n) = \Omega_{\epsilon} \left(\exp(\frac{\log(2-\epsilon)\log n}{\log\log n}) \right).$$

One could also consider a solution where at least one of x, y, z is equal to 2, in which case we can see

$$R_3(n) \ge 6\tau(n) + 6\tau(2n-3) + O(1).$$

It is not clear to us how big $\tau(n) + \tau(2n-3)$ can be. One could do similarly for 3, 4, ... and obtain apparently stronger omega results. In particular, let

$$\tau(n;q,a) = \sum_{\substack{d \mid n \\ d \equiv a \bmod q}} 1.$$

By (3) (below), we see that

$$R_3(n) \ge \sum_{1 \le x \le n^{\frac{1}{3}}} \max(\tau(xn - (x^2 - 1); x, 1) - 2, 0)$$

Note Proposition 1 is an immediate corollary of this.

We wonder whether

$$R_3(n) = \Omega_{\epsilon} \left(\exp(\frac{\log(3-\epsilon)\log n}{\log\log n}) \right),\,$$

which would be the case if $R_3(n)$ behaves like $\tau_3(n)$.

2.2. Algorithm for calculating $R_3(n)$. In order to find all the $n \le 5 \times 10^9$ with $R_3(n) = 0$, we need a reasonable algorithm for determining when $R_3(n) > 0$. Suppose that xyz + x + y + z = n and that $x \le y \le z$. Multiplying through by x and adding 1 to both sides yields (3) $(xy+1)(xz+1) = nx - x^2 + 1$.

Therefore, if we can find the proper divisors d of $nx - x^2 + 1$ such that $d \equiv 1 \pmod{x}$, then from this set of divisors we can easily determine the complete set of solutions of $f_3(x, y, z) = n$. If there are no proper divisors $d \equiv 1 \mod x$, then $R_3(n) = 0$. The probabilistic time complexity for prime factorization of a number m is $O(\exp((\log m)^{\frac{1}{2}+\epsilon}))$. We apply this to $m = nx - x^2 + 1$, for all $x \leq n^{\frac{1}{3}}$ and so for each n, there are $n^{\frac{1}{3}+\epsilon}$ steps to determine the value of $R_3(n)$.

For the primes p up to 5336500537, we checked using this method whether or not $R_3(p) = 0$. We computed all 2014 primes up to 5336500537 such that $R_3(p) = 0$:

2, 3, 5, 7, 11, 13, 17, 23, 31, 37, 41, 43, 53, 67, 71, 83, 97, 101, 107, 113,..., 5178563387, 5220047297, 5284333573, 5322410117

See this link for the full list.

2.3. Algorithm for calculating $R_4(n)$. Let

$$f_4(x, y, z, w) = xyzw + x + y + z + w.$$

Suppose $f_4(x, y, z, w) = n$ and $x \leq y \leq z \leq w$. Multiplying through by xy and rearranging yields

$$(xyz+1)(xyw+1) = nxy + 1 - x^2y - xy^2.$$

So, to find $R_4(n)$, it is a matter of finding proper divisors d of $nxy + 1 - x^2y - xy^2$ such that $d \equiv 1 \pmod{xy}$. This involves $\ll n^{\frac{1}{2}} \log n$ factorizations of numbers that are less than $n^{\frac{3}{2}}$. So, the total time to calculate $R_4(n)$ is $O(n^{\frac{1}{2}+\epsilon})$.

Since the number of divisors of $nxy + 1 - x^2y - xy^2$ is $O((nxy + 1 - x^2y - xy^2)^{\epsilon}) = O(n^{\epsilon})$, this argument also proves Theorem 5.

Next, note that

$$f_4(x, y, z, 1) = f_3(x, y, z) + 1$$

Therefore, if $R_3(n) > 0$, then $R_4(n+1) > 0$. Thus, to verify that $R_4(n) > 0$ for $45752 < n \le 5336500538$, it suffices to check that $R_4(p+1) > 0$ for the 2014 values of $p \le 5336500537$ for which $R_3(p) = 0$. For each of these, $R_4(p+1) > 0$.

We believe that the complete list of numbers n for which $R_4(n) = 0$ is:

 $1, 2, 3, 4, 6, 8, 12, 14, 18, 32, 38, 44, 54, \\68, 102, 108, 182, 192, 194, 224, 252, 374, 422, \\432, 908, 1092, 1202, 1278, 2468, 2768, \\3182, 4508, 7208, 16104, 21998, 26348, 45752$

BRIAN CONREY & NEIL SHAH

2.4. Residue Classes. The expression $f_3(x, y, z) = xyz + x + y + z$ is also equal to

$$z(xy+1) + x + y.$$

As a result, every positive integer $n \equiv x + y \pmod{xy + 1}$ with n > xy + 1 has $R_3(n) > 0$. For example, if x = 2 and y = 2, we see that every n > 5 such that $n \equiv 4 \pmod{5}$ has $R_3(n) > 0$. Similarly, if $n \equiv 5 \pmod{7}$, if $n \equiv 7 \pmod{11}$ or if $n \equiv 7$ or 8 (mod 13), then $R_3(n) > 0$. In general, there are

(4)
$$f_3(p) = \frac{1}{2}(\tau(p-1)-2)^1$$

residue classes modulo any odd prime p such that if n is congruent to one of these residue classes mod p, then $R_3(n) > 0$.

Consequently, to give a bound on the number of $n \leq N$ for which $R_3(n) = 0$, we can count how many $n \leq N$ are not in any of these residue classes. This idea suggests using the large sieve to estimate $U_3(n)$, which we will do in the proof of Theorem 3.

Now, we do the same thing for $f_4(x, y, z, w) = xyzw + x + y + z + w$. Note that

$$xyzw + x + y + z + w = w(xyz + 1) + x + y + z$$

Just as for $f_3(x, y, z)$, this shows every positive integer $n \equiv x + y + z \pmod{xyz+1}$ with n > xyz + 1 has $R_4(n) > 0$. For example, if x = 2, y = 2, and z = 3, we see that every n > 9 such that $n \equiv 7 \pmod{13}$ has $R_4(n) > 0$. In general, there are

(5)
$$f_4(p) = \frac{1}{6}(\tau_3(p-1) - 3)$$

residue classes modulo any odd prime p such that if n is congruent to one of these residue classes mod p, then $R_4(n) > 0$. We use this again in the proof of Theorem 6.

3. Proofs

3.1. Proofs of Theorems 1 and 4.

Proof. Observe that we can split the sum into two parts:

$$\sum_{n \le N} R_3(n) = \sum_{xyz + x + y + z \le N} 1 = \sum_{xyz + x + y + z \le \frac{N}{\log N}} 1 + \sum_{\frac{N}{\log N} < xyz + x + y + z < N} 1$$

For the second part,

$$\sum_{\frac{N}{\log N} < xyz + x + y + z < N} 1 = \sum_{xy < N} \sum_{\frac{N}{\log N} - x - y \\ \frac{xy < N}{xy + 1} < z < \frac{N - x - y}{xy + 1}} 1$$

 ${}^{1}f_{3}(p)$ and $f_{3}(x, y, z)$ denote different functions. The same is true for $f_{4}(p)$ and $f_{4}(x, y, z, w)$.

$$N \gg xyz \gg xy \frac{N}{\log N}$$

Therefore,

$$x + y \ll xy \ll \log N.$$

Thus,

$$\sum_{\substack{xy < N}} \sum_{\frac{N}{\log N} - x - y \\ xy + 1} < z < \frac{N - x - y}{xy + 1}} 1 \ll \sum_{\substack{xyz \le 2N \\ xy \ll \log N}} 1 \ll \sum_{\substack{xy \ll \log N}} \frac{N}{xy} \ll N (\log \log N)^2$$

Therefore, the terms with $x+y+z \gg \frac{N}{\log N}$ are negligible, and we can assume that $x+y+z \ll \frac{N}{\log N}$. Then, we have

$$\sum_{xyz+x+y+z < N} 1 = \sum_{xyz < N+O(\frac{N}{\log N})} 1 = \sum_{n < N+O(\frac{N}{\log N})} \tau_3(n) \sim \frac{N \log^2 N}{2}$$

(see 12.1.4 of [T]), which proves Theorem 1.

The proof of Theorem 4 is similar.

3.2. Proof of Theorem 2.

Proof. If n = xyz + x + y + z, then $nx - x^2 + 1 = (xy + 1)(xz + 1)$. Therefore,

$$R_3(n) \le 6 \sum_{x \le n^{\frac{1}{3}}} \tau(nx - x^2 + 1).$$

We now use the following estimate, which is a consequence of Henriot ([H1] and [H2]):

$$\sum_{x < n^{\frac{1}{3}}} \tau(nx - x^2 + 1) \ll n^{\frac{1}{3}} \log n \prod_{p \mid n^2 + 4} (1 + 4/p).$$

Therefore,

$$R_3(n) \ll n^{\frac{1}{3}} \log n \prod_{p|n^2+4} (1+1/p)^4.$$

By the prime number theorem,

$$\prod_{p|n^2+4} (1+1/p) \le \prod_{p\le 3\log n} (1+1/p).$$

And by Merten's Theorem,

 $\prod_{p \le 3 \log n} (1 + 1/p) \ll \log \log n$

and the result follows.

3.3. Proof of Theorem 3.

Proof. Define

$$a_n = \begin{cases} 0 & \text{if } R_3(n) > 0\\ 1 & \text{if } R_3(n) = 0 \end{cases}$$

and define $Z = \sum_{n=1}^{N} a_n$. Note that $Z = U_3(N)$. By the discussion in section 1.4 and the large sieve inequality (see Theorem 7.11 of [O]), we have that

(6)
$$Z \le \frac{(N^{\frac{1}{2}} + X)^2}{Q}$$

where X is a free parameter and

$$Q = \sum_{q \le X} \mu^2(q) \prod_{p|q} \frac{f_3(p)}{p - f_3(p)},$$

where μ denotes the Möbius function and $f_3(p)$ is defined as in (4). We now prove the following lemma:

Lemma 1. Let

$$f_3(p) = \frac{\tau(p-1) - 2}{2}$$

for a prime p > 3. Then, there exists a c > 0 such that as $X \to \infty$, we have

$$Q := \sum_{\substack{q \le X \\ (q,6)=1}} \mu^2(q) \prod_{p|q} \frac{f_3(p)}{p - f_3(p)} \gg \exp\left(c\sqrt{\log X}\right).$$

Proof. First of all

$$Q \ge \sum_{q \le X} \mu^2(q) \prod_{p|q} \frac{f_3(p)}{p}.$$

Let $g_3(p) = \tau(p-1)$. Then for p > 3 we have $f_3(p) \ge \frac{g_3(p)}{6}$. Therefore, for any k, by restricting only to squarefree integers q with k distinct prime factors, we have

$$Q \ge 6^{-k} \prod_{\substack{p_1 \cdot p_2 \cdots p_k \le X\\ 3 < p_1 < p_2 < \cdots < p_k}} \frac{g_3(p_1) \dots g_3(p_k)}{p_1 \dots p_k}$$

Now the Titchmarsh divisor problem (as found in Theorem 3.9 of [HR]) asserts that

$$\sum_{3$$

$$\sum_{X^a$$

Therefore,

$$Q \ge 6^{-k} \sum_{3 < p_1 < X^{\frac{1}{k^2}}} \sum_{X^{\frac{1}{k^2} < p_2 < X^{\frac{2}{k^2}}}} \cdots \sum_{X^{\frac{k-1}{k^2} < p_k < X^{\frac{k}{k^2}}}} \frac{g_3(p_1) \dots g_3(p_k)}{p_1 \dots p_k} \gg (\frac{C}{6k^2} \log X)^k$$

for $X \ge 10$ and any $k < \sqrt{\frac{\log X}{\log 3}}$. Now we choose $k = b\sqrt{\log X}$

with a sufficiently small b and have

$$Q \gg \exp\left(\sqrt{\log X} \left(b \log \frac{C}{6b^2}\right)\right) \gg \exp(c\sqrt{\log X})$$

for some small c > 0 as claimed.

Using the result of Lemma 1 in (6) and choosing $X = N^{\frac{1}{2}}$, we obtain the bound

$$U_3(N) \ll \frac{N}{e^{c\sqrt{\log N}}}.$$

3.4. **Proof of Theorem 6.** The proof is similar to the proof of Theorem 3 except that we use (1)

$$\sum_{0$$

and $f_4(p)$ as defined in (5). We leave the details to the reader.

4. CONCLUSION AND OPEN QUESTIONS

Our proofs for bounds on $U_3(N)$ and $U_4(N)$ are sieve problems with two different numbers of residue classes. $U_3(N)$ has $c \log p$ residue classes per prime p on average that get sieved out and $U_4(N)$ has $c \log^2 p$ residue classes per prime p on average that get sieved out. It is interesting to note that $U_3(N)$ appears to go to infinity and $U_4(N)$ appears to be bounded.

Another problem which bears similarity is expressing n as the sum of three positive integer cubes. Similar to the proof of Theorem 2, one can obtain $r_3(n) \ll n^{\frac{1}{3}} \log n (\log \log n)^a$ for some a. By Hooley's paucity results, one can obtain $r_3(n) \ll n^{\frac{1}{3}}$. Hardy and Littlewood's [HL] Hypothesis K was that $r_3(n) \ll n^{\epsilon}$. However, in 1936, Mahler found a parametric family of solutions for n's which are perfect twelfth powers and proved that $r_3(n) = \Omega(n^{\frac{1}{12}})$.

BRIAN CONREY & NEIL SHAH

Here are some related problems we think are worth exploring further:

- (1) Prove Conjecture 1, that $R_3(n)$ grows slower than n^{ϵ} for any $\epsilon > 0$.
- (2) Prove Conjecture 4.
- (3) Prove that there are infinitely many positive integers n for which $R_3(n) = 0$.
- (4) Does $R_4(n)$ go to ∞ with n?
- (5) Give any improvement on the omega result in (2).
- (6) Give omega results for $\tau(n) + \tau(n+1)$ and $\tau(n^2+1)$.

We hope that this work stimulates the reader to pursue some of these questions further.

5. Acknowledgements

After posting the first version of this paper to the arXiv, we had a number of email exchanges. We would like to thank Olivier Bordelles, James Maynard, Bill McEachen, Tomás Oliviera e Silva, Gerald Tenenbaum, and Trevor Wooley for their helpful comments. We would also like to thank Michel Marcus for adding sequence A350535 to the OEIS.

References

- [E] Erdös, P. On the sum $\sum_{k=1}^{x} d(f(k))$. London Math. Soc. 27 (1952), 7-15.
- [H1] Henriot, K. Nair-Tenenbaum bounds uniform with respect to the discriminant. Math. Proc. Cambridge Philos. Soc. 152 (2012), no. 3, 405–424.
- [H2] Henriot, K. Nair-Tenenbaum uniform with respect to the discriminant—Erratum [MR2911138]. Math. Proc. Cambridge Philos. Soc. 157 (2014), no. 2, 375–377.
- [HL] Hardy, G. H.; Littlewood, J. E. Some problems of 'Partitio numerorum' (VI): Further researches in Waring's Problem. Math. Z. 23 (1925), no. 1, 1–37.
- [HR] Halberstam, H.; Richert, H.-E. Sieve methods. London Mathematical Society Monographs, No. 4. Academic Press [Harcourt Brace Jovanovich, Publishers], London-New York, 1974. xiv+364 pp.
- [M] Mahler, Kurt Note on Hypothesis K of Hardy and Littlewood. J. London Math. Soc. 11 (1936), no. 2, 136–138.
- [O] Odlyzko, Andrew Michael SIEVE METHODS. Thesis (B.S.) California Institute of Technology. 1971. 146 pages.
- [S] Shiu, P. A Brun-Titchmarsh theorem for multiplicative functions. J. Reine Angew. Math. 313 (1980), 161–170.
- [Sh] Shevelev, V. Representation of positive integers by the form $x_1 \dots x_k + x_1 + \dots + x_k$. arXiv:1508.03970
- [T] Titchmarsh, E. C. The theory of the Riemann zeta-function. Second edition. Edited and with a preface by D. R. Heath-Brown. The Clarendon Press, Oxford University Press, New York, 1986. x+412 pp.

American Institute of Mathematics, San Jose, CA 95112, USA

Email address: conrey@aimath.org

MORGAN HILL, CA 95037, USA

Email address: neilhemang@gmail.com

10