Integer Complexity: Experimental and Analytical Results

Jānis Iraids[1], Kaspars Balodis[2], Juris Čerņenoks[2], Mārtiņš Opmanis[1], Rihards Opmanis[1], and Kārlis Podnieks[1]

[1] Institute of Mathematics and Computer Science, University of Latvia, Raiņa bulvāris 29, Riga, LV-1459, Latvia
[2] University of Latvia, Raiņa bulvāris 19, Riga, LV-1586, Latvia

Abstract. We consider representing of natural numbers by arithmetical expressions using ones, addition, multiplication and parentheses. The (integer) complexity of n – denoted by $\|n\|$ – is defined as the number of ones in the shortest expressions representing n. We arrive here very soon at the problems that are easy to formulate, but (it seems) extremely hard to solve. In this paper we represent our attempts to explore the field by means of experimental mathematics. Having computed the values of $\|n\|$ up to 10^{12} we present our observations. One of them (if true) implies that there is an infinite number of Sophie Germain primes, and even that there is an infinite number of Cunningham chains of length 4 (at least). We prove also some analytical results about integer complexity.

1 Introduction

The field explored in this paper is represented most famously in F26 of Guy [1], and as the sequence A005245 in "The On-Line Encyclopedia of Integer Sequences"[2].

We consider representing of natural numbers by arithmetical expressions using ones, addition, multiplication and parentheses. Let's call this "representing numbers in basis $\{1, +, \cdot\}$". For example,

$$2 = 1 + 1;$$
$$3 = 1 + 1 + 1;$$
$$4 = 1 + 1 + 1 + 1 = 2 \cdot 2 = (1 + 1) \cdot (1 + 1);$$
$$5 = 1 + 1 + 1 + 1 + 1 = 1 + 2 \cdot 2 = 1 + (1 + 1) \cdot (1 + 1);$$
$$6 = 1 + 1 + 1 + 1 + 1 + 1 = 2 \cdot 3 = (1 + 1) \cdot (1 + 1 + 1);$$
$$7 = 1 + 1 + 1 + 1 + 1 + 1 + 1 = 1 + 2 \cdot 3 = 1 + (1 + 1) \cdot (1 + 1 + 1);$$
$$8 = 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 =$$
$$= 2 + 2 \cdot 3 = 1 + 1 + (1 + 1) \cdot (1 + 1 + 1) =$$
$$= 2 \cdot 2 \cdot 2 = (1 + 1) \cdot (1 + 1) \cdot (1 + 1);$$
$$9 = 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 = 3 + 2 \cdot 3 = 1 + 2 \cdot 2 \cdot 2 =$$

$$= 3 \cdot 3 = (1 + 1 + 1) \cdot (1 + 1 + 1);$$
$$10 = 1 + 1 + \ldots + 1 = 2 \cdot 2 + 2 \cdot 3 = 2 + 2 \cdot 2 \cdot 2 =$$
$$= 2 \cdot 5 = (1 + 1) \cdot (1 + 1 + 1 + 1 + 1);$$
$$11 = \ldots = 2 + 3 \cdot 3 = 1 + 2 \cdot 5;$$
$$12 = \ldots = 2 \cdot 2 \cdot 3;$$
$$13 = \ldots = 2 \cdot 2 + 3 \cdot 3 = 3 + 2 \cdot 5;$$
$$14 = \ldots = 2 \cdot (1 + 2 \cdot 3);$$
$$\ldots$$

As we see, most numbers can be represented in several ways that may differ in size. For example, the number 8 is represented above by three different expressions containing 8, 7 and 6 ones respectively.

We will measure the size of an expression by the number of ones it contains. We don't need counting neither of operations (if an expression contains k ones, then it contains k–1 operations), nor of parentheses (the postfix notation might be used).

The size of the shortest expressions representing a particular number n can be considered as the "complexity" of n. Hence, the term "integer complexity". Some numbers allow for several shortest expressions (for examples, see above: 4, 5, 10, 11, 13).

Definition 1. *Let's denote by $\|n\|$ the number of ones in the shortest expressions representing n in basis $\{1, +, \cdot\}$. We will call it the **integer complexity** of n.*

For example, as we see above:

$\|1\| = 1;$	$\|5\| = 5;$	$\|9\| = 6;$	$\|13\| = 8;$
$\|2\| = 2;$	$\|6\| = 5;$	$\|10\| = 7;$	$\|14\| = 8;$
$\|3\| = 3;$	$\|7\| = 6;$	$\|11\| = 8;$	$\|15\| = 8;$
$\|4\| = 4;$	$\|8\| = 6;$	$\|12\| = 7;$	$\ldots$

This definition corresponds to the sequence A005245 in "The On-Line Encyclopedia of Integer Sequences". In [1], $\|n\|$ is denoted by $f(n)$, the notation $\|n\|$ is due to Arias de Reyna [3].

In a similar fashion, representation of natural numbers in other bases, for example, $\{1, +, \cdot, -\}$, $\{1, +, \cdot, \uparrow\}$ and $\{1, +, \cdot, -, \uparrow\}$ could also be considered (sequences A091333, A025280 and A091334 [4,5,6], $\uparrow$ stands for exponentiation).

As a function of n, in average, $\|n\|$ is growing logarithmically, namely, one can prove easily:

Theorem 1. *For all $n > 1$,*

$$3 \log_3 n \le \|n\| \le 3 \log_2 n \approx 4.755 \log_3 n$$

In [1], Guy attributes this result to Dan Coppersmith.

The lower bound of Theorem 1 is reached by infinitely many numbers n, exactly – by the powers of three. For example,

$$3^4 = 81 = (1+1+1) \cdot (1+1+1) \cdot (1+1+1) \cdot (1+1+1),$$

and, indeed, products of $1+1+1$s is the best way of representing powers of three in basis $\{1, +, \cdot\}$:

Theorem 2. $\|n\| = 3\log_3 n$, if and only if $n = 3^b$ for some $b \geq 1$. In particular, for all $b \geq 1$, $\|3^b\| = 3b$ (moreover, the product of $1+1+1$s is shorter than any other representation of 3^b).

Similarly, the product of $1+1$s seems to be the best way of representing powers of two in basis $\{1, +, \cdot\}$. For example,

$$2^5 = 32 = (1+1) \cdot (1+1) \cdot (1+1) \cdot (1+1) \cdot (1+1).$$

Hypothesis 1. For all $a \geq 1$, $\|2^a\| = 2a$ (moreover, the product of $1+1$s is shorter than any other representation of 2^a).

$\|2^a\| = 2a$ is true for all powers $2^a < 10^{12}$, i.e. for all a, $0 < a \leq 39$ – as verified by Jānis Iraids. We consider proving or disproving of Hypothesis 1 as one of the biggest challenges of number theory.

The upper bound of Theorem 1 doesn't seem to be exact. As observed by Rawsthorne [7], the "worst" value of $\frac{\|n\|}{\log_3 n}$ seems to be ≈ 3.928 (and not 4.755 of Theorem 1), and it is reached only by a single number, namely, by 1439 ($\|1439\| = 26$):

Hypothesis 2. For all $n > 1$,

$$\|n\| \leq \frac{\|1439\|}{\log_3 1439} \log_3 n \approx 3.928 \log_3 n$$

Hypothesis 2 is true for all $n \leq 10^{12}$ – as verified by Jānis Iraids.

Thus, we arrive here very soon at the problems that are easy to formulate, but (it seems) extremely hard to solve. In this paper we represent our attempts to explore the field by means of experimental mathematics. We managed to prove analytically only a few of the experimental observations.

In Section 2 we explain the basic concepts and their simplest properties. Section 3 represents our analytical results. Section 4 considers algorithms allowing to calculate the values of $\|n\|$. The best of them was used by Jānis Iraids to calculate $\|n\|$ up to $n = 10^{12}$. In Section 5 we present our experimental observations (several confirmed and refuted hypotheses included).

Wolfram—Alpha (by Wolfram Alpha LLC, 2009) was used for some of the computations in this paper.

2   Basic concepts and related work

2.1   The largest and smallest numbers of complexity n

For a given n, there exists only a finite number of expressions of size $\leq n$. Hence,

Definition 2. *a) Let's denote by $E(n)$ the **largest** m such that $\|m\| = n$.*
*b) Let's denote by $E_k(n)$ the k-**th largest** m such that $\|m\| \leq n$ (if it exists).*
  Thus, $E(n) = E_0(n)$.
*c) Let's denote by $e(n)$ the **smallest** m such that $\|m\| = n$.*

In this definition, $E(n)$ corresponds to the sequence A000792[8], and $e(n)$ – to A005520[9].

Proposition 1. *As a function of n, $E(n)$ is monotonically increasing.*

Proof. Take one of the shortest expressions for $E(x)$ and append $+1$ to it. Now, it contains $x + 1$ ones, and its value is $E(x) + 1$. Therefore the greatest number of complexity $x + 1$ is no less than $E(x) + 1$. $\qquad\square$

Proposition 2. *As a function of n, $e(n)$ is monotonically increasing.*

Proof. Assume that there exists n, such that $e(n) < e(n-1)$. Take the smallest such n. Then it would mean that before $e(n)$ there are no numbers of complexity $\geq n$, in particular, $\|e(n)\| - \|e(n) - 1\| > 1$. But it is impossible since $e(n)$ can be written as $e(n) + 1$. $\qquad\square$

Proposition 3. *For all $n \geq 1$, $n \leq E(\|n\|)$.*

Proof. By definition – $E(y)$ is the greatest number whose complexity is y. Thus for all n, if $\|n\| = y$ then $n \leq E(y)$. $\qquad\square$

Proposition 4. *For all $x, y \geq 1$, $E(x) \cdot E(y) \leq E(x + y)$.*

Proof. Take the product of shortest expressions for $E(x)$ and $E(y)$. The value of this product is $E(x) \cdot E(y)$ and it contains $x + y$ ones, so it cannot be greater than $E(x + y)$. $\qquad\square$

Theorem 3. *For all $k \geq 0$:*

$$E(3k + 2) = 2 \cdot 3^k;$$

$$E(3k + 3) = 3 \cdot 3^k;$$

$$E(3k + 4) = 4 \cdot 3^k.$$

Guy [10] attributes this result to John L. Selfridge.

Theorem 4. *For all $n \geq 8$:*

$$E_2(n) = \frac{8}{9} E(n).$$

This result is due to Rawsthorne[7].

The behaviour of $e(n)$ appears to be more complicated, for details see Section 5.1.

Lemma 1. *If $a+b=n$ and $\|a\|+\|b\|=\|n\|$ and $\|a\| \leq \|b\|$ and given $\|n\| \leq N$, then $E(\|a\|) \leq \frac{n-\sqrt{n^2-4E(N)}}{2}$.*

Proof. From $a+b=n$ we get $E(\|a\|)+E(\|b\|) \geq n$ using Proposition 3. Further, as $\|a\|+\|b\|=\|n\|$, we get $E(\|a\|)+\frac{E(\|n\|)}{E(\|a\|)} \geq n$ by Proposition 4. Still more due to the monotonicity of $E(x)$ we can substitute for the estimate of n's complexity: $E(\|a\|)+\frac{E(N)}{E(\|a\|)} \geq n$. For convenience let us denote $E(\|a\|)$ by x and $E(N)$ by y obtaining $x+\frac{y}{x} \geq n$. Solving the quadratic inequality for x we get $x \leq \frac{n-\sqrt{n^2-4y}}{2}$. To complete the proof insert back the original values of x and y. □

Corollary 1. *For $n \geq 29$, if $a+b=n$ and $\|a\|+\|b\|=\|n\|$ and $\|a\| \leq \|b\|$, then $a \leq 2n^{\log_2 3 - 1} \approx 2n^{0.585}$.*

Proof. By Theorem 1, $N \leq 3 \log_2 n$. Furthermore, we use the convenient fact that $E(n) \leq 3^{\frac{n}{3}}$.

$$a \leq E(\|a\|) \leq \frac{n-\sqrt{n^2-4E(N)}}{2} \leq$$
$$\leq \frac{n-\sqrt{n^2 - 4\cdot 3^{\log_2 n}}}{2} \leq \frac{n(1-\sqrt{1-4\cdot n^{\log_2 3 - 2}})}{2} \leq$$
$$\leq \frac{n(1-(1-4\cdot n^{\log_2 3 - 2}))}{2} \leq 2n^{\log_2 3 - 1}.$$

□

A similar proof of Corollary 1 is given in [11].

2.2 Ranking numbers

Consider an expression in basis $\{1, +, \cdot\}$ drawn as a rooted n-ary tree, its leaves containing ones and inner nodes containing either $+$ or $\cdot$. Since both addition and multiplication is associative and commutative, let us merge any adjacent additions and multiplications. For example, the shortest expression for 5^6 can be obtained as follows, it contains 29 ones (not $5 \cdot 6 = 30$, as one might expect):

$$5^6 = 15625 = 1 + 2^3 \cdot 3^2 \cdot 217 = 1 + 2^3 \cdot 3^2(1 + 2^3 \cdot 3^3)$$

The corresponding tree is drawn in Figure 1.

Definition 3. *The **height of an expression** is the height of the corresponding tree.*
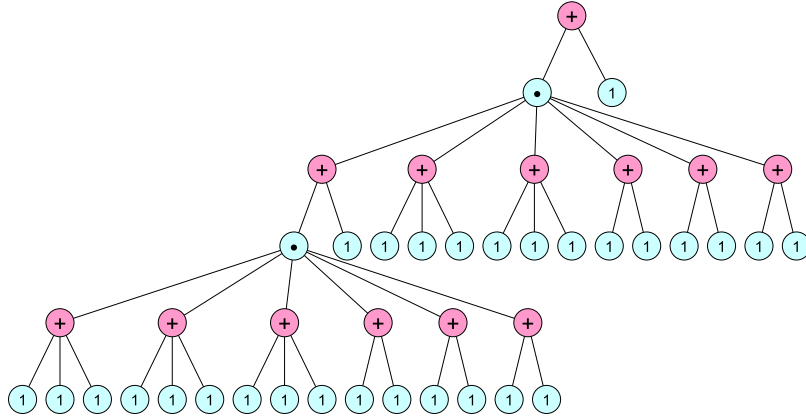
Fig. 1. The tree of the shortest expression for 5^6

In general, for a given n, there can be several shortest expressions of different height, for example, of height 2 and 1:

$$4 = (1 + 1) \cdot (1 + 1) = 1 + 1 + 1 + 1,$$

or of height 4, 3 and 2:

$$10 = (1+1)\cdot(1+(1+1)\cdot(1+1)) = 1+(1+1+1)\cdot(1+1+1) = (1+1)\cdot(1+1+1+1+1).$$

We will be interested in classifying the shortest expressions by introducing expression height as another complexity measure of positive integers:

Definition 4. $rank(n)$ – the ***rank of a number*** n – *is defined as the minimum height among the shortest expressions representing* n.

Examples:

* the only number of rank 0 is 1;
* the only numbers of rank 1 are: $2, 3, 4, 5$;
* numbers of rank 2: $6, 8, 9, 10, 12, 15, ...$; see Hypothesis 5 below;
* $rank(7) = rank(11) = rank(13) = 3$;
* $rank(14) = 4$.

For $n \leq 10^9$, $\max rank(n) = 19$ and $\max \|n\| = 67$. For other observations – see Section 5.4.

Obviously $rank(n) = 2$ for infinitely many numbers n. Namely, for all the powers of 3.

Hypothesis 3. *For all $r > 2$ there exists an infinite amount of numbers having rank r.*

This hypothesis implies

Hypothesis 4. *As a function of n, rank(n) is unlimited.*

Definition 5. *Let's denote by $r(n)$ the **smallest** m such that $rank(m) = n$.*

The values of $r(n)$ up to n=19 are represented in Table 4.
Now, let's try exploring more closely the numbers of rank 2.

Lemma 2. *For $n > 1$, the shortest expression of height ≤ 2 representing n, is obtained from the prime decomposition $p_1 \cdot \ldots \cdot p_k$ of n as $(1 + \ldots + 1) \cdot (1 + \ldots + 1) \cdot \ldots \cdot (1 + \ldots + 1)$.*

The number of ones in this expression, i.e. $p_1 + \ldots + p_k$, is called, sometimes, the **integer logarithm** of n (see also the sequence A001414[12]).

Theorem 5. *If $rank(n) = 2$, then $n = 2^a 3^b 5^c$ for some a, b, c with $a + b + c > 0$ and $c < 6$.*

Jānis Iraids, Rihards Opmanis. If n is divisible by a prime $p > 5$, then (since $\|p\| < p$) one can convert the expression of Lemma 2 into a shorter expression representing n. Hence, if $rank(n) = 2$, then $n = 2^a 3^b 5^c$ for some a, b, c with $a + b + c > 0$. But $\|5^6\| < 5 \cdot 6$ with the shortest expression being of height 5, as seen in Figure 1. Hence, if n is divisible by 5^6, then, again, one can convert the expression of Lemma 2 into a shorter expression representing n. $\square$

Hypothesis 5. *All and only numbers of rank 2 are the numbers $2^a 3^b 5^c > 5$ with $c < 6$. An equivalent form, extending Hypothesis 1:*

$$\|2^a 3^b 5^c\| = 2a + 3b + 5c$$

for all a, b, c with $a + b + c > 0$ and $c < 6$ (moreover, the product of $1 + \ldots + 1s$ is shorter than any other representation of $2^a 3^b 5^c$).

As verified by Jānis Iraids, $\|2^a 3^b 5^c\| = 2a + 3b + 5c$ is true for all $2^a 3^b 5^c \leq 10^{12}$ with $a + b + c > 0$ and $c < 6$.
For $c = 0$, Hypothesis 5 appears in [7].

2.3 Logarithmic complexity

Because of Theorem 1, the values of $\frac{\|n\|}{\log_3 n}$ are located within the segment $[3, 4.755]$. Hence, the

Definition 6. *The **logarithmic complexity** of $n > 1$ is defined as*

$$\|n\|_{\log} = \frac{\|n\|}{\log_3 n}.$$

For example, $\|2\|_{\log} = \frac{\|2\|}{\log_3 2} \approx 3.1699$.

Hypothesis 6. *It would follow from Hypothesis 5, that the values of $\|n\|_{\log}$ are* ***dense*** *across the segment $[3, \|2\|_{\log}]$, i.e. no subsegment of it is free from the values of $\|n\|_{\log}$.*

Proof. According to Hypothesis 5, $\left\|2^a 3^b\right\| = 2a + 3b$, for all $a + b > 0$. Hence, by choosing a and b appropriately, one can locate the value of $\left\|2^a 3^b\right\|_{\log}$ within any given subsegment of $[3, \|2\|_{\log}]$. $\qquad\square$

As observed by Mārtiņš Opmanis (and confirmed by Jānis Iraids for all the numbers $\leq 10^{12}$), it seems, the largest values of $\|n\|_{\log}$ are taken by single numbers, see Table 1. The lists in braces represent Cunningham chains of primes [13].

Table 1: Largest values of $\|n\|_{\log}$

n	$\|n\|$	$\approx \|n\|_{\log}$	$rank(n)$	Other properties
1439	26	3.928	9	$e(26), r(9), \{89, 179, 359, 719, 1439, 2879\}$
23	11	3.854	5	$e(11), r(5), \{2, 5, 11, 23, 47\}$
719	23	3.841	7	$e(23), \{89, 179, 359, 719, 1439, 2879\}$
179	18	3.812	7	$e(18), r(7), \{89, 179, 359, 719, 1439, 2879\}$
4283	29	3.809	7	$e(29), \{2141, 4283\}$
1438	25	3.777	8	$e(25), 2 \times 719$
59	14	3.772	5	$e(14), \{29, 59\}$
6299	30	3.767	7	$e(30)$, prime
15287	33	3.763	9	$e(33), \{3821, 7643, 15287\}$
107	16	3.762	5	$e(16), \{53, 107\}$
347	20	3.756	7	$e(20), \{173, 347\}$
1499	25	3.756	7	prime
467	21	3.754	5	$e(21), \{233, 467\}$
11807	32	3.749	7	$e(32), \{5903, 11807\}$
263	19	3.746	5	$e(19), \{131, 263\}$
21599	34	3.743	7	$e(34), \{2699, 5399, 10799, 21599\}$

The values of $\|n\|_{\log}$ become (and, it seems, stay) less than 3.60 approximately at $n = 2 \cdot 10^9$.

Let's consider the subsegments $[C, D], C < D$ of the segment $[3, 4.755]$ that do not contain the values of $\|n\|_{\log}$ at all. Of course, according to Hypothesis 6, then $C > \|2\|_{\log}$. Let's denote by C_1 the infimum of these numbers. I.e. C_1 is the point separating the area where the values of $\|n\|_{\log}$ are dense, from the area, where these values are not dense.

Hypothesis 7. *It would follow from Hypothesis 5, that $C_1 \geq \|2\|_{\log}$.*

On the other hand, for some numbers C, $\|n\|_{\log} > C$ only for finitely many values of n. Let's denote by C_2 the infimum of these numbers. This is also known

as $\limsup_{n\to\infty} \|n\|_{\log}$. I.e. C_2 is the point separating the area where the values of $\|n\|_{\log}$ are "absolutely sparse", from the area, where these values are not sparse.

Of course, $C_1 \leq C_2$. Hence,

Hypothesis 8. *It would follow from Hypothesis 5, that $C_2 \geq \|2\|_{\log}$.*

More about the possible value of C_2 – in Section 5.1.

3   Analytical results

3.1   Complexity of $2^n - 1$

For the sake of brevity let us introduce $A(n) = \|2^n - 1\| - 2n$ and $B(n) = \|2^n + 1\| - 2n$. We can then establish the following facts.

Theorem 6. *For $n \geq 1$,*

a) $A(2n) \leq A(n) + B(n)$;
b) $A(3n) \leq A(n) + B(n) + 1$;
c) $A(n + 1) \leq A(n) + 1$.

Proof. We shall provide the expressions that will result in these upper bounds.

a) $2^{2n} - 1 = (2^n - 1)(2^n + 1)$. If we take the complexity of both sides and subtract $4n$ we get:

$$A(2n) = \left\|2^{2n} - 1\right\| - 4n \leq \|2^n - 1\| - 2n + \|2^n + 1\| - 2n = A(n) + B(n).$$

b) $2^{3n} - 1 = (2^n - 1)((2^n + 1)2^n + 1)$. Similarly, we get

$$
\begin{aligned}
A(3n) = \left\|2^{3n} - 1\right\| - 6n &\leq \\
&\leq \|2^n - 1\| - 2n + \|2^n + 1\| - 2n + \|2^n\| - 2n + 1 \leq \\
&\leq A(n) + B(n) + 1.
\end{aligned}
$$

c) $2^{n+1} - 1 = (2^n - 1) \cdot 2 + 1$. Once again we have

$$A(n + 1) = \left\|2^{n+1} - 1\right\| - 2n - 2 \leq \|2^n - 1\| - 2n + 2 - 2 + 1 = A(n) + 1.$$

This method can be extended for numbers other than 2, but then it yields significantly less interesting results because of the very inefficient "$n + 1$" step. $\qquad\square$

Corollary 2. *(Kaspars Balodis) If $n > 1$, then*

$$\|2^n - 1\| \leq 2n + \lfloor \log_2 n \rfloor + H(n) - 3,$$

where $H(n)$ is the number of 1-s in the binary representation of n, i.e., $H(n)$ is the Hamming weight of n in binary.

Proof. Using the above theorem we can obtain an upper bound of $A(n)$, setting $B(n) = 1$ and $A(1) = A(2) = -1$ and ignoring the rule b) altogether. We will use rule c) per every 1 in the binary representation of n except the leftmost digit, in total, $H(n) - 1$ times. We will use rule a) per every digit, except the two leftmost digits, in total, $\lfloor \log_2 n \rfloor - 1$ times. In this way we will reduce n to 2 at which point $A(n) = -1$ having "paid" 1 for each application of any of the rules a) and c). To sum up:

$$A(n) \leq H(n) - 1 + \lfloor \log_2 n \rfloor - 1 + (-1).$$

$\square$

Corollary 3. *If $n > 1$ then $\|2^n - 1\| \leq 2n + 2\lfloor \log_2 n \rfloor - 2$.*

3.2 Connection of rank and defect

Definition 7. *(Harry Altman, Joshua Zelinsky) The **defect of a number** n is*

$$d(n) \overset{\text{def}}{=} \|n\| - 3 \cdot \log_3 n.$$

Proposition 5. *(Jānis Iraids)*

$$d(n) \geq \left\lfloor \frac{rank(n) - 1}{2} \right\rfloor \left(1 + 3\log_3 \frac{6}{7} \right).$$

Proof. We will prove this by induction on the rank. First of all, for all n having $rank(n) < 3$ the proposition is true trivially since

$$\left\lfloor \frac{rank(n) - 1}{2} \right\rfloor \left(1 + 3\log_3 \frac{6}{7} \right) \leq 0.$$

Now assuming that it is true for all n having $rank(n) < r$. Suppose r is an even number, then we again trivially have

$$\left\lfloor \frac{r - 1}{2} \right\rfloor \left(1 + 3\log_3 \frac{6}{7} \right) = \left\lfloor \frac{r - 2}{2} \right\rfloor \left(1 + 3\log_3 \frac{6}{7} \right).$$

If on the other hand, r is an odd number greater than 2, then the shortest expression for n has height at least 3, addition being the outermost operation. Now write n as a sum of numbers of even rank lower than r and order them in non-increasing fashion. This can be done since the numbers of odd rank can be split down further merging them with the n's outermost addition.

$$n = \sum_{i=1}^{k} a_i, \quad rank(a_1) \geq rank(a_2) \geq \cdots \geq rank(a_k).$$

Note that $rank(a_1) = r - 1$ and $k \geq 2$. For defects, we have

$$d(n) = \sum_{i=1}^{k} d(a_i) + 3\log_3 \frac{\prod_{i=1}^{k} a_i}{\sum_{i=1}^{k} a_i}.$$

Following the induction the defect of a_1 is at least $\frac{r-3}{2}\left(1 + 3\log_3 \frac{6}{7}\right)$. If $rank(a_k) \geq 2$ and so all a_i must necessarily be at least 6. The expression $3\log_3 \frac{\prod_{i=1}^{k} a_i}{\sum_{i=1}^{k} a_i}$ is minimised when $k = 2$ and $a_1 = a_2 = 6$ producing a minimum of 3. However, if there are l numbers of rank 0: $rank(a_{k-l+1}) = \ldots = rank(a_k) = 0$ then $d(a_{k-l+1}) = \ldots = d(a_k) = 1$ and $\sum_{i=k-l+1}^{k} d(a_i) + 3\log_3 \frac{\prod_{i=1}^{k} a_i}{\sum_{i=1}^{k} a_i}$ is minimised at $l = 1$, $a_1 = 6$ and $a_2 = 1$ giving the $1 + 3\log_3 \frac{6}{7}$. Consequently,

$$d(n) \geq \frac{r-3}{2}\left(1 + 3\log_3 \frac{6}{7}\right) + \min\left(3, 1 + 3\log_3 \frac{6}{7}\right) =$$

$$= \frac{r-1}{2}\left(1 + 3\log_3 \frac{6}{7}\right).$$

$\square$

4   Algorithms for computing the complexity of a number

The purpose of this chapter is to describe several algorithms for computing $\|n\|$. This is useful for both exploring the behaviour of the complexity and as well as a tool for verifying whether some hypotheses about the value of $\|n\|$ hold for large n.

It appears that computing the complexity of number n is a relatively difficult task. At the time of writing we are unaware of any algorithm that provably works faster than $O(n^{\log_2 3 - 1})$. Yet there exists an algorithm, that in practice runs in approximately $\Theta(n \log n)$ time.

The **first and simplest algorithm** is essentially as follows: evaluate all the possible expressions of $\{1, +, \cdot\}$ with increasing number of ones until an expression giving n is found. We can do so taking as basis the postfix notation, also known as Reverse Polish notation. Note, that in the postfix notation the first (leftmost) symbol will always be 1.

The possible expression generator will use three main rules:

a) After 1 only $+$ or 1 can follow. $\cdot$ can not follow because that would mean multiplication with 1 that is meaningless;

b) After $+$ only $\cdot$ or 1 can follow. $+$ can not follow because that would mean that it was possible to make the previous addition earlier and thus it is not postfix notation;

c) After $\cdot$ only $+$ or 1 can follow. $\cdot$ can not follow because that would mean that it was possible to make the previous multiplication earlier and thus it is not postfix notation.

For each total number of ones x in the expression excluding the leading one we can count the number of distinct expressions as 4^x – postfix notation of an expression with $x+1$ ones contains $2x+1$ symbols and for each symbol there are two possible symbols that can follow. Assuming, that evaluation of the expression takes roughly x time, the overall time complexity can be estimated as $O(4^x)$.

Thus the overall running time of the algorithm will be $O\left(4^{\|n\|}\right)$, which taking into account $\|n\| \leq 3\log_2 n$ will yield $O\left(n^6\right)$.

However, the expression generating algorithm can be improved by eliminating similar expressions of the same number. For instance, 6 can be represented as

* $(1+1)\cdot(1+1+1)$ – postfix notation $11+11+1+\cdot$ or
* $(1+1+1)\cdot(1+1)$ – postfix notation $11+1+11+\cdot$.

Since multiplication is associative, we can omit one of these representations restricting that the first multiplier has to be smaller or equal to the second multiplier. The same principle applies to additions.

The **second algorithm** uses the idea of sieving. We will use an array $f[1..n]$ of integer values where the complexity of the number n will be stored as $f[n]$. Before the main routine starts, precalculate the values of $E(n)$ in an array $E[1..m]$.

> **for** $i = 1$ to n **do**
> $\quad f[i] \leftarrow i$
> $updated \leftarrow True$
> $height \leftarrow 2$
> **while** $updated$ **do**
> $\quad updated \leftarrow False$
> $\quad$ **if** $height \equiv 1 \pmod 2$ **then**
> $\quad\quad$ **for** $i = 2$ to n **do**
> $\quad\quad\quad a \leftarrow \frac{n-\sqrt{n^2-4E[f[n]]}}{2}$
> $\quad\quad\quad$ **for** $j = 1$ to a **do**
> $\quad\quad\quad\quad$ **if** $f[i] > f[j] + f[i-j]$ **then**
> $\quad\quad\quad\quad\quad f[i] \leftarrow f[j] + f[i-j]$
> $\quad\quad\quad\quad\quad a \leftarrow \frac{n-\sqrt{n^2-4E[f[n]]}}{2}$
> $\quad\quad\quad\quad\quad updated \leftarrow True$
> $\quad$ **else**
> $\quad\quad$ **for** $i = 2$ to $\lfloor\sqrt{n}\rfloor$ **do**
> $\quad\quad\quad j \leftarrow i + i$
> $\quad\quad\quad k \leftarrow 2$
> $\quad\quad\quad$ **while** $j \leq n$ **do**
> $\quad\quad\quad\quad$ **if** $f[j] > f[k] + f[i]$ **then**
> $\quad\quad\quad\quad\quad f[j] \leftarrow f[k] + f[i]$
> $\quad\quad\quad\quad\quad updated \leftarrow True$
> $\quad\quad\quad\quad j \leftarrow j + i$
> $\quad\quad\quad\quad k \leftarrow k + 1$
> $\quad height \leftarrow height + 1$

If the array f is initialized as shown each pass will only update the complexity of numbers that have $rank(n) \geq height$. As a consequence, this algorithm produces the rank of a number; the rank can be stored if necessary. Note that the array f could be initialized with the upper bound provided by Theorem 1 because at no point would $f[i]$ be smaller than $\|i\|$ and yet exceed the value at the corresponding point in unmodified algorithm. To further reduce the running

time, one can use a bootstrapping step where numbers that could potentially be used as the smallest of two addends are computed. These are exactly the numbers that cannot be represented best as sums.

While the second algorithm is as fast as any we know, it uses a linear amount of memory – the array where the complexity values are stored. For $n > 10^{11}$ the calculation thus becomes unfeasible. We used this algorithm to calculate $\|n\|$ and $rank(n)$ up to $n = 1.5 \cdot 10^9$.

The **third and final algorithm** – the one we used for calculating $\|n\|$ for n up to 10^{12} – is conceptually very simple: for every natural number compute the complexity by definition and store for subsequent steps.

$$\|1\| = 1 \tag{1}$$

$$\|n\| = \min_{a+b=n \vee a \cdot b=n} \{\|a\| + \|b\|\} \tag{2}$$

The techniques used are identical to what Fuller [14] describes in the comments of his program. For factorisation we used an approach similar to the one described in [15]. The core idea is to maintain a priority queue of the so called eliminators – at any point n for each prime the priority is the smallest integer multiple that is no less than n.

5 Experimental results

5.1 $e(n)$ – the least number of complexity n

Function $e(n)$ corresponds to the sequence A005520 [9]. Our observations up $n = 89$ are represented in Table 2.

Hypothesis 9. *$e(n)$ is prime for all n, except $n \in \{1, 4, 7, 11, 25\}$.*

Observation 1. *For $k \leq 3$, the number $\frac{e(n)-k}{k+1}$ is prime for almost all n.*

If, for $k = 1$, Observation 1 holds for an infinite number of values of n, it would imply that there is an infinite number of **Sophie Germain primes** – these are defined as integers p such that p and $2p + 1$ are both primes.

Moreover, it seems that the sequence of e(n) contains primes which are the end numbers of increasingly long Cunningham chains. **Cunningham chain** (CC) of length k is defined as a sequence of k primes $\{p_1, p_2, ..., p_k\}$ such that $p_{i+1} = 2 \cdot p_i + 1, 1 \leq i < k$ [13]. In particular,

 ∗ $e(13)$ is the end number of the first CC of length 5: $\{2, 5, 11, 23, 47\}$;
 ∗ $e(26)$ is the end number of another CC of length 5: $\{89, 179, 359, 719, 1439\}$;
 ∗ $e(27)$ is the end number of the first CC of length 6:

$$\{89, 179, 359, 719, 1439, 2879\};$$

 ∗ $e(80)$ is the end number of another CC of length 5.

The above-mentioned are the only CCs of length ≥ 5 backward-generated by $e(n), n \leq 89$.

For the following 19 values of $n \leq 89$, $e(n)$ generates CCs of length 4:

$$\{11, 23, 34, 49, 51, 60, 61, 65, 66, 67, 70, 72, 73, 74, 77, 84, 86, 87, 89\}.$$

If, for $k = 2; 3$, Observation 1 holds for an infinite number of values of n, it would imply that there is an infinite number of integers p such that p and $3p+2$ (or, correspondingly, p and $4p + 3$) are both primes [16].

Table 2: Prime factorizations of numbers close to $e(n)$

n	$\frac{e(n)-2}{3}$	$\frac{e(n)-1}{2}$	$e(n)$	$e(n)+1$
1		$-$	$-$	1	2
2		$-$	$-$	2	3
3		$-$	1	3	2^2
4		$-$	$-$	2^2	5
5	1	2	5	$2 \cdot 3$
6	5	3	7	2^3
7	$-$	$-$	$2 \cdot 5$	11
8	3	5	11	$2^2 \cdot 3$
9	5	2^3	17	$2 \cdot 3^2$
10	$-$	$-$	$2 \cdot 11$	23
11	7	11	23	$2^3 \cdot 3$
12	13	$2^2 \cdot 5$	41	$2 \cdot 3 \cdot 7$
13	$3 \cdot 5$	23	47	$2^4 \cdot 3$
14	19	29	59	$2^2 \cdot 3 \cdot 5$
15	29	$2^2 \cdot 11$	89	$2 \cdot 3^2 \cdot 5$
16	$5 \cdot 7$	53	107	$2^2 \cdot 3^3$
17	$5 \cdot 11$	83	167	$2^3 \cdot 3 \cdot 7$
18	59	89	179	$2^2 \cdot 3^2 \cdot 5$
19	$3 \cdot 29$	131	263	$2^3 \cdot 3 \cdot 11$
20	$5 \cdot 23$	173	347	$2^2 \cdot 3 \cdot 29$
21	$5 \cdot 31$	233	467	$2^2 \cdot 3^2 \cdot 13$
22	227	$11 \cdot 31$	683	$2^2 \cdot 3^2 \cdot 19$
23	239	359	719	$2^4 \cdot 3^2 \cdot 5$
24	$11 \cdot 37$	$13 \cdot 47$	1223	$2^3 \cdot 3^2 \cdot 17$
25	$-$	$-$	$2 \cdot 719$	1439
26	479	719	1439	$2^5 \cdot 3^2 \cdot 5$
27	$7 \cdot 137$	1439	2879	$2^6 \cdot 3^2 \cdot 5$
28	$5 \cdot 251$	$7 \cdot 269$	3767	$2^3 \cdot 3 \cdot 157$
29	1427	2141	4283	$2^2 \cdot 3^2 \cdot 7 \cdot 17$
30	2099	$47 \cdot 67$	6299	$2^2 \cdot 3^2 \cdot 5^2 \cdot 7$
31	3359	5039	10079	$2^5 \cdot 3^2 \cdot 5 \cdot 7$

32	$5 \cdot 787$	5903	11807	$2^5 \cdot 3^2 \cdot 41$
33	$5 \cdot 1019$	7643	15287	$2^3 \cdot 3 \cdot 7^2 \cdot 13$
34	$23 \cdot 313$	10799	21599	$2^5 \cdot 3^3 \cdot 5^2$
35	$3 \cdot 3733$	$107 \cdot 157$	33599	$2^6 \cdot 3 \cdot 5^2 \cdot 7$
36	$5 \cdot 23 \cdot 131$	$2 \cdot 11299$	45197	$2 \cdot 3^6 \cdot 31$
37	18679	28019	56039	$2^3 \cdot 3 \cdot 5 \cdot 467$
38	$5 \cdot 5443$	40823	81647	$2^4 \cdot 3^6 \cdot 7$
39	32999	49499	98999	$2^3 \cdot 3^2 \cdot 5^3 \cdot 11$
40	54419	81629	163259	$2^2 \cdot 3^2 \cdot 5 \cdot 907$
41	$53 \cdot 1283$	101999	203999	$2^5 \cdot 3 \cdot 5^3 \cdot 17$
42	80627	120941	241883	$2^2 \cdot 3^2 \cdot 6719$
43	$5 \cdot 24763$	185723	371447	$2^3 \cdot 3^2 \cdot 7 \cdot 11 \cdot 67$
44	180179	270269	540539	$2^2 \cdot 3^3 \cdot 5 \cdot 7 \cdot 11 \cdot 13$
45	196799	295199	590399	$2^6 \cdot 3^2 \cdot 5^2 \cdot 41$
46	302399	453599	907199	$2^6 \cdot 3^4 \cdot 5^2 \cdot 7$
47	$173 \cdot 2083$	540539	$1\,081\,079$	$2^3 \cdot 3^3 \cdot 5 \cdot 7 \cdot 11 \cdot 13$
48	617039	925559	$1\,851\,119$	$2^4 \cdot 3^3 \cdot 5 \cdot 857$
49	680399	$1\,020\,599$	$2\,041\,199$	$2^4 \cdot 3^6 \cdot 5^2 \cdot 7$
50	$1\,081\,079$	$1\,621\,619$	$3\,243\,239$	$2^3 \cdot 3^4 \cdot 5 \cdot 7 \cdot 11 \cdot 13$
51	$1\,280\,159$	$1\,920\,239$	$3\,840\,479$	$2^5 \cdot 3^3 \cdot 5 \cdot 7 \cdot 127$
52	$2\,187\,359$	$3\,281\,039$	$6\,562\,079$	$2^5 \cdot 3^3 \cdot 5 \cdot 7^2 \cdot 31$
53	$2\,735\,519$	$4\,103\,279$	$8\,206\,559$	$2^5 \cdot 3^2 \cdot 5 \cdot 41 \cdot 139$
54	$3\,898\,919$	$5\,848\,379$	$11\,696\,759$	$2^3 \cdot 3^2 \cdot 5 \cdot 32491$
55	$4\,882\,919$	$7\,324\,379$	$14\,648\,759$	$2^3 \cdot 3^2 \cdot 5 \cdot 7 \cdot 5813$
56	$13 \cdot 59 \cdot 9697$	$11\,156\,399$	$22\,312\,799$	$2^5 \cdot 3^3 \cdot 5^2 \cdot 1033$
57	$9\,164\,959$	$13\,747\,439$	$27\,494\,879$	$2^5 \cdot 3 \cdot 5 \cdot 7^3 \cdot 167$
58	$13\,915\,439$	$20\,873\,159$	$41\,746\,319$	$2^4 \cdot 3^3 \cdot 5 \cdot 7 \cdot 11 \cdot 251$
59	$17\,417\,399$	$26\,126\,099$	$52\,252\,199$	$2^3 \cdot 3^2 \cdot 5^2 \cdot 7 \cdot 11 \cdot 13 \cdot 29$
60	$26\,110\,559$	$39\,165\,839$	$78\,331\,679$	$2^5 \cdot 3^2 \cdot 5 \cdot 7 \cdot 19 \cdot 409$
61	$36\,202\,319$	$54\,303\,479$	$108\,606\,959$	$2^4 \cdot 3^3 \cdot 5 \cdot 7 \cdot 11 \cdot 653$
62	$6577 \cdot 7247$	$71\,495\,279$	$142\,990\,559$	$2^5 \cdot 3^2 \cdot 5 \cdot 109 \cdot 911$
63	$67\,699\,439$	$101\,549\,159$	$203\,098\,319$	$2^4 \cdot 3^3 \cdot 5 \cdot 17 \cdot 5531$
64	$91\,328\,639$	$136\,992\,959$	$273\,985\,919$	$2^7 \cdot 3^2 \cdot 5 \cdot 13 \cdot 3659$
65	$127\,340\,639$	$191\,010\,959$	$382\,021\,919$	$2^5 \cdot 3^4 \cdot 5 \cdot 7 \cdot 4211$
66	$165\,145\,679$	$247\,718\,519$	$495\,437\,039$	$2^4 \cdot 3^3 \cdot 5 \cdot 7^2 \cdot 31 \cdot 151$
67	$227\,109\,119$	$340\,663\,679$	$681\,327\,359$	$2^8 \cdot 3^2 \cdot 5 \cdot 7^2 \cdot 17 \cdot 71$
68	$335\,430\,119$	$503\,145\,179$	$1\,006\,290\,359$	$2^3 \cdot 3^2 \cdot 5 \cdot 601 \cdot 4651$
69	$468\,798\,119$	$703\,197\,179$	$1\,406\,394\,359$	$2^3 \cdot 3^3 \cdot 5 \cdot 7 \cdot 17 \cdot 31 \cdot 353$
70	$619\,264\,799$	$928\,897\,199$	$1\,857\,794\,399$	$2^5 \cdot 3^3 \cdot 5^2 \cdot 7 \cdot 11 \cdot 1117$
71	$909\,474\,719$	$1\,364\,212\,079$	$2\,728\,424\,159$	$2^5 \cdot 3^2 \cdot 5 \cdot 7 \cdot 11^2 \cdot 2237$
72	$1\,247\,732\,639$	$1\,871\,598\,959$	$3\,743\,197\,919$	$2^5 \cdot 3^4 \cdot 5 \cdot 7 \cdot 11^3 \cdot 31$
73	$40499 \cdot 41221$	$2\,504\,113\,919$	$5\,008\,227\,839$	$2^9 \cdot 3^3 \cdot 5 \cdot 7 \cdot 11 \cdot 941$
74	$2\,290\,896\,719$	$3\,436\,345\,079$	$6\,872\,690\,159$	$2^4 \cdot 3^3 \cdot 5 \cdot 7 \cdot 454543$
75	$3\,279\,830\,399$	$7643 \cdot 643693$	$9\,839\,491\,199$	$2^7 \cdot 3^4 \cdot 5^2 \cdot 7 \cdot 11 \cdot 17 \cdot 29$
76	$4\,495\,159\,679$	$6\,742\,739\,519$	$13\,485\,479\,039$	$2^7 \cdot 3^2 \cdot 5 \cdot 11^3 \cdot 1759$

77	5 574 925 439	8 362 388 159	16 724 776 319	$2^7 \cdot 3^8 \cdot 5 \cdot 7 \cdot 569$
78	8 226 486 239	12 339 729 359	24 679 458 719	$2^5 \cdot 3^2 \cdot 5 \cdot 7 \cdot 19 \cdot 128861$
79	11 841 566 159	17 762 349 239	35 524 698 479	$2^4 \cdot 3^2 \cdot 5 \cdot 49339859$
80	14 737 208 639	22 105 812 959	44 211 625 919	$2^6 \cdot 3^2 \cdot 5 \cdot 7^2 \cdot 11 \cdot 19 \cdot 1499$
81	20 797 230 719	31 195 846 079	62 391 692 159	$2^7 \cdot 3^2 \cdot 5 \cdot 7^2 \cdot 221059$
82	31 251 071 039	46 876 606 559	93 753 213 119	$2^6 \cdot 3^2 \cdot 5 \cdot 7^2 \cdot 61 \cdot 10891$
83	40 517 305 919	60 775 958 879	121 551 917 759	$2^6 \cdot 3^4 \cdot 5 \cdot 7 \cdot 13 \cdot 29 \cdot 1777$
84	54 513 320 399	81 769 980 599	163 539 961 199	$2^4 \cdot 3^3 \cdot 5^2 \cdot 7 \cdot 11 \cdot 196657$
85	83 528 413 919	125 292 620 879	250 585 241 759	$2^5 \cdot 3^4 \cdot 5 \cdot 7 \cdot 2762183$
86	106 809 776 639	160 214 664 959	320 429 329 919	$2^9 \cdot 3^3 \cdot 5 \cdot 7^2 \cdot 37 \cdot 2557$
87	141 615 840 239	212 423 760 359	424 847 520 719	$2^4 \cdot 3^4 \cdot 5 \cdot 7 \cdot 9366127$
88	210 123 688 319	315 185 532 479	630 371 064 959	$2^7 \cdot 3^3 \cdot 5 \cdot 7 \cdot 13 \cdot 17 \cdot 23581$
89	290 857 880 879	436 286 821 319	872 573 642 639	$2^4 \cdot 3^3 \cdot 5 \cdot 7^3 \cdot 19 \cdot 61987$

The behaviour of $e(n)$ provides some evidence that the logarithmic complexity of n does not tend to 3, and even that the constant $C_2 = \limsup_{n \to \infty} \|n\|_{\log}$ is greater than $\|2\|_{\log}$. It is useful to note the following fact:

Proposition 6.

$$C_2 = \limsup_{n \to \infty} \|n\|_{\log} = \limsup_{n \to \infty} \|e(n)\|_{\log}$$

Proof. Since $\|e(n)\|_{\log}$ is a subsequence of $\|n\|_{\log}$ the $\geq$ follows obviously.

The $\leq$ is proven from the contrary. Assume that there exists an infinite list of numbers $\{x_i\}$, that only have a finite number of elements from sequence $e(n)$ of logarithmic complexity $\geq \min_i\{\|x_i\|_{\log}\}$. Furthermore, there is an infinite subsequence $\{y_i\}$ of $\{x_i\}$'s such that the complexity of numbers strongly increases. But any number of this sequence has a corresponding number in $e(n)$, namely $e(\|y_i\|)$ that has greater or equal logarithmic complexity.

Alternatively, on the axis of logarithmic complexity, $e(n)$ yields the rightmost point of numbers of complexity n. □

Hypothesis 10. *The limit* $\lim_{n \to \infty} \|e(n)\|_{\log}$ *exists and*

$$\lim_{n \to \infty} \|e(n)\|_{\log} = \limsup_{n \to \infty} \|e(n)\|_{\log}$$

As one can see in Figures 2 and 3:

Hypothesis 11. $\log_3 e(n)$ *behaves almost linearly, namely,*

$$\log_3 e(n) \approx 0.297n - 1.55.$$

Hence,

$$\|e(n)\|_{\log} = \frac{n}{\log_3 e(n)} \approx 3.37 + \frac{5.2}{0.297n - 1.55},$$
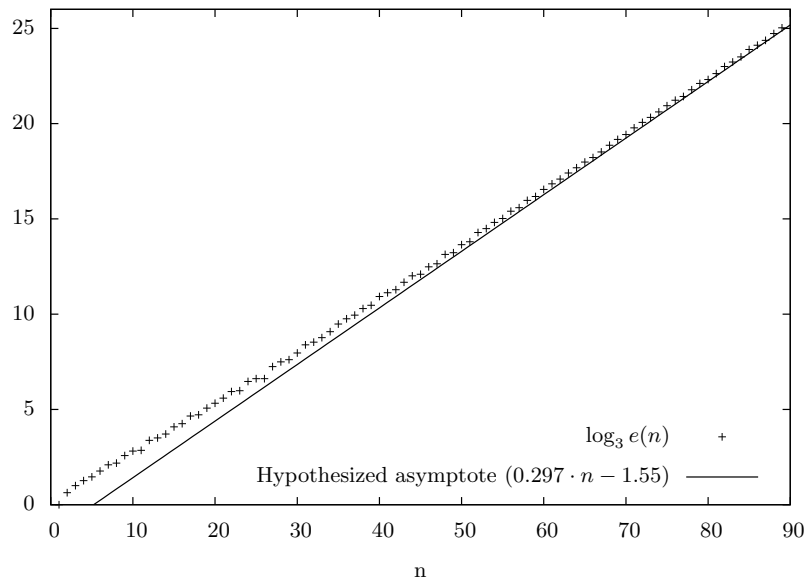
and $\lim_{n \to \infty} \|e(n)\|_{\log} \approx 3.37$.

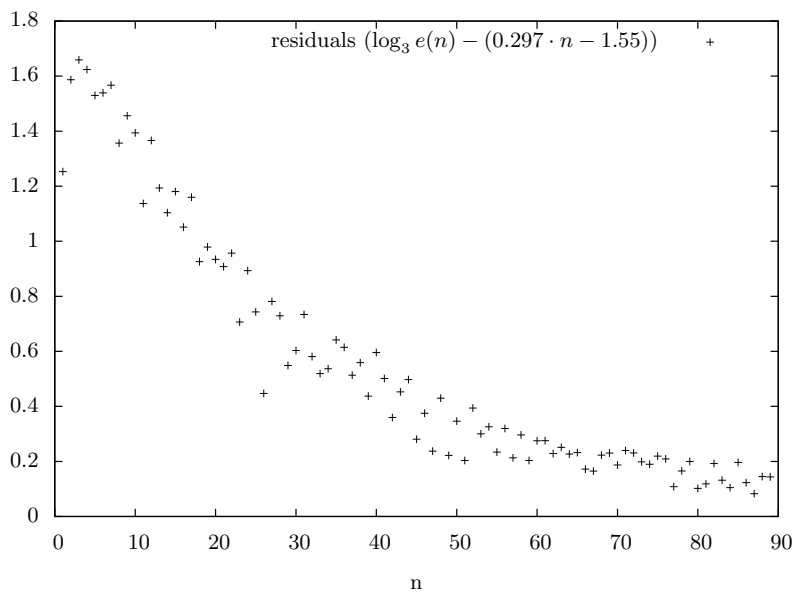**Fig. 2.** Values of $\log_3 e(n)$ compared with hypothesized asymptote

Fig. 3. Residuals of the asymptote

Lemma 3. *For all $n > 1$:*

$$\|n\|_{\log} \leq \|e(\|n\|)\|_{\log}.$$

Proof. Obviously, $\|n\| = \|e(\|n\|)\|$, and $n \geq e(\|n\|)$. Hence, $\log_3 n \geq \log_3 e(\|n\|)$, and $\|n\|_{\log} \leq \|e(\|n\|)\|_{\log}$. □

Thus, if $\epsilon > 0$, then, it seems, $\|n\|_{\log} > 3.37 + \epsilon$ can be true only for finitely many values of n. Thus, in terms of Section 2.3:

Hypothesis 12. *It would follow from Hypothesis 11, that $C_2 \leq 3.37$ (approximately).*

5.2 Structure of shortest expressions

Analyzing the structure of the shortest expressions representing numbers, we have come to the conclusion that we should not only abandon any attempts to obtain a "deterministic" method for construction of shortest expressions, and turn to "nondeterministic" methods. We should abandon also radical attempts to predict regularities in the structure of shortest expressions.

For example, one might propose the following hypothesis: if p is the smallest prime divisor of $n > 1$, then

$$\|n\| = \min\left(1 + \|n - 1\|, \|p\| + \|n/p\|\right).$$

For $p = n$, and $p = 2$ (with prime n/p) this hypothesis appears in Guy [10]: for any prime p,

$$\|p\| = 1 + \|p - 1\|,$$
$$\|2p\| = \min\left(1 + \|2p - 1\|, 2 + \|p\|\right).$$

Similarly, one could suppose:

$$\|3p\| = \min\left(1 + \|3p - 1\|, 3 + \|p\|\right).$$

The first hypothesis fails, the smallest counterexample being

$$p = 353\,942\,783 = 2 \cdot 3 + (1 + 2^2 \cdot 3^2)(2 + 3^4(1 + 2 \cdot 3^{10})),$$

$\|p\| = 63, 1 + \|p - 1\| = 64$, found by Martin N. Fuller, 2008, see [2].

The second hypothesis fails as well, the smallest counterexample being

$$2p = 10\,278\,600\,694 = 2 \cdot 3 + (1 + 2 \cdot 3^2(1 + 2^4))(1 + 3^{14}(1 + 2 \cdot 3)),$$

$\|2p\| = 72, 1 + \|2p - 1\| = 2 + \|p\| = 73$, found by Jānis Iraids, 2010.

By analogy, it seems, the third hypothesis also should fail, but this does not happen for $3p \leq 10^{12}$.

Thus, when trying to build the shortest expression representing a number, subtraction of 1 and division by primes are not universal candidates for the first operation. How about subtraction of other numbers?

Subtractions of 3, 4, 5, 7, 10, 11, 13, etc. include subtraction of 1. Thus, it remains to consider only subtractions of 6, 8, 9, 12, etc.

The first number, for which subtraction of 6 is necessary as the first operation, is the above prime found by Fuller:

$$353\,942\,783 = 2 \cdot 3 + (1 + 2^2 \cdot 3^2)(2 + 3^4(1 + 2 \cdot 3^{10})).$$

Until 10^{12}, there are only 21360 numbers for which subtraction of 6, 8 or 9 is necessary as the first operation.

Until 10^{12}, there are exactly 3 numbers for which the first operation must be subtraction of 8:

$$341\,317\,451\,698 = 2 \cdot prime = 2^3 + (1 + 2^4(1 + 2 \cdot 3^3))(1 + 3^{18});$$

$$474\,934\,483\,834 = 2 \cdot 6011 \cdot 39505447 = 2^3 + (1 + 2^4)(1 + 2^3 \cdot 3^4)(1 + 3^{16});$$

$$782\,747\,233\,558 = 2 \cdot prime = 2^3 + (1 + 2^4(1 + 2 \cdot 3^5))(1 + 3^{15}(1 + 2 \cdot 3)).$$

Until 10^{12}, there are 119 numbers for which the first operation must be subtraction of 9, the first three ones being:

$$16\,534\,727\,299 = 103 \cdot 160531333 = 3^2 + (1 + 2^7 \cdot 3^3)(1 + 3^{14});$$

$$68\,238\,632\,999 = prime = 3^2 + (1 + 1 + 3^2)(1 + 2^4 \cdot 3^4)(1 + 3^{14});$$

$$85\,619\,928\,299 = prime = 3^2 + (1 + 2^2 \cdot 3)(1 + 2^4)(1 + 3^{18}).$$

Necessity for subtraction of 12 (or larger addendum) was not detected for numbers until 10^{12}.

According to Corollary 1, if $n \geq 29$ and the shortest expression for n is a sum $n = a + b$, then the smaller addendum $a \leq 2n^{\log_2 3 - 1} \approx 2n^{0.585}$. However, the above observations show that for $n \leq 10^{12}$ the smaller addendum does not exceed 9.

5.3 Complexity of $2^n + 1$ and $2^n - 1$

Since, it seems, $\|2^n\| = 2n$, one can suppose

Hypothesis 13. *For all $n \geq 0$, except 3 and 9,*

$$\|2^n + 1\| = 2n + 1.$$

Hypothesis 13 is true for all $2^n + 1 \leq 10^{12}$, i.e. for all $n \leq 39$ – as verified by Jānis Iraids.

Both exceptions are due to relatively massive divisibility by 3:

$$2^3 + 1 = 9 = 2 \cdot 2 \cdot 2 + 1 = 3 \cdot 3;$$

$$2^9 + 1 = 513 = (3 \cdot 3 \cdot 2 + 1) \cdot 3 \cdot 3 \cdot 3.$$

On the other hand, since we do not have subtraction in our expression basis $\{1, +, \cdot\}$, the numbers $2^n - 1$ seem to be more complicated than 2^n. In Theorem 2, an upper bound of $\|2^n - 1\|$ was proved. In Table 3 this result is compared with experimental data for $n \leq 39$.

Table 3: Complexity of $\|2^n - 1\|$

n	$\|2^n - 1\| - 2n$	$\lfloor \log_2 n \rfloor + H(n) - 3$	n	$\|2^n - 1\| - 2n$	$\lfloor \log_2 n \rfloor + H(n) - 3$
1	-1	$-$	21	2	4
2	-1	-1	22	3	4
3	0	0	23	4	5
4	0	0	24	2	3
5	1	1	25	3	4
6	0	1	26	3	4
7	1	2	27	2	5
8	1	1	28	3	4
9	1	2	29	4	5
10	2	2	30	3	5
11	3	3	31	4	6
12	1	2	32	3	3
13	2	3	33	4	4
14	2	3	34	4	4
15	2	4	35	4	5
16	2	2	36	2	4
17	3	3	37	2	5
18	1	3	38	2	5
19	2	4	39	3	6
20	2	3	$-$	$-$	$-$

Thus, it seems, the upper bound of Theorem 2 is exact for all $n = 2^k, k > 0$:

Observation 2. *For all $0 \le k \le 5$,*

$$\left\|2^{2^k} - 1\right\| = 2 \cdot 2^k + k - 2.$$

5.4   Observing ranks

The first number for which subtraction of 6 is necessary as the first operation to obtain the shortest expression of minimum height, is

$$22\,697\,747 = prime = 2 \cdot 3 + (2 + 3^7)(1 + 2^7 \cdot 3^4),$$

complexity 55, rank 5.

The values of $r(n)$ up to $n = 19$ are represented in Table 4. The lists in braces represent Cunningham chains of primes [13].

Table 4: $r(n)$ – the least number of rank n

n	$r(n)$	$\|r(n)\|$	Other properties
1	2	2	$e(2), \{2, 5, 11, 23, 47\}$
2	6	5	$2 \cdot 3$

3	7	6	$e(6), \{3, 7\}$
4	14	8	$2 \cdot 7$
5	23	11	$e(11), \{2, 5, 11, 23, 47\}$
6	86	14	$2 \cdot 43$
7	179	18	$e(18), \{89, 179, 359, 719, 1439, 2879\}$
8	538	21	$2 \cdot 269$
9	1439	26	$e(26), \{89, 179, 359, 719, 1439, 2879\}$
10	9566	30	$2 \cdot 4783$
11	21383	33	$\{10691, 21383, 42767\}$
12	122847	37	$3 \cdot 40949$
13	777419	44	prime
14	1965374	46	$2 \cdot 982687$
15	6803099	51	$\{3401549, 6803099\}$
16	19860614	53	$2 \cdot 9930307$
17	26489579	55	$\{13244789, 26489579, 52979159\}$
18	269998838	61	$2 \cdot 4093 \cdot 32983$
19	477028439	64	$14207 \cdot 33577$

In Figure 4, the values of $\log_3 r(n)$ are compared with n.

Observation 3. $\log_3 r(n)$ *tends to* n*, hence, it seems,* $r(n) \approx 3^n$*.*

5.5 Collapse of powers

While attempting to prove or disprove the Hypothesis 1 one might try to generalize the hypothesis by looking for other numbers n, such that the shortest expressions for all the integer powers of n: n^2, n^3, n^4, ... can be obtained as products of shortest expressions for n.

Definition 8. *If this property holds, then the number n is called **resistant**.*

Obviously, if n is resistant, then for all $k > 0$,

$$\left\| n^k \right\| = k \left\| n \right\|;$$

$$\left\| n^k \right\|_{\log} = \left\| n \right\|_{\log}.$$

Currently, only the number 3 is proved to be resistant. According to Hypothesis 1, another resistance candidate is the number 2. Existence of resistant numbers other than powers of 3 would provide a lower bound on C_2:

Proposition 7. *If $n = 2$ or $n \neq 3^k$ is resistant, $C_2 \geq \left\| n \right\|_{\log} > 3$.*

Definition 9. *If $k > 0$ is the least number such that $\left\| n^k \right\| < k \left\| n \right\|$, let us say that the number n **collapses at** k.*
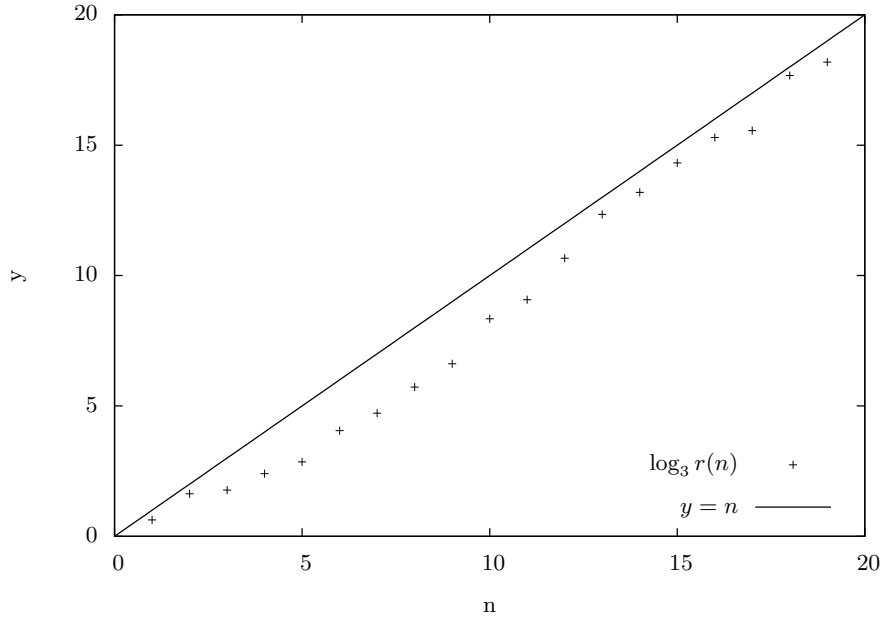
Fig. 4. Values of $\log_3 r(n)$ compared with n.

For example, the number 5 collapses at 6 (see Section 2.2):

$$\left\|5^6\right\| = 29 < 6\left\|5\right\| = 6 \cdot 5 = 30.$$

However, it seems, most primes collapse already at 2, for example,

$$\left\|11^2\right\| = 15 < 2\left\|11\right\| = 2 \cdot 8 = 16;\ 11^2 = 121 = 1 + 2^3 \cdot 3 \cdot 5.$$

Of the 168 primes until 10^3, 120 primes collapse at 2 (71%), 24 – at 3 (14%), 3 – at 4 (2%), the remaining 21 do not collapse until 4 (13%), and, currently, 12 of them are not known to collapse at all, namely,

$$2, 3, 109, 163, 379, 433, 487, 541, 577, 733, 739, 811.$$

For details, see Table 5. The rest of numbers all collapse at 2 or 3, and their logarithmic complexity exceeds 3.417. It seems, primes having larger logarithmic complexity are more likely to collapse.

Table 5: Powers in which primes until 1000 collapse, sorted by increasing logarithmic complexity

p	$collapses$	$\|p\|$	$rank(p)$	$\approx \|p\|_{\log}$
3	–	3	1	3.000

2	?(> 39)	2	1	3.170
487	?(> 4)	18	3	3.196
163	?(> 5)	15	3	3.235
433	?(> 4)	18	3	3.257
109	?(> 5)	14	3	3.278
811	?(> 4)	20	3	3.280
577	?(> 4)	19	3	3.283
769	3	20	3	3.307
757	≤ 6	20	5	3.314
541	?(> 4)	19	3	3.317
739	?(> 4)	20	5	3.326
73	6	13	3	3.329
379	?(> 4)	18	5	3.331
733	?(> 4)	20	5	3.331
271	4	17	3	3.334
193	4	16	3	3.340
991	≤ 12	21	5	3.344
37	5	11	3	3.347
977	2	21	5	3.351
19	6	9	3	3.358
97	6	14	3	3.362
257	≤ 6	17	3	3.366
937	3	21	5	3.372
673	3	20	5	3.374
919	3	21	5	3.381
181	3	16	3	3.381
661	3	20	5	3.384
7	9	6	3	3.387
653	2	20	5	3.390
337	3	18	5	3.398
641	4	20	3	3.400
883	3	21	5	3.401
127	2	15	5	3.402
881	2	21	5	3.402
877	3	21	5	3.405
241	3	17	3	3.405
631	2	20	5	3.408
457	3	19	5	3.408
331	3	18	5	3.408
5	6	5	1	3.413

Of the 1229 primes until 10^4, 1030 primes collapse at 2 (84%), 122 – at 3 (10%), the remaining 77 do not collapse until 3 (6%).

Of the 78498 primes until 10^6, 71391 primes collapse at 2 (91%), the remaining 7107 do not collapse until 2 (9%), and, currently, only 2722 are not known to collapse at all (4%).

The following observations are true for primes less than 10^6.

Observation 4. *Almost all primes collapse at* 2.

Observation 5. *If a prime p does not collapse at all, then* $\|p\|_{\log} < 3.364$. *If a prime p collapses, then* $\|p\|_{\log} > 3.180 > \|2\|_{\log}$.

If we turn to composite numbers, we can encounter numbers that collapse because one ore more of their prime divisors collapse – even when their initial shortest expression did not directly contain the prime as a multiplier. $3^4 + 1 = 82$ is an example of such a number:

* $\|82\| = 13$ and $\|41\| = 12$
* $\left\|82^{12}\right\| < 12 \cdot 13$, because $\left\|41^{12}\right\| \leq 131 < 12 \cdot 11$:

$$41^{12} = [2^3 \cdot 3^2 \cdot \left(2^2 + 1\right) \left(3^7 + 1\right)$$
$$\left(2 \cdot 3^4 \cdot \left(2^3 \cdot 3^2 \cdot \left(2 \cdot 3^3 + 1\right) + 1\right) + 1\right)$$
$$\left(2 \cdot 3^4 \cdot \left(2 \cdot 3^9 \cdot (2 \cdot 3 + 1) + 1\right) + 1\right)] + 1$$

On the other hand, there could possibly be composite numbers that do not collapse even though some of their prime divisors do. Obviously, a necessary condition is that the shortest expression for the composite number does not directly contain the collapsing prime as a multiplier. One candidate is $3^5 + 1 = 244 = 2^2 \cdot 61$; $\|244\| = 16$ but $\|61\| = 13$. Using a heuristic algorithm Juris Čerņenoks was able to produce expressions for 61^k for $k = 7..15$. His results suggest that 61 does not collapse well enough, i.e., the number of ones saved from collapsing is less than required to catch up with the expression $244 = 3^5 + 1$.

On a side note, the above mentioned 82 and 244 are interesting because if they proved to be resistant, then C_2 would exceed $\|2\|_{\log}$.

6   Conclusions

Trying to explore representing of natural numbers by arithmetical expressions using ones, addition, multiplication and parentheses, one arrives very soon at the problems that are easy to formulate, but (it seems) extremely hard to solve.

Consider, for example, the above Hypothesis 1 stating that the best way of representing of 2^n is $(1+1)(1+1)....$ We consider proving or disproving of Hypothesis 1 as one of the biggest challenges of number theory.

Almost as challenging seems Hypothesis 4 stating that as a function of n, $rank(n)$ is unlimited. Rank is an additional (to $\|n\|$) measure of integer complexity introduced in this paper.

As another challenge we regard determining of the distribution of the values of logarithmic complexity $\|n\|_{\log}$ which are located within the segment $[3, 4.755]$.

First, denote by C_1 the point separating the area (on the left) where the values of $\|n\|_{\log}$ are dense, from the area, where these values are not dense. On the other hand, on the right, the values of $\|n\|_{\log}$ are "absolutely sparse": for most C, $\|n\|_{\log} > C$ only for finitely many values of n. Denote by C_2 the point separating the area (on the right) where the values of $\|n\|_{\log}$ are "absolutely sparse", from the area, where these values are not sparse. Of course, $C_1 \leq C_2$. Our Hypotheses 5 and 11 (if true) imply that

$$3.1699 \approx \|2\|_{\log} \leq C_1 \leq C_2 \leq 3.37.$$

Our main experimental "device" is the database containing the values of $\|n\|$ up to $n = 10^{12}$ calculated by Jānis Iraids. The database can be accessed by using an online calculator page linked from [2].

And finally, our Hypothesis 9 and Observation 1 (if true) imply that there is an infinite number of Sophie Germain primes, and even that there is an infinite number of Cunningham chains of length 4 (at least).

References

1. Guy, R.K.: F26 Expressing numbers with just ones. In: Unsolved Problems in Number Theory. Springer Science+Business Media, Inc. (2004)
2. Sloane, N.J.A.: The On-Line Encyclopedia of Integer Sequences. A005245 Complexity of n: number of 1's required to build n using $+$ and $\cdot$.
3. Arias de Reyna, J.: Complejidad de los números naturales. Gaceta de la Real Sociedad Matemática Española (2000)
4. Voss, J.: The On-Line Encyclopedia of Integer Sequences. A091333 Number of 1's required to build n using $+$, $-$, $\cdot$ and parentheses.
5. N. J. A. Sloane, David W. Wilson: The On-Line Encyclopedia of Integer Sequences. A025280 Complexity of n: number of 1's required to build n using $+$, $\cdot$ and $\uparrow$.
6. Voss, J.: The On-Line Encyclopedia of Integer Sequences. A091334 Number of 1's required to build n using $+$, $-$, $\cdot$, $\uparrow$ and parentheses.
7. Rawsthorne, D.A.: How many 1's are needed? Fibonacci Quarterly (1989)
8. Sloane, N.J.A.: The On-Line Encyclopedia of Integer Sequences. A000792 $a(n) = max(n - i)a(i) : i < n; a(0) = 1$.
9. Sloane, N.J.A.: The On-Line Encyclopedia of Integer Sequences. A005520 Smallest number of complexity n: smallest number requiring n 1's to build using $+$ and $\cdot$.
10. Guy, R.K.: Some suspiciously simple sequences. The American Mathematical Monthly (1986)
11. Srinivas Vivek V., Shankar B. R.: Integer complexity: Breaking the $\theta(n^2)$ barrier. World Academy of Science, Engineering and Technology (2008)
12. Sloane, N.J.A.: The On-Line Encyclopedia of Integer Sequences. A001414 Integer log of n: sum of primes dividing n (with repetition).
13. Caldwell, C.K.: Cunningham chain. `http://primes.utm.edu/glossary/xpage/CunninghamChain.html` [Last accessed 27 January 2012]
14. Fuller, M.N.: Program to calculate A005245, A005520, A005421. `http://oeis.org/A005245/a005245.c.txt` [Last accessed: 30 December 2011]
15. Bennet, T.W.: Prime Generator Algorithm. `http://sandbox.mc.edu/~bennet/cs220/codeex/pgenintro.html` [Last accessed 13 January 2012]

16. Dickson, L.E.: A New Extension of Dirichlet's Theorem on Prime Numbers. Messenger Math (1904)