COMPLEXITY OF NATURAL NUMBERS

Juan Arias de Reyna
Facultad de Matemáticas, Universidad de Sevilla, Sevilla, Spain
arias@us.es

Received: 10/12/22, Revised: 11/26/23, Accepted: 1/22/24, Published: 2/19/24

Abstract

This is the English version of the paper: *Complejidad de los números naturales*, Gac. R. Soc. Mat. Esp. **3** (2000), 230–250. The complexity $\|n\|$ of a natural number n is the least number of 1's needed to write an expression for n using only the addition $+$, the product $*$, the unit 1, and parentheses ( ). We study its main properties and end up formulating a set of conjectures. These conjectures are now mainly resolved by the work of Harry Altman. We include an appendix explaining these later results.

1. Introduction

Our purpose is to explore what seems to be a trivial question that may be understood by high school students in their early teens, but with very deep relationships.

Lately we have been interested in one of the mathematical problems that the author consider most important: the $\mathbf{P} \overset{?}{=} \mathbf{NP}$ problem. In this case the first difficulty is to explain the problem to a professional mathematician, say to an expert in analysis. This is not a minor issue, but the author think that the problem $\mathbf{P} \overset{?}{=} \mathbf{NP}$ may be reformulated as an inequality. Hence to explain the question adequately, so that it is understood by an expert in Analysis, such a reformulation is maybe the first step in the solution of the problem. The question we shall discuss here arose while trying to obtain this explanation.

We start with the main question: Given a natural number n, how many 1's are needed to write n? For example,

$$19 = 1 + (1+1)(1+1+1)(1+1+1)$$

so that nine 1's suffice to write 19. We shall say that the complexity of 19 is less than or equal to 9, and we shall write this as $\|19\| \leq 9$. Of course, the complexity

DOI: 10.5281/zenodo.10680389

of 19 will be the number of 1's in the most economical representation of 19. We only admit expressions with sums and products.

The first values of the complexity function may be easily computed,

$$1, 2, 3, 4, 5, 5, 6, 6, 6, 7, 8, 7, 8, 8, 8, 8, 9, 8, 9, 9, \dots .$$

We see that this is not a monotonic sequence: $8 = \|11\| > \|12\| = 7$.

When in our investigations we find any sequence of natural numbers, there is something we must do: look in *The On-Line Encyclopedia of Integer Sequences* (`http://oeis.org`) of Sloane and Plouffe [18]. In it we find this sequence and a reference to a paper by Guy [13] where it is defined and analyzed.

2. Complexity of a Natural Number

We have defined the complexity as a function $n \mapsto \|n\|$ of $\mathbb{N} \to \mathbb{N}$ such that for every pair of natural numbers m and n, we have

$$\|1\| = 1, \qquad \|m+n\| \leq \|m\| + \|n\|, \qquad \|m \cdot n\| \leq \|m\| + \|n\|.$$

In fact, it is the largest function satisfying these conditions (see Propositon 2). To prove this and other assertions it is useful to introduce the concept of expression.

Definition. (expression) An *expression* is a sequence of symbols. The allowed symbols are `x, +, ( )`. Not every sequence of these symbols is an expression. Examples of expressions are

`(x + x)`, `(x+(xx))`, `(x+((x+x)((x+(x+x))(x+(x+x)))))`.

The formal definition is inductive:

(a) The variable `x` is an expression.

(b) If `A` and `B` are expressions, then `(A+B)` and `(AB)` are also expressions.

(c) The only expressions are those obtained by repeated applications of rules (a) and (b).

We define the value of an expression `A` as the number $v(\mathtt{A})$ that results when replacing `x` by 1. Again we use induction to define the value of an expression: $v(\mathtt{x}) = 1$, and if `A` and `B` are expressions then, $v(\mathtt{(A+B)}) = v(\mathtt{A}) + v(\mathtt{B})$ and $v(\mathtt{(AB)}) = v(\mathtt{A})v(\mathtt{B})$.

Given an expression, we may define its complexity as the number of letters `x` it contains. For example $\|\mathtt{(x+(xx))}\| = 3$. Let $\mathcal{E}$ be the set of expressions. We may translate the definition of the complexity of n as

$$\|n\| = \inf\{\|\mathtt{A}\| : \mathtt{A} \in \mathcal{E} \text{ and } v(\mathtt{A}) = n\}.$$

If we want to compute the value of $\|n\|$ we may use the following proposition.

Proposition 1. *For each natural number $n > 1$,*

$$\|n\| = \min_{\substack{2\le d\le\sqrt{n},\ d|n\\ 1\le j\le n/2}} \{\|d\| + \|n/d\|, \quad \|j\| + \|n-j\|\}.$$

Proof. Let E be an optimal expression for n, i.e., one that gives its complexity $\|n\| = \|\mathtt{E}\|$. As an expression that is not x, we will have E=(A+B), or E=(AB). Let $a = v(\mathtt{A})$ and $b = v(\mathtt{B})$. Then either $n = a + b$ and $\|n\| = \|a\| + \|b\|$ or $n = ab$ and $\|n\| = \|a\| + \|b\|$. In the first case, if j is the least of a and b, we will have $1 \le j \le n/2$, and in the second case, if d is the least of a and b, then d will be a divisor of n with $2 \le d \le \sqrt{n}$. Of course for the reasoning to be valid we must check that if E is an optimal expression for n, then A and B must be optimal expressions for a and b, respectively. We leave this check to the reader. □

Using the above proposition and the mathematical software Mathematica, we have computed the values of $\|n\|$ for $1 \le n \le 200\,000$.

3. Bounds

Proposition 2. *Let $P\colon \mathbb{N} \to \mathbb{R}$ be a function satisfying*

$$P(1) = 1, \quad P(n+m) \le P(n) + P(m), \quad P(n\cdot m) \le P(n) + P(m).$$

Then for each $n \in \mathbb{N}$ we have $P(n) \le \|n\|$.

Proof. It is easy to see by induction that for each expression A, we have $P(v(\mathtt{A})) \le \|\mathtt{A}\|$. It is true for A = x, and, if it is true for A and B, then it is true for (A+B) and (AB). For example, for the product we have

$$P\big(v(\mathtt{(AB)})\big) = P(v(\mathtt{A})v(\mathtt{B})) \le P(v(\mathtt{A})) + P(v(\mathtt{B})) \le \|\mathtt{A}\| + \|\mathtt{B}\| = \|\mathtt{(AB)}\|,$$

and a similar argument is valid for the sum. (Observe that by the definition of v we have $v(\mathtt{(A+B)}) = v(\mathtt{A}) + v(\mathtt{B})$ and $v(\mathtt{(AB)}) = v(\mathtt{A})v(\mathtt{B})$.)

Now in $P(v(\mathtt{A})) \le \|A\|$ we take the minimum over all expressions A such that $n = v(\mathtt{A})$. In this way we get $P(n) \le \|n\|$. □

Corollary 1. *For each natural number n we have $\log_2(1+n) \le \|n\|$.*

Proof. It is sufficient to check the properties of $P(n) = \log_2(1+n)$. □

Later, in Corollary 3, we will obtain a better inequality.

3.1. Upper Bounds

Now we get an upper bound. To this end we define a new function $L\colon \mathbb{N} \to \mathbb{N}$.

Definition 1. We define the function L inductively:

(a) Let $L(1) = 1$.

(b) If p is a prime number, then $L(p) = 1 + L(p-1)$.

(c) If $n = p_1 p_2 \cdots p_k$ is a product of prime numbers (which may be repeated), then $L(p_1 p_2 \cdots p_k) = L(p_1) + L(p_2) + \cdots + L(p_k)$.

It is clear from this definition that if $n = ab$ with a and $b \geq 2$, then we will have $L(ab) = L(a) + L(b)$.

Proposition 3. *For each $n \in \mathbb{N}$ we have*

$$\|n\| \leq L(n).$$

Proof. We may prove this by induction. For $n = 1$ we have $\|1\| = L(1) = 1$. Assume that $\|k\| \leq L(k)$ for each $k < n$. There are two possibilities. If $n = p$ is a prime number

$$\|p\| \leq \|p-1\| + \|1\| = \|p-1\| + 1 \leq L(p-1) + 1 = L(p).$$

If n is composite $n = ab$ with a and $b > 2$,

$$\|n\| \leq \|a\| + \|b\| \leq L(a) + L(b) = L(ab) = L(n). \qquad \square$$

Proposition 4. *For each $n \geq 2$ we have*

$$L(n) \leq \frac{3}{\log 2}(\log n).$$

Proof. Since $L(2) = 2$ and $L(3) = 3$ the result is true for $n = 2$ and $n = 3$.

Assume now that $n > 3$ and that the Proposition is true for all natural numbers strictly less than n.

If $n = p$ is a prime number we have

$$L(p) = 1 + L(p-1) = 1 + 2 + L\Big(\frac{p-1}{2}\Big) \leq 3 + \frac{3}{\log 2}\log\Big(\frac{p-1}{2}\Big).$$

We want this to be

$$L(p) \leq \frac{3}{\log 2}(\log p).$$

Hence we must check that

$$3 \leq \frac{3}{\log 2}\log\Big(\frac{2p}{p-1}\Big),$$

which is easily proved for $p \geq 3$.

If $n = ab$ with a and $b \geq 2$, we have

$$L(ab) = L(a) + L(b) \leq \frac{3}{\log 2}(\log a) + \frac{3}{\log 2}(\log b) = \frac{3}{\log 2}(\log ab). \qquad \square$$

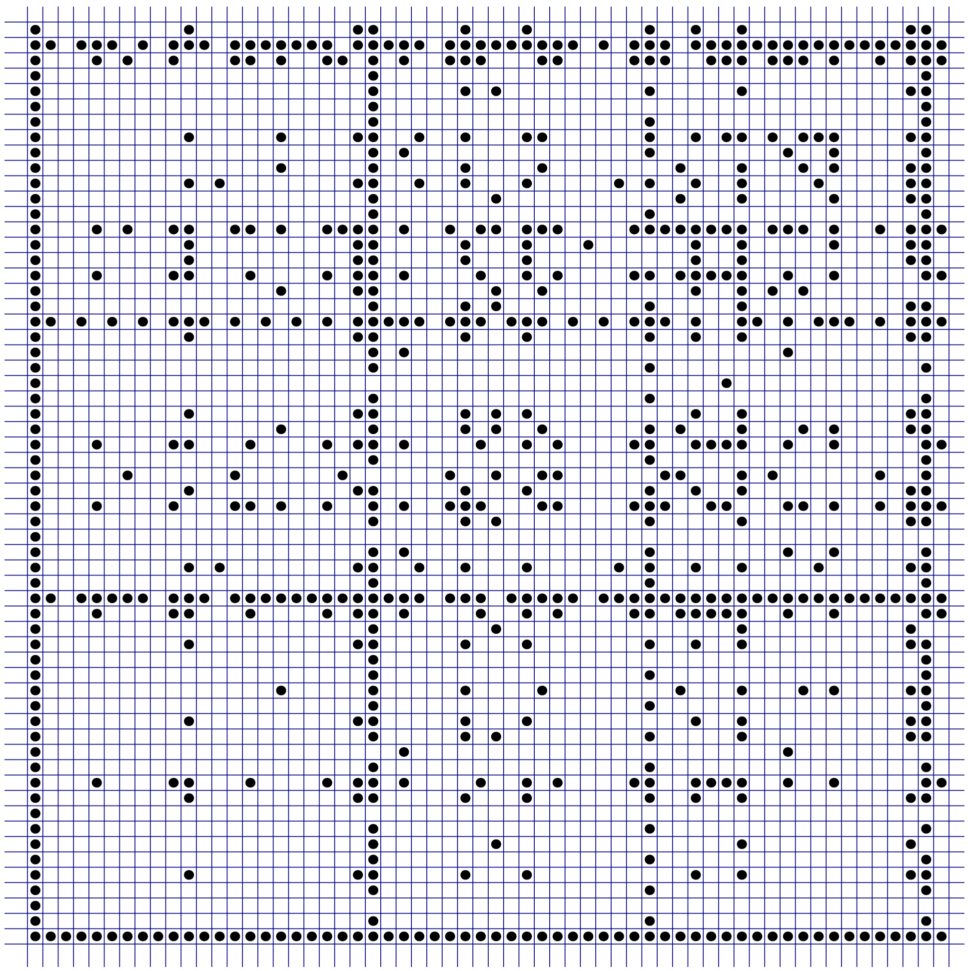

Figure 1: Bad Factors. (The lower left corner dot is at $(1, 1)$).

Remark 1. We do not know if the constant $3/\log 2$ in the above theorem is optimal. The proof makes one suspect that the quotient $L(n)/\log n$ may be large when $n = p_k$ is a prime such that there exists a sequence of primes $(p_j)_{j=1}^{k}$ with $p_{j+1} = 2p_j + 1$. For example, 89, 179, 359, 719, 1439, 2879 is such a sequence of prime numbers, and the maximum value of the quotient $L(n)/\log n$ that we know is

$$\frac{L(2879)}{\log 2879} = 3.766384578\cdots < 4.328085123\cdots = \frac{3}{\log 2}.$$

The main difference between the two functions $L(\cdot)$ and $\|\cdot\|$ is that $L(\cdot)$ is additive and $\|\cdot\|$ is not. For each pair of numbers n and m greater than 1 we have $L(mn) = L(m) + L(n)$. On the other hand there exist pairs n, m of numbers greater than 1 and such that $\|mn\| < \|m\| + \|n\|$. In such a case we shall say that $n \cdot m$ is a bad factorization.

In Figure 1 we put a dot at each point (n, m) such that $n \cdot m$ is a *bad factorization*. The figure contains all the factors n and $m \leq 60$.

The product $1 \cdot n$ is always a bad factorization. In Figure 3.1 we see some other surprising regularities. There are some conspicuous (vertical and horizontal) aligned points. Especially note the verticals at $n = 23$, 41, 59, which deserve an explanation.

These numbers, we may call them *bad factors*, appear to have great complexity. We define the *number with great complexity* n_k as the number n_k that is the least solution to $\|n\| = k$. The first values of this sequence are

$$1,\quad 2,\quad 3,\quad 4,\quad 5,\quad 7,\quad 10,\quad 11,\quad 17,\quad 22,\quad 23,\quad 41,\quad 47,\quad 59,$$
$$89,\quad 107,\quad 167,\quad 179,\quad 263,\quad 347,\quad 467,\quad 683,\quad 719,\quad 1223,$$
$$1438,\quad 1439,\quad 2879,\quad 3767,\quad 4283,\quad 6299,\quad 10079,\quad 11807,$$
$$15287,\quad 21599,\quad 33599,\quad \dots .$$

This sequence appears in [18] `OEIS:A005520`. In this way we find the reference to Rawsthorne [17].

4. Mean Values

There is another proof of $\|n\| \leq 3\log n/\log 2$. We observe that if we write n in binary $n = \sum_{j=0}^{k-1} \varepsilon_j 2^j + 2^k$ we have a means to express n:

$$n = \varepsilon_0 + 2(\varepsilon_1 + 2(\varepsilon_2 + \cdots + 2(\varepsilon_{k-2} + 2(\varepsilon_{k-1} + 2))\cdots)).$$

If we substitute each 2 by $1 + 1$ and observe that each ε_j is equal to 0 or 1, we have an expression for n that uses at most $2k + k$ ones, and where k is determined by $2^k \leq n < 2^{k+1}$. It follows that $\|n\| \leq 3\log n/\log 2$.

The above reasoning proves that the function $L_2(n) = 2k + \varepsilon_0 + \varepsilon_1 + \cdots + \varepsilon_{k-1}$ is another upper bound for $\|n\|$. The relationship between $L_2(n)$ and $L(n)$ is not very simple. Amongst the first 1000 numbers we generally have $L(n) \leq L_2(n)$ but this inequality has exceptions. The first one is $L_2(161) = 16 < 17 = L(161)$. In this range the difference is small.

The function $L_2(n)$ allows us to obtain information about $\|\cdot\|$. Consider the numbers n that in binary take the form $1\varepsilon_{k-1}\cdots\varepsilon_0$, i.e., numbers than in binary have $k+1$ digits. By the above expression we have

$$\|n\| \le 2k + \varepsilon_0 + \cdots + \varepsilon_{k-1}.$$

We may suppose that the ε_k are independent random variables with mean $1/2$. The inequality of Chernoff (see [11] or [1] for a simple exposition) says that

$$\mathbb{P}\Big(\Big|\sum \varepsilon_j - k/2\Big| < x\sqrt{k}\Big) \ge 1 - 2e^{-2x^2}.$$

It follows that $\mathbb{P}(\|n\| \le 2k + k/2 + x\sqrt{k}) \ge 1 - 2e^{-2x^2}$, and taking $x = \sqrt{\log k}$ we get

$$\mathbb{P}\Big(\|n\| > 5k/2 + \sqrt{k\log k}\Big) \le 2k^{-2}.$$

Hence between the 2^k values of n with $2^k \le n < 2^{k+1}$ at most $(2/k^2)2^k$ satisfy $\|n\| > 5k/2 + \sqrt{k\log k}$. The other ones, most of them, satisfy

$$\|n\| \le \frac{5k}{2} + \sqrt{k\log k} = \frac{5}{2}\frac{\log n}{\log 2} + O(\sqrt{\log n \log\log n}).$$

Therefore, for almost all large values of n we have

$$\|n\| \le \frac{5}{2}\frac{\log n}{\log 2} + O(\sqrt{\log n \log\log n}).$$

The upper bound $L(n)$ is very good for small values of n. For example for the first 220 values of n, $L(n) = \|n\|$, except for the values in the following table.

n	$\|n\|$	$L(n)$
46	12	13
47	13	14
55	12	13
82	13	14
83	14	15
92	14	15
94	15	16
110	14	15

n	$\|n\|$	$L(n)$
115	15	16
118	15	16
121	15	16
138	15	16
139	16	17
141	16	17
145	15	16
161	16	17

n	$\|n\|$	$L(n)$
164	15	16
165	15	16
166	16	17
167	17	18
184	16	17
188	17	18
217	16	17
220	16	17

In these cases the bound $L_2(n)$ is equal or greater than $L(n)$, except for the case $n = 161$.

The two functions $L(n)$ and $\|n\|$ coincide in 771 values of n in the range $1 \le n \le 1000$, the difference being equal to 1 for the 229 other values in this range with a few exceptions.

5. Particular Values

Numbers with Small Complexity. A good lower bound for $\|n\|$ is obtained from the knowledge of the largest number we may write with m ones. That is, given m, we want to find which is the largest natural number N with $\|N\| = m$. The answer roughly is that we must group the m ones in groups of three and multiply them. To show this we define the concept of *extremal expression*. Let $\mathtt{M}_m$ be an expression with $\|\mathtt{M}_m\| = m$ (that is $\mathtt{M}_m$ is formed with m symbols x and the operations of sum and product), and such that its value $v(\mathtt{M}_m)$ is the maximum of all the expression with m ones, i.e.,

$$N = v(\mathtt{M}_m) = \sup_{\|\mathtt{A}\|=m} v(\mathtt{A}).$$

We say that such an expression $\mathtt{M}_m$ is *extremal*.

In the above situation $\|N\| = m$. In fact, since $N = v(\mathtt{M}_m)$ and $\|\mathtt{M}_m\| = m$, we have $\|N\| \le m$. Assume, by contradiction, that $\|N\| < m$. Then there will exist an expression B such that $v(\mathtt{B}) = N$ and $\|\mathtt{B}\| = \|N\| < m$. Let d be such that $m = d + \|\mathtt{B}\|$. We may construct an expression C such that $\mathtt{C} = \mathtt{B} + \mathtt{x} + \cdots + \mathtt{x}$ and such that $\|\mathtt{C}\| = \|\mathtt{B}\| + d = m$ and $v(\mathtt{C}) = v(\mathtt{B}) + d > N$. This contradicts the definition of $\mathtt{M}_m$.

It is easy to see that the following expressions are extremal

$$\mathtt{M}_1 = \mathtt{x}, \quad \mathtt{M}_2 = \mathtt{(x + x)}, \quad \mathtt{M}_3 = \mathtt{(x + (x+x))},$$
$$\mathtt{M}_4 = \mathtt{(x+x)(x+x)}, \quad \mathtt{M}_5 = \mathtt{(x+(x+x))(x+x)}, \dots .$$

We see that given m, the extremal expression $\mathtt{M}_m$ is not unique. For example, for $m = 4$ the expression $\mathtt{M}_4 = \mathtt{(x+(x+(x+x)))}$ is another possibility.

We shall use here a notation that is not very precise. For example, we shall write $\mathtt{M}_3^a\mathtt{M}_2$ to denote any expression having this form, not defining how the product is constructed from its factors. So, $\mathtt{M}_3^4$ denotes any of the expressions $\mathtt{((M_3M_3)(M_3M_3))}$, $\mathtt{(M_3(M_3(M_3M_3)))}$ or any other form of grouping the factors.

Proposition 5. *Let $\mathtt{M}_2 = \mathtt{(x + x)}$, $\mathtt{M}_3 = \mathtt{(x + (x+x))}$ and $\mathtt{M}_4 = \mathtt{(x+x)(x+x)}$. For $n > 1$, the expressions*

$$\mathtt{M}_n = \begin{cases} \mathtt{M}_3^k & \text{if } n = 3k, \\ \mathtt{M}_3^{k-1}\mathtt{M}_4 & \text{if } n = 3k+1, \\ \mathtt{M}_3^k\mathtt{M}_2 & \text{if } n = 3k+2, \end{cases}$$

are extremal.

Proof. We may check the proposition for $n = 2$, 3 and 4 directly.

We assume the assertion is true for all $s < n$ and try to prove it for $n \ge 5$. Certainly there is one extremal expression K with $\|\mathtt{K}\| = n$. Then there are two expressions $\mathtt{A}$ and $\mathtt{B}$ such that $\mathtt{K} = (\mathtt{A}+\mathtt{B})$ or $\mathtt{K} = (\mathtt{AB})$. $\mathtt{A}$ and $\mathtt{B}$ are extremal expressions because $\mathtt{K}$ is extremal. We may replace $\mathtt{A}$ and $\mathtt{B}$ by extremal expressions of the same complexity and value, and the resulting expression $\mathtt{K}'$ will also be extremal. Hence, without loss of generality, we may assume, using the induction hypothesis, that $\mathtt{A}$ and $\mathtt{B}$ are of the form given in the proposition or that $\mathtt{A} = \mathtt{x}$ and $\mathtt{B}$ is as in the proposition.

The case $\mathtt{K} = (\mathtt{A}+\mathtt{B})$ is only possible if $v(\mathtt{A})$ or $v(\mathtt{B}) = 1$, because, in other cases, the expression $(\mathtt{AB})$ contradicts the extremality of $\mathtt{K}$. But $\mathtt{K} = (\mathtt{x}+\mathtt{M}_3^k)$, $\mathtt{K} = (\mathtt{x}+\mathtt{M}_3^{k-1}\mathtt{M}_4)$, or $\mathtt{K} = (\mathtt{x}+\mathtt{M}_3^k\mathtt{M}_2)$ are impossible with $n \geq 5$. These expressions are clearly not extremal. Compare with $\mathtt{M}_3^{k-1}\mathtt{M}_4$, $\mathtt{M}_3^k\mathtt{M}_2$ or $\mathtt{M}_3^{k+1}$, respectively.

Therefore $\mathtt{K} = (\mathtt{AB})$ where $\mathtt{A}$ and $\mathtt{B}$ are like those in the proposition. Some of the combinations are not possible: for example $\mathtt{A} = \mathtt{M}_3^k\mathtt{M}_2$ and $\mathtt{B} = \mathtt{M}_3^{j-1}\mathtt{M}_4$ are not possible since $\mathtt{M}_3^{k+j-1}\mathtt{M}_4\mathtt{M}_2$ is improved by $\mathtt{M}_3^{k+j+1}$ and $\mathtt{K}$ will not be extremal. A case by case analysis proves that $\mathtt{K}$ is one of the three forms in the proposition. □

Corollary 2. *For $a = 0$, 1, or 2 and all $b \in \mathbb{N}$ we have*

$$\|2^a 3^b\| = 2a + 3b, \qquad a = 0, 1, 2.$$

All natural numbers $n > 1$ may be written in a unique way as $n = 2a + 3b$ with $a = 0$, 1 or 2. In this case $2^a 3^b$ is the greatest number m with $\|m\| = n$. Hence $m > 2^a 3^b$ implies $\|m\| > 2a + 3b$.

We define g by

$$g(n) = \begin{cases} 3a & \text{if } n \in [3^a, 3^a + 3^{a-1}), \\ 3a+1 & \text{if } n \in [3^a + 3^{a-1}, 2\cdot 3^a), \\ 3a+2 & \text{if } n \in [2\cdot 3^a, 3^{a+1}), \end{cases}$$

we then have $g(n) \leq \|n\|$ for each n.

Corollary 3. *For any $n \geq 2$ we have*

$$3\frac{\log n}{\log 3} \leq \|n\| \leq L(n) \leq 3\frac{\log n}{\log 2}.$$

Proof. We only need to prove the first inequality. If $n = 3^a$, we see directly that the inequality is true. If $x \in (3^a, 3^a + 3^{a-1}]$, we have $\|x\| \geq 3a + 1$. Then

$$\|x\| \geq \|3^a\| + 1 = 3a + 1 \geq 3\frac{\log(4\cdot 3^{a-1})}{\log 3} \geq 3\frac{\log x}{\log 3}.$$

Similarly, for $x \in (4\cdot 3^{a-1}, 2\cdot 3^a]$ we have

$$\|x\| \geq \|4\cdot 3^{a-1}\| + 1 \geq 3a + 2 \geq 3\frac{\log(2\cdot 3^a)}{\log 3}.$$

Finally, for $x \in \big(2\cdot 3^a, 3^{a+1}\big]$, we only need to check that

$$\|x\| \geq \|2\cdot 3^a\| + 1 = 3a + 3 \geq 3\frac{\log(3^{a+1})}{\log 3}.$$ □

6. The Problem $\mathbf{P} \stackrel{?}{=} \mathbf{NP}$ and the Complexity of the Natural Numbers

Before explaining the problem we must describe the classes **P** and **NP**. Consider a finite alphabet A, and let A^* be the set of *words*, that is, the set of finite sequences of elements of A.

We call a subset $S \subset A^*$ a *language* . We say that S is in the *class* **P** if there is an algorithm T and a polynomial $p(t)$ such that with a word x as input, T gives an output $T(x)$, such that $T(x) = 1$ if $x \in S$ and $T(x) = 0$ if $x \notin S$. Also, T gives the output $T(x)$ in a time bounded by $p(|x|)$ (here $|x|$ denotes the length of the word x). We then say that T is a *polynomial algorithm.* In a few words we may say that **P** is the class of languages recognizable in polynomial time. It is important to notice that this concept is very stable with respect to the diverse definitions of what is an algorithm, how we compute the "time" that the algorithm T takes to give the output, or even if we consider the same language in a different alphabet (as when we consider a set of natural numbers written in different basis). In other words, the concept does not change if we give proper definitions of these concepts.

The *class* **NP** consists of the languages recognized by *non deterministic polynomial algorithms.* That is $S \subset A^*$ is in **NP** if there exists an algorithm T and a polynomial $p(x)$ such that for each $x \in S$ there is $y \in A^*$ with $|y| \le p(|x|)$ and such that with the input (x, y) the algorithm gives the output $T(x, y) = 1$ in time bounded by $p(|x|)$. On the other hand, if $x \notin S$ we have $T(x, y) = 0$ for all y with $|y| \le p(|x|)$.

We say that in this case T is a *non-deterministic algorithm* since to obtain $x \in S$ we must first choose y. If we know which y to use, then this process is fast, but if we do not know y, we may try each possible y, but this will need a time greater than or equal to $|A|^{p(|x|)}$ which in practice is impossible.

Again, the class **NP** is very stable with respect to possible changes in the definitions. Also many practical problems are in this class.

It is easy to check that $\mathbf{P} \subset \mathbf{NP}$. The question is whether these two classes are the same. To understand a bit more of the difficulty, observe the following.

Our experience as mathematicians teaches us that to understand a proof, or better yet, to check the correctness of a proof, is a task of type **P**. That is, the time needed is proportional to the length of the proof.

On the other hand, to determine if a conjecture x is a Theorem we need first to write the proof y and then to apply the above procedure to check the correctness of the pair (x, y). The set of all theorems is not in the class **NP** since we know the length of the proof $|y|$ is not bounded by the length of the theorem x, that is, $|y| \not\le p(|x|)$. But for each polynomial $p(t)$, the following set is in **NP**:

$$\mathcal{T}_p = \{x : x \text{ is a theorem with a proof of length } \le p(|x|)\}.$$

Maybe someone finds these definitions rather vague, but the formal logic allows one to make things precise.

If $\mathbf{P} = \mathbf{NP}$ and the proof were sufficiently constructive (technically, that we can find a polynomial algorithm for an **NP**-complete problem), then there would exist a polynomial algorithm that would allow us not only to decide if $x \in \mathcal{T}_p$, but also to find in this case a proof for x in polynomial time. The mathematicians would not be needed any more.

When one recalls the achievements of the 20^{th} century – proof of Fermat's theorem, classification of finite simple groups, pointwise convergence of Fourier series of function in L^p, Riemann's hypothesis for algebraic varieties over fields of characteristic p, independence of continuum hypothesis, and many more – one gets the impression that there exists an algorithm to decide $x \in \mathcal{T}_p$ by searching directly for a proof, not by trial and error. This algorithm consists of taking promising students, giving them the possibility to travel and speak with specialists on the topic in question, letting them try to solve analogous questions, study the solution of related problems, and so on . . . .

7. Connection of the Complexity of Natural Numbers and the Problem $\mathbf{P} \stackrel{?}{=} \mathbf{NP}$

Consider the assertion $\|4787\| = 28$. We may decompose it in two parts. The first, $\|4787\| \leq 28$, has a very easy proof:

$$4787 = 2 + 3(2 + 3^2)(1 + 2^4 3^2). \qquad (\star)$$

The other part of the assertion $\|4787\| \geq 28$, has a much more laborious proof. At present we do not know any way, other than computing the values of $\|n\|$ for all $n \leq 4787$, a task that, on a personal computer, took several hours. Of course this does not imply that it is easy to find a proof as in $(\star)$.

Consider the sets

$$A = \{(n, c) \in \mathbb{N}^2 : \|n\| \leq c\}, \qquad B = \{(n, c) \in \mathbb{N}^2 : \|n\| > c\}.$$

The fact, as we have remarked, that if $(n, c) \in A$, then there is a relatively short proof of it, shows us that A is in the class **NP**.

Roughly, a set A is in **NP**, if to prove that $x \in A$ an exhaustive search is required. This search in principle is exponential in the size of x. Nevertheless, once the proof has been found, it is easily recognized. Easily meaning in polynomial time with respect to the size of x. Complete information may be found in the book [12]. These problems bring to mind the one of finding a needle in a haystack. Once we have found the needle there is no doubt that the task is done, but at first it appears unreachable since the straw is so similar to the needle that we do not see any means other than search methodically.

The core of the problem $\mathbf{P} \stackrel{?}{=} \mathbf{NP}$ is whether in situations where there exists a short proof, there is always a direct path to find it. If $\mathbf{P} = \mathbf{NP}$, then there is always a direct path to the proof without hesitations. At first sight this appears a wild assumption, but the rigorous proof of $\mathbf{P} \neq \mathbf{NP}$ eludes us still after twenty seven years of study.

Recently Microsoft has funded an investigation center and has contracted Michael Friedman (Fields medal in 1986). Friedman has the intention of trying to solve the question $\mathbf{P} \stackrel{?}{=} \mathbf{NP}$. Microsoft will invest 2.6 million dollars each year in this program.

It appears that $\mathbf{P} = \mathbf{NP}$ is false, but not all is so simple. Sometimes tasks that appear to need an exhaustive search have been proved simple. We shall give an example.

Let $\mathcal{C} \subset \mathbb{N}$ be the set of composite numbers. At first sight it appears that the only means to prove that n is composite is to divide n by each number $m \leq \sqrt{n}$ and check if some remainder equals 0. The size of n is of the order of the number of digits needed to write it, i.e., of the order $\log n$. The number of needed checks maybe $\sqrt{n} = e^{(\log n)/2}$, which grows exponentially with $\log n$. And if really n is composite there is a short proof: to exhibit a proper divisor d of n. That is $\mathcal{C}$ is in the class $\mathbf{NP}$.

But it is not so difficult to decide whether n is composite. If n is prime and b is prime with n we have $b^{n-1} \equiv 1 \pmod{n}$. An idea somewhat more elaborate: if n is a prime and $n-1 = 2^s t$, then in the sequence of the rests of b^t, $b^{2t}, \dots, b^{2^s t}$ modulo n, the last different from 1 must be -1. In the other case it is certain that n is composite. This is the famous Miller-Rabin test. It is known that if the generalized Riemann hypothesis is true, then if n is composite, the test of Miller-Rabin is not satisfied for some $b < 2(\log n)^2$. Hence, under the mentioned hypothesis, we have a fast algorithm (polynomial) to decide whether n is composite: to do the test of Miller-Rabin for all $b < 2(\log n)^2$.

Another incentive to pose the problem $\mathbf{P} \stackrel{?}{=} \mathbf{NP}$ is the existence of $\mathbf{NP}$-*complete problems*. That is, sets $B \subset \mathbb{N}$ such that B is in the class $\mathbf{NP}$ and, for which from $B \in \mathbf{P}$ it follows that $\mathbf{P} = \mathbf{NP}$.

Since Euclid's times, mathematicians have had a clear concept of an algorithm. Turing goes a step further, and by an effort of introspection, gives us a precise definition. Turing's mental image is that of a mathematician, notebook in hand, computing. By abstracting the procedure, Turing created the idea of a modern computer. Starting from Turing's definitions, it is possible to quantify the time a computer will spend on a given task, giving a precise definition of the classes $\mathbf{P}$ and $\mathbf{NP}$.

The first connection of the complexity of the natural numbers with the problem $\mathbf{P} \stackrel{?}{=} \mathbf{NP}$ is the fact that $\mathbf{P} = \mathbf{NP}$ implies the existence of a fast algorithm to compute $\|n\|$. There will be constants C and $k \in \mathbb{N}$ and an algorithm that will compute $\|n\|$ in time less than or equal to $C(\log n)^k$.

8. Complexity of Boolean Functions

There is another connection, this time structural, between the complexity of natural numbers and the problem $\mathbf{P} \stackrel{?}{=} \mathbf{NP}$. To explain this connection we must define a related concept, that of the complexity of a Boolean function.

The set $\{0,1\}$ is a field when we consider the composition laws sum and product modulo 2. For each number n let $\mathcal{F}_n$ be the set of functions $f\colon \{0,1\}^n \to \{0,1\}$. The set $\mathcal{F}_n$ is a ring if we take sum and product with respect to the field in the image $\{0,1\}$.

For example, consider the constant functions $\mathbf{1}$, $\mathbf{0}$ and the components π_j defined by $\pi_j(\mathbf{x}) = \pi_j(x_1, x_2, \dots, x_n) = x_j$. The ring $\mathcal{F}_n$ is generated by these functions, i.e., we may write any function $f \in \mathcal{F}_n$ as a polynomial of the above functions. To see this, given $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n) \in \{0,1\}^n$, we define the function $f_\varepsilon = \prod_j(\delta_j + \pi_j)$, where, for each j, $\delta_j = 1 + \varepsilon_j$. Then $f_\varepsilon(\mathbf{x}) = 0$, except for $\mathbf{x} = \varepsilon$. Hence, any function g may be written

$$g = \sum_{\varepsilon \in S} f_\varepsilon,$$

where S is the set of ε such that $g(\varepsilon) = 1$.

As in the case of the natural numbers, we may define the complexity of the elements of $\mathcal{F}_n$. It will be the greatest function $f \mapsto \|f\|$ such that

$$\|\mathbf{0}\| = \|\mathbf{1}\| = 0, \quad \|\pi_j\| = 1, \quad \|f+g\| \le \|f\| + \|g\|, \quad \|fg\| \le \|f\| + \|g\|.$$

For any $\theta \in (0,1)$, most of the elements of $\mathcal{F}_n$ have complexity greater than or equal to $2^{\theta n}$. The proof of this result is done by counting how many elements have complexity k, say a_k. It is easy to see that $a_0 = 2$, $a_1 = 2n$. From f and g with $\|f\| = j$ and $\|g\| = k - j$ we get, at most, four elements with complexity less than or equal to k. They are $f+g$, fg, $1+f+g$, $1+fg$. With these observations we get $a_k \le 4(a_1 a_{k-1} + a_2 a_{k-2} + \cdots + a_{k-1}a_1)$. It follows that $a_k \le A_k$, where A_k is defined by

$$A_0 = 2, \quad A_1 = 2n, \quad A_k = 4\sum_{j=1}^{k-1} A_j A_{k-j}.$$

From this definition we get

$$\sum_{k=0}^{\infty} A_k x^k = \frac{17 - \sqrt{1 - 32nx}}{8}, \qquad A_k = \frac{1}{2(2k-2)}\binom{2k-2}{k}(8n)^k.$$

Hence

$$a_k \le A_k \sim \frac{2^{5k}}{8\sqrt{2\pi}k^{3/2}} n^k.$$

Therefore, for x large,

$$\sum_{k=0}^{x} A_k \le c\sum_{k=0}^{x}(32n)^k \le c'(32n)^x \le Ae^{Bx\log n},$$

and hence if $x < 2^{\theta n}$, with $0 < \theta < 1$, we get

$$\sum_{k=0}^{x} A_k \ll \operatorname{card}(\mathcal{F}_n) = 2^{2^n},$$

proving our assertion.

Each construction of $f(x_1, \dots, x_n)$ as a polynomial allows one to prove an assertion of type $\|f\| \le a$. But from the polynomial expression we may get something more practical: a circuit that allows to compute $f(x_1, \dots, x_n)$ starting from the inputs x_j.

As in the case of natural numbers, it is difficult to prove inequalities of type $\|f\| > a$. In fact the situation is surprising: we have seen that in the set of functions with n variables, the complexity is usually larger than $2^{\theta n}$. Hence one would expect to have an easy task in defining a sequence of functions (f_n), where f_n depends on n variables and such that $\|f_n\| > 2^{\theta n}$. On the contrary it has only been achieved that $\|f_n\| > p(n)$, where p is a polynomial of small degree (see [19], [14]). The problem here is not to prove that there exist sequences with $\|f_n\| > 2^{\theta n}$, which, as we have seen is easy, but to define explicitly a concrete sequence of functions for which this is so. When we speak of "define explicitly" we refer to a technical concept that needs some explanation. We must exclude easy solutions such as *let* f_n *the first function of* n *variables with maximum complexity*. We say that (f_n) is given explicitly if there is an algorithm that computes the value of $f_n(x_1, \dots, x_n)$ in a reasonable time.

The problem $\mathbf{P} \stackrel{?}{=} \mathbf{NP}$ induces one to consider a special sequence of Boolean functions. Let a be a natural number and consider $n = \binom{a}{2}$ the number of pairs. Our variables will be

$$x_{12},\ x_{13},\ x_{23},\ x_{14},\ x_{24},\ x_{34},\ \dots,\ x_{1a},\ x_{2a},\ \cdots,\ x_{a-1\,a}.$$

In this way, each set of values of these variables in $\{0,1\}^n$ may be seen as a graph with a vertices, and where $x_{jk} = 1$ if and only if the vertices j and k are connected by an edge of the graph. For each $b \le a$, let $f_b^a(x_{12}, \dots, x_{a-1\,a})$ be the function that is equal to 1 if and only if there is a set of b vertices such that all of them are connected in the graph.

It is plausible that $\|f_b^a\| \ge \binom{a}{b}$, since to compute the value of f_b^a in a given graph we need to check each set of b vertices. It can be shown that, if this is so, then $\mathbf{P} \ne \mathbf{NP}$. In this way, to prove $\|f_b^a\| \ge \binom{a}{b}$, may be, the most promising path to solve the $\mathbf{P} \stackrel{?}{=} \mathbf{NP}$ question.

In the case of the complexity of natural numbers, an analogous question is the following. Posed by Guy [13]:

Problem 1. Is there a sequence of natural numbers (a_n) such that

$$\lim_{n\to\infty} \frac{\|a_n\|}{\log a_n} > \frac{3}{\log 3}? \tag{1}$$

A good candidate is the sequence 2^n. All computed values satisfy $\|2^n\| = 2n$. Selfridge asks (see [13]) whether there exists any n with $\|2^n\| < 2n$.

If for some n and k we would have $2^n = 3^k$ (which is clearly impossible), the second expression would give us $\|2^n\| < 2n$. Of course the advantage would be greater for big n than for small n. Although the above is impossible, maybe another type of equality would yield $\|2^n\| < 2n$. For example, if for some n, 2^n written in base 3 has small digits. Again, this is unlikely but not impossible. Also, there may exist another type of expression for 2^n. The question here is whether a number of the form $(1+1)(1+1)\cdots(1+1)$ may be written in some way with fewer 1's. We almost have a trivial example $4 = (1+1)(1+1) = 1+1+1+1$. Here we have the same number of 1's so that we call it an almost-example. Maybe there are non-trivial almost-examples, for example

$$2^{27} = 1 + (1+2\cdot 3)(1+2^3\cdot 3^2)(1+2^9\cdot 3^3(1+2\cdot 3^2)).$$

If we replace each 2 by $1+1$ and each 3 by $1+1+1$ we get an expression for 2^{27} with 57 ones, in which the multiplicative structure of 2^{27} is not used.

The above equality proves that $\|2^{27}-1\| \le 56$. In spite of an intense search we have not found an $n > 2$ such that $\|2^n - 1\| < 2n - 1$, but we think this may happen.

The evidence appears to be in favor of the existence of a sequence that satisfies (1). For example, we may look at Figure 2. There we have put a little disk with center at each point $(n, \|n\|)$ with $1 \le n \le 2000$ and also we have drawn the smooth curves that bound $\|n\|$, i.e., $3(\log t)/\log 3$ and $3(\log t)/\log 2$, and also the curve $5\log t/2\log 2$. The points overlap and we see some lines parallel to the x-axis. We see that the upper bound appears to be bad and that apparently $\|n\| \le 5\log n/2\log 2$, whereas in reality we have only proved that this inequality is true for almost all $n \in \mathbb{N}$.

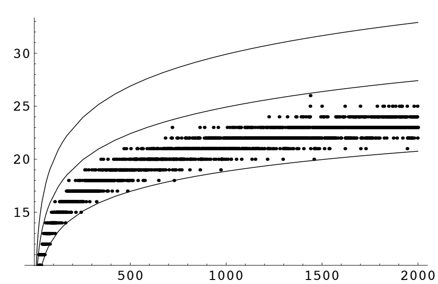

Figure 2: Graph of $\|n\|$.

But this figure says nothing about the limit $\lim \|n\|/\log n$, in which we are interested in. We only see that for the first 2000 values of n this sequence is bounded by the limits $5/2\log 2$ and $3/\log 3$.

9. Conjectures

We have computed, using Proposition 1, the complexity of the first 200 000 natural numbers. Looking at these numbers, one sees many regularities. We will call them conjectures about the behavior of the function $\|\cdot\|$, although we do not have much confidence that they persist for larger numbers.

The following conjectures were derived from tables such as Table 1. In this table we have written in columns the numbers with complexity $3n$ ($n = 1, 2, \ldots, 8$), written in base 3 and in decreasing order.

The first observation $\|3n\| = 3 + \|n\|$ is wrong. For example, $\|107\| = 16$ and $\|321\| = \|1 + 2^6 5\| = 18$. But the following conjectures seem to be true.

3	6	9	12	15	18	21	24
10	100	1000	10000	100000	1000000	10000000	100000000
	22	220	2200	22000	220000	2200000	22000000
	21	210	2101	21010	210100	2101000	21010000
		202	2100	21000	210000	2100000	21000000
		201	2020	20200	202000	2020000	20200000
		$\overline{122}$	2010	20100	201000	2010000	20100000
			2002	20020	200222	2002220	20022200
			2001	20010	200200	2002000	20020000
			$\overline{1221}$	20002	200100	2001000	20010000
			1220	20001	200020	2000200	20002000
			1212	$\overline{12221}$	200010	2000100	20001000
			1211	12210	200002	2000020	20000200
			1201	12200	200001	2000010	20000100
			1122	12122	$\overline{122210}$	2000002	20000020
			1121	12120	122100	2000001	20000010
			1112	12111	122000	$\overline{1222100}$	20000002
				12110	121220	1221000	20000001
				12102	121200	1220000	$\overline{12221000}$
				12101	121121	1212200	12210000
				12012	121110	1212000	12200000
				12010	121100	1211210	12122000
				12001	121022	1211100	12121201
				11221	121020	1211000	12120000

Table 1: Numbers with complexity $3n$ written in base 3

Conjecture 1. For each natural number n, there is an integer $a \geq 0$ such that $\|3^j n\| = 3(j-a) + \|3^a n\|$ for each natural number $j \geq a$.

Let us define the set $A = \{n \in \mathbb{N} : \|3^j n\| = 3j + \|n\| \text{ for all } j\}$[1].

Conjecture 2. For each pair of natural numbers p and q, there exists $a \geq 0$ such that, for $j \geq a$, we have $\|p(q3^j + 1)\| = 3j + 1 + \|p\| + \|q\|$.

The main observation in Table 1 is that the greatest numbers with complexity $3n$ are those natural numbers contained in the sequence $(3^n a_n)$, where a_n is given by

$$1, \frac{2(3+1)}{3^2}, \frac{2^6}{3^4}, \frac{2\cdot 3+1}{3^2}, \frac{2(3^2+1)}{3^3}, \frac{2\cdot 3^2+1}{3^3}, \frac{2^9}{3^6}, \frac{2(3^3+1)}{3^4}, \frac{2\cdot 3^3+1}{3^4}, \dots, \frac{2(3^k+1)}{3^{k+1}}, \frac{2\cdot 3^k+1}{3^{k+1}}, \dots.$$

Conjecture 3. There exist three transfinite sequences $(a_\alpha)_{\alpha<\xi}$, $(b_\alpha)_{\alpha<\xi}$, $(c_\alpha)_{\alpha<\xi}$ of rational numbers, such that the (greatest) numbers of complexity $3n$ (respectively $3n+1$, $3n+2$) are the (first) natural numbers contained in the sequence $(3^n a_\alpha)$, (resp. $(3^n b_\alpha)$, $(3^n c_\alpha)$).

The ordinal ξ is an infinite numerable ordinal such that $\omega\xi = \xi$.

These sequences start in the following way:

$$\begin{array}{llllll}
(a_\alpha), & 1, \frac{8}{9}, \frac{64}{81}, \frac{7}{9}, \frac{20}{27}, \dots \to \frac{2}{3} & \frac{160}{243}, \frac{52}{81}, \dots \to \frac{16}{27} & \frac{1280}{2187}, \frac{140}{243}, \dots \to \frac{5}{9} & \dots \\
(b_\alpha), & \frac{4}{3}, \frac{32}{27}, \frac{10}{9}, \frac{256}{243}, \frac{28}{27}, \dots \to 1 & \frac{80}{81}, \frac{26}{27}, \dots \to \frac{8}{9} & \frac{640}{729}, \frac{70}{81}, \dots \to \frac{64}{81} & \dots \\
(c_\alpha), & 2, \frac{16}{9}, \frac{5}{3}, \frac{128}{81}, \frac{14}{9}, \dots \to \frac{4}{3} & \frac{320}{243}, \frac{35}{27}, \dots \to \frac{32}{27} & \frac{95}{81}, \frac{2560}{2187}, \dots \to \frac{10}{9} & \dots
\end{array}$$

where the dots indicate infinite sequences, and where the indicated limits are not terms of the sequences.

Conjecture 4. The three sequences are decreasing. The denominators of each term a_α, b_α or c_α are powers of 3.

Conjecture 5. The numbers of the sequence (a_α) are the numbers of the set

$$\left\{\frac{n}{3^{\|n\|/3}} : \|n\| \equiv 0 \mod 3 \quad \text{and} \quad n \in A\right\},$$

ordered decreasingly.

[1] With the notation that was introduced after this was published, A is the set of *stable numbers*.

Conjecture 6. The numbers of the sequence (b_α) are the numbers of the set

$$\Big\{\frac{n}{3^{(\|n\|-1)/3}} : \|n\| \equiv 1 \mod 3 \quad \text{and} \quad n \in A\Big\},$$

ordered decreasingly.

Conjecture 7. The numbers of the sequence (c_α) are the numbers of the set

$$\Big\{\frac{n}{3^{(\|n\|-2)/3}} : \|n\| \equiv 2 \mod 3 \quad \text{and} \quad n \in A\Big\},$$

ordered decreasingly.

The following conjectures are more doubtful. They are only based on a few cases.

Conjecture 8. For all ordinals $\beta < \xi$ we have

$$\lim_{n\to\infty} a_{\omega\beta+n} = c_\beta/3, \quad \lim_{n\to\infty} b_{\omega\beta+n} = a_\beta, \quad \lim_{n\to\infty} c_{\omega\beta+n} = b_\beta.$$

This is the basis of the assertion about the value of ξ, which appears to be at least $\xi = \omega^\omega$, since this is the least solution of $\omega\xi = \xi$.

The following assertions, along with conjecture 8, allow us to predict, with some accuracy, the values of the transfinite sequences.

Conjecture 9. The numbers of the sequence $b_{\omega\beta+n}$ that converges to $a_\beta = b/3^a$ (with $\|b\| = 3a$) are numbers from the sequences

$$\frac{p(q3^j+1)}{3^{a+j}}, \quad \text{where} \quad b = pq, \text{ and, } \|p(q3^j+1)\| = 3a+3j+1,$$

and those sporadic terms of the sequence $2^{3j+2}/3^{2j+1}$ contained between $\sup_{\gamma<\beta} a_\gamma$ and a_β.

Conjecture 10. The numbers of the sequence $c_{\omega\beta+n}$ that converges to $b_\beta = b/3^a$ (with $\|b\| = 3a+1$) are numbers from the sequences

$$\frac{p(q3^j+1)}{3^{a+j}}, \quad \text{where} \quad b = pq, \text{ and, } \|p(q3^j+1)\| = 3a+3j+2,$$

and those sporadic terms of the sequence $2^{3j+1}/3^{2j}$ contained between $\sup_{\gamma<\beta} b_\gamma$ and b_β.

Conjecture 11. The numbers of the sequence $a_{\omega\beta+n}$ that converges to $c_\beta/3 = b/3^a$ (with $\|b\| = 3a-1$) are numbers from the sequences

$$\frac{p(q3^j+1)}{3^{a+j}}, \quad \text{where} \quad b = pq, \text{ and, } \|p(q3^j+1)\| = 3a+3j,$$

and those sporadic terms of the sequence $2^{3j}/3^{2j}$ contained between $\sup_{\gamma<\beta} \frac{1}{3}c_\gamma$ and $\frac{1}{3}c_\beta$.

In Conjecture 9, 10 and 11 we observe that some terms come from subsequent sequences. For example, the term $c_\omega = 320/243$ is the term corresponding to $j=0$ of the sequence $2^6(4\cdot 3^j+1)/3^{j+5}$, that converges to $b_3 = 256/243$.

The above conjectures allow one to predict, for example, the 200 largest numbers with complexity 30.

The numbers with complexity 14 divided by 81, are

$$\begin{aligned}
c_0 &= 162/81, & c_1 &= 144/81, & c_2 &= 135/81, & c_3 &= 128/81, & c_4 &= 126/81,\\
c_5 &= 120/81, & c_6 &= 117/81, & c_7 &= 114/81, & c_9 &= 112/81, & c_{10} &= 111/81,\\
c_{11} &= 110/81, & c_{13} &= 109/81, & c_{\omega+1} &= 105/81, & c_{\omega+2} &= 104/81, & c_{\omega+3} &= 102/81,\\
c_{\omega+6} &= 100/81, & c_{\omega+8} &= 99/81, & c_{\omega+10} &= 98/81, & c_{\omega+14} &= 97/81, & c_{\omega 2} &= 95/81,\\
c_{\omega 2+3} &= 93/81, & c_{\omega 2+5} &= 92/81, & c_{\omega 2+8} &= 91/81, & c_{\omega 3+4} &= 88/81, & c_{\omega 3+7} &= 87/81,\\
c_{\omega 3+15} &= 86/81, & c_{\omega 4+2} &= 85/81, & c_{\omega 5+1} &= 83/81, & & & c_{\omega^2+\omega+2} &= 79/81,\\
& & & & c_{\omega^2+\omega 2+3} &= 77/81,
\end{aligned}$$

and

$$\frac{71}{81},\quad \frac{69}{81},\quad \frac{67}{81},\quad \frac{59}{81}.$$

For the last four numbers we do not have enough data to know the corresponding ordinal.

References

[1] N. Alon and J. H. Spencer, *The Probabilistic Method*, John Wiley and Sons, New York, 1992.

[2] H. Altman, Integer complexity and well-ordering, *Michigan Math. J.* **64** (2015), 509-538.

[3] H. Altman, Internal Structure of addition chains: well-ordering, *Theoret. Comput. Sci.* **721** (2018), 54-69.

[4] H. Altman, Integer complexity: representing numbers of bounded defect, *Theoret. Comput. Sci.* **652** (2016), 64-85.

[5] H. Altman, Integer Complexity: algorithms and computational results, *Integers* **18** (2018), #A45.

[6] H. Altman, Integer complexity: the integer defect, *Mosc. J. of Comb. Number Theory* **8**(3) (2019), 193-217.

[7] H. Altman and J. Arias de Reyna, Integer complexity, stability, and self-similarity, `https://arxiv.org/abs/2111.00671v5`

[8] H. Altman and J. Zelinsky, Numbers with integer complexity close to the lower bound, *Integers* **12** (2012), 1093-1125.

[9] J. Arias de Reyna, Complexity of natural numbers, Translation of Complejidad de los números naturales, *Gac. R. Soc. Mat. Esp.* **3** (2000), 230–250.

[10] J. Arias de Reyna, Arithmetical self-similar compact sets, `https://arxiv.org/abs/2302.06224`.

[11] H. Chernoff, A measure of the asymptotic efficiency for test of a hypothesis based on the sum of observations, *Ann. of Math. Statist.*, **23** (1952), 493-509.

[12] M. R. Garey and D. S. Johnson, *Computers and Intractability, a Guide to the Theory of NP-completeness*, W. H. Freeman and Co., 1979

[13] R. K. Guy, What is the least number of ones needed to represent n using only $+$ and $\times$ (and parentheses)?, *Amer. Math. Monthly* **93** (1986), 189-190.

[14] J. Hastad, The Shrinkage exponent of the Morgan formulas is 2, *Siam J. Comput.* **27** (1998), 48-64.

[15] J. Iraids, K Balodis, J Čerņenoks, M Opmanis, R Opmanis and K. Podnieks, Integer complexity: Experimental and analytical results, `https://arxiv.org/abs/1203.6462`.

[16] K. Mahler and J. Popken, On a maximum problem in arithmetic, (Dutch), *Nieuw Arch. Wiskd.* (3) **1** (1953), 1-15.

[17] D. A. Rawsthorne, How many 1's are needed?, *Fibonacci Quart.* **27** (1989), 14-17.

[18] OEIS Foundation Inc. (2023), The On-Line Encyclopedia of Integer Sequences, Published electronically at `https://oeis.org`.

[19] U. Zwick, A $4n$ lower bound on the combinatorial complexity of certain symmetric Boolean functions over the basis of unate dyadic Boolean functions, *Siam J. Comput.* **20** (1991), 499-505.

Addendum to: Complexity of Natural Numbers

A1. Introduction

This addendum describes the current status of eleven conjectures posed in the paper: *Complejidad de los numeros naturales*, Gac. R. Soc. Mat. Esp. **3** (2000), 230–250, whose English translation is given above.

The complexity $\|n\|$ of a natural number n is the least number of 1's needed to write an expression for n in the language using the addition $+$, the product $*$, the unit 1 and parentheses as the only symbols. So $\|9\| = 6$ since $9 = (1+1+1)*(1+1+1)$ and there is no other expression for 9 with less than 6 unit symbols. Richard Guy popularized several problems on this concept, such as asking if $\|2^n\| = 2n$ for all $n \geq 1$.

Frequently we have $\|3n\| = \|n\| + 3$. Therefore, it is natural to think about the fractions $n/3^{\lfloor \|n\|/3 \rfloor}$ that remain invariant when we change n to $3n$ and $\|3n\| = \|n\| + 3$. A number is called stable when $\|n3^k\| = \|n\| + 3k$ for all $k \geq 1$. In [9] we made some conjectures about these fractions.

As a young man, the author lived isolated from other mathematicians and was always eager to learn about open problems. When in 2000, as a professional mathematician, we published a paper [9] in the 'Gaceta' in Spanish full of conjectures. In writing the Gaceta paper the authr was thinking of a young person looking for a topic to research. We wanted to provide a fruitful topic. We was not interested in asking fully accurate questions, but left to this young person the formulation of correct statements. We knew that what we wrote was not exactly true, but the author was sure that we were directing this young audience to a rich topic.

These objectives were amply met when Harry Altman and Joshua Zelinsky submitted a paper [8] to *Integers* in late 2011, resolving some of the conjectures. In particular they showed that for each $m \geq 1$ there is some $a \geq 0$ such that $n = m3^a$ is stable. In 2015 Altman [2] resolved more of them. We ended up later collaborating with Altman on [7] in which we reformulate and resolve the rest of our conjectures. In the remainder of this addendum we summarize the current status of these conjectures. Most of them are true after some reformulation and have been proved.

A2. State of the Conjectures

The main contribution of Altman and Zelinsky [8] is the definition of the *defect* $\delta(n) := \|n\| - 3\log_3 n$, and their proof that for any $x > 0$ the set

$$A_x = \{n \in \mathbb{N} \colon \delta(n) < x\}$$

has a simple alternate description that can be computed efficiently.

A2.1. Conjecture 1

(True). A number n is called *stable* if $\|3^k n\| = \|n\| + 3k$ for all $k \geq 0$. The conjecture says that for any natural number n there is K such that $n3^K$ is stable.

Conjecture 1 is proved in [8, Theorem 13]. The *stable complexity* of n is then defined by $\|n\|_{\text{st}} := \|3^K n\| - 3K$, and the *stable defect* is defined by $\delta_{\text{st}}(n) := \|n\|_{\text{st}} - 3\log_3 n$.

A2.2. Conjecture 2

(True after reformulation). The Conjecture states that $\|p(q3^j + 1)\| = \|p\| + \|q\| + 3j+1$. This is true under the extra hypotheses that pq is stable and $\|pq\| = \|p\|+\|q\|$.

The revised statement is proved in [7, Theorem 1.18]. The two extra conditions were not mentioned in the original statement of Conjecture 2. In [7, Theorem 1.20] another version of Conjecture 2 is proved under the extra hypotheses that pq is stable and $\|pq\| + 1 = \|p\| + \|q\|$.

Without extra hypotheses, in [7] the counterexample $p = 2$, $q = 1094$ is mentioned, where

$$\|2 \cdot 1094\| = \|3^7 + 1\| = 22 < \|2\| + \|1094\| = 24.$$

And we have $\|2(1094 \cdot 3^k + 1)\| = \|2188 \cdot 3^k + 2\| \leq 24 + 3k$, instead of $\|2(1094 \cdot 3^k + 1)\| = 25 + 3k$.

A2.3. Conjecture 3

(Partly true; false as stated). Three sets are defined in Conjectures 5, 6 and 7. These sets are well-ordered, which is proved in [2, Theorem A.5]. However a main assertion of Conjecture 3, that the (largest) numbers of complexity $3n+u$ are exactly the first natural numbers contained in these sequences when its terms are multiplied by 3^n is not true, because the numbers in these sequences are not all stable.

We may reformulate Conjecture 3, by stating that the stable numbers of complexity $3n + u$ are equal to the stable numbers in these sequences when its terms are multiplied by 3^n. But then this revised conjecture is almost tautological.

With good will, we could say that the well-ordering part in Conjecture 3 is true but not the rest.

A2.4. Conjectures 4, 5, 6, 7

All these conjectures are true. Conjecture 4 is true by definition if we take the definition in Conjectures 5, 6 and 7 of the sequences. The three defined sets are

$$S_u := \Big\{\frac{n}{3^k} \colon n \text{ stable and } \|n\| = 3k + u\Big\}, \qquad u \in \{0, 1, 2\}. \tag{1}$$

Conjectures 3 and 4 together state that the sets S_u are each well-ordered by the reverse of the usual order of $\mathbb{R}$.

Conjectures 5, 6, and 7 are proved in [2, Theorem A.5].

We may summarize as follows. Conjecture 3 says that there are certain sequences with two properties: well-ordering and representation of certain numbers. The second part is not true. Conjectures 5, 6 and 7 define some sequences and say that they coincide with those of Conjecture 3. The property of well-ordering predicted in Conjecture 3 is fulfilled by the sequences defined in Conjectures 5, 6 and 7.

There is a change of language between this paper [9] and that used in the subsequent papers [8], [2], and [7]. The conjectures in [9] speak of the sets S_u defined in (1); the later papers refer to the sets

$$\mathscr{D}^u_{\text{st}} := \{\delta_{\text{st}}(n) \colon n \text{ stable number with } \|n\|_{\text{st}} \equiv u \pmod 3\}.$$

The map $x \mapsto u - 3\log_3 x$ maps S_u into $\mathscr{D}^u_{\text{st}}$. To prove this, consider an element $\frac{n}{3^k} \in S_u$ with n stable and $\|n\|_{\text{st}} = 3k + u$; then

$$u - 3\log_3 \frac{n}{3^k} = 3k + u - 3\log_3 n = \delta_{\text{st}}(n) \in \mathscr{D}^u_{\text{st}}.$$

This map is clearly bijective between these two sets. This map reverses the order (it is strictly decreasing). Therefore it is an isomorphism of well-ordered sets.

Therefore, for any ordinal $0 \leq \alpha < \omega^\omega$ and any $u \in \{0,1,2\}$, we have the following relationship, between the α-th element $S_u[\alpha] = \frac{n}{3^\ell}$ of S_u and the α-th element $\mathscr{D}^u_{\text{st}}[\alpha] = \delta_{\text{st}}(n)$ of $\mathscr{D}^u_{\text{st}}$

$$u - 3\log_3 S_u[\alpha] = \mathscr{D}^u_{\text{st}}[\alpha].$$

This relationship will make translation easy in what follows.

A2.5. Conjecture 8

This conjecture is true. The sets S_u are well-ordered by $x \preccurlyeq y$ if and only if $x > y$. For a well-ordered set L and an ordinal α we denote by $L[\alpha]$ the element of L in position α. With this notation, Conjecture 8, asserts that

$$\lim_{n\to\infty} S_0[\omega\alpha + n] = S_2[\alpha]/3, \quad \lim_{n\to\infty} S_1[\omega\alpha + n] = S_0[\alpha], \tag{2}$$

$$\lim_{n\to\infty} S_2[\omega\alpha + n] = S_1[\alpha]. \tag{3}$$

Conjecture 8 is proved in [7, Thmeorem 5.1]. (The change of language in the corollary offers no serious difficulty).

A2.6. Conjectures 9, 10, and 11

These conjectures are true after reformulation. Theorem 5.3 in [7] gives the true version of these conjectures. We may state it in our language as the following.

Theorem 1. *For any ordinal $\beta < \omega^\omega$, let $S_u[\beta] = \frac{n}{3^k}$ with $\|n\|_{\text{st}} = 3k + u$; then the two sets*

$$\{S_{u+1}[\omega\beta + r] \colon r \in \mathbb{Z}_{\geq 0}\},$$
$$\Big\{\frac{b(a3^r + 1)}{3^{r+k+\varepsilon}} \colon n = ab, \quad \|n\|_{\text{st}} = \|a\|_{\text{st}} + \|b\|_{\text{st}}, \quad r \in \mathbb{Z}_{\geq 0}\Big\}$$

have a finite symmetric difference. (Here we take $u + 1 \pmod 3$ and $\varepsilon = 0$ for $u = 0$, 1 and $\varepsilon = 1$ for $u = 2$).

The conditions on $n = ab$, with a and b in the second set, were not correct in the statement of Conjectures 9, 10 and 11. Conjectures 9, 10 and 11 speak of the difference between the two sets as consisting of the fractions corresponding to powers of 2. This is not true; there are powers of two, but for example $S_0(\omega 2) = \frac{1280}{2187}$ is a counterexample not of this form already appearing in the Gaceta paper!

A2.7. Table

We include a table summarizing the situation.

Conjecture	Truth value	Solution found in
1	True	[8, Theorem 13]
2	True (re-stated)	[7, Theorem 1.18 and 1.20]
3(wo), 4, 5, 6, 7	True	[2]
3(r)	False	
8	True	[7, Theorem 5.1]
9, 10, 11 main part	True (re-stated)	[7, Theorem 5.3]
9, 10, 11 sporadics	False	

Table 2: Gaceta conjectures

In [10] we give a new interpretation of the conjectures as asserting the existence of a surprising object encoding these properties, which we call an *arithmetic compact set*.