NUMBERS WITH INTEGER COMPLEXITY CLOSE TO THE LOWER BOUND I

HARRY ALTMAN AND JOSHUA ZELINSKY

Abstract. Define $f(n)$ to be the *integer complexity* of n, the smallest number of ones needed to write n using an arbitrary combination of addition and multiplication. John Selfridge showed that $f(n) \geq 3\log_3 n$ for all n. Define $d(n) = f(n) - 3\log_3 n$; in this paper we present a method for classifying all n with $d(n) < r$ for a given r. From this, we derive a number of consequences. We prove that $f(2^m 3^k) = 2m + 3k$ for $m \leq 31$ with m and k not both zero, and present a method that can, with more computation, prove the same for larger m. We extend a result of Daniel Rawsthorne by finding formulae for the r'th largest number with complexity k, so long as k is sufficiently large relative to r. Defining A_r to be the set of all n with $d(n) < r$, we prove that $A_r(x) = \Theta((\log x)^{\lfloor r \rfloor + 1})$. Finally we prove that the set of all values of d is well-ordered, with order type ω^ω, as was conjectured earlier by Juan Arias de Reyna.

1. Introduction and motivation

In this paper we consider the notion of *integer complexity*, as was introduced by Mahler and Popken in 1953 in [6], and later popularized by Richard Guy in [3]; it appears as problem F26 in his *Unsolved Problems in Number Theory* [4]. We say the complexity of a natural number n is the smallest number ones needed to write it using any combination of addition and multiplication. For instance, 7 has a complexity of 6, since it can be written using 6 ones as $(1+1+1)(1+1)+1$, but not with any fewer. We will refer to a representation of n that uses the smallest number of ones as *most-efficient*, and we will denote the complexity of n by $f(n)$. Note that we can compute $f(n)$ recursively; $f(1) = 1$, and for $n > 1$,

$$f(n) = \min_{\substack{a,b \in \mathbb{N} \\ a+b=n \text{ or } ab=n}} f(a) + f(b).$$

(For those interested in actually computing f, there is an algorithm slightly faster than the naïve one, due to Srinivas and Shankar and described in [8]; we used this in our own calculations.)

The complexity of n is approximately logarithmic in n. For $n > 1$, we have $f(n) \leq 3\log_2 n$ as can be seen by writing n in binary. John Selfridge showed (as noted in [7]) that we also have $f(n) \geq 3\log_3 n$. In more detail, let us denote by $E(k)$ the largest number writable with k ones. What Selfridge showed is that $E(k)$ is given by the formulae $E(1) = 1$; $E(3k) = 3^k$; $E(3k{+}2) = 2{\cdot}3^k$; and $E(3k{+}4) = 4{\cdot}3^k$. (Note, by the way, that for $k > 1$, we have $E(k+3) = 3E(k)$.) In particular, for any k we have $E(k) \leq 3^{k/3}$, and so for any n we have $f(n) \geq 3\log_3 n$.

Obviously a slightly better lower bound can be obtained from Selfridge's result, namely, $f(n) \geq L(n)$ where $L(n)$ is given by

$$L(n) = \begin{cases} 1 & \text{if} \quad n = 1 \\ 3k & \text{if} \quad 3^k \le n < 4 \cdot 3^{k-1} \text{ and } k \neq 0 \\ 3k+2 & \text{if} \quad 2 \cdot 3^k \le n < 3^{k+1} \\ 3k+4 & \text{if} \quad 4 \cdot 3^k \le n < 2 \cdot 3^{k-1} \end{cases}$$

Note that $L(n)$ is equal to the smallest number of ones needed to make a number that is at least n, as well as the smallest k such that $E(k) \geq n$. We also have, for $n > 1$, $L(3n) = L(n)+3$. Also observe that, past the initial 1, the ratios of successive values of E are $3/2, 4/3, 3/2, 3/2, 4/3, 3/2, \ldots$; this sequence of ratios will reoccur later.

Not much more about the growth rate of f is known, though better upper bounds have recently been proven. Indeed, it is not even known that $f(n) \not\approx 3\log_3(n)$, a rather weak assertion. We will suggest a potential approach to this problem in Part II.

The notion of integer complexity is somewhat similar to the more well-known notion of addition chain length. An addition chain for n of length k is an increasing sequence $(a_0, \ldots, a_k)$, starting with $a_0 = 1$, ending with $a_k = n$, and such that, for $i > 0$, each a_i is the sum of two (possibly equal) previous terms. The length of the shortest addition chain for n is denoted $l(n)$ [9]. Both are ways of measuring the complexity of a given number, and both are approximately logarithmic in n. However they get there a bit differently – in an addition chain, once a number is constructed once, it can then be used repeatedly at no additional cost. This is not allowed in integer complexity; $f(5) = 5$, as every 1 in $1+1+1+1+1$ must be paid for. Instead of allowing free reuse, though, we do allow multiplication. It seems that allowing either multiplication or free reuse is enough to make the resulting complexity measure approximately logarithmic. Interestingly, if we allow both of these by considering $\tau(n)$, the length of the shortest addition-multiplication chain for n, while obviously $\tau(n) = \Omega(\log\log n)$, it is not known that $\tau(n) = O(\log\log n)$, nor is it expected to be true. In fact, the growth rate of τ – specifically, $\tau(n!)$ – is related to the P-vs.-NP problem for Blum-Shub-Smale machines over the complex numbers. If there is no sequence a_n such that $\tau(a_n n!)$ grows only polylogarithmically, then $P \neq NP$ for such machines. [2]

Unfortunately, integer complexity seems to be lacking in nice properties. For instance, it is not monotonic; $f(11) = 8$, but $f(12) = 7$ (via the representation $(1+1+1+1)(1+1+1)$). One case where it does behave nicely is powers of 3; for $n > 0$ we have $f(3^n) \leq 3n$, since we can write 3^n as the product of n copies of $1+1+1$; and by Selfridge's bound above, we do indeed have that $f(3^n) = 3n$, and there is no shorter way of writing 3^n. However, if we attempt to generalize this to, say, powers of 5, this fails; while $f(5) = 5$, and $f(25) = 10$, and $f(125) = 15$, it turns out that $f(5^6) = 29$ instead of the expected 30, as $5^6 = 15626 = 1+(1+1)(1+1)(1+1)(1+1+1)(1+1+1)(1+(1+1)(1+1)(1+1)(1+1+1)(1+1+1)(1+1+1))$.

The corresponding question for powers of 2, whether $f(2^m) = 2m$ for $m > 0$, is open. More generally, is $f(2^m 3^k) = 2m + 3k$, if m and k are not both zero? Jānis Iraids has verified this for $2^m 3^k \leq 10^{12}$ [5], so in particular $f(2^m) = 2m$ for $1 \leq m \leq 39$. Note that while handling general powers of 2 is hard, fixed powers of 2 with varying powers of 3 turn out to be considerably more workable. In Section 5 of this paper we will prove:

Theorem 1.1. *For $m \le 31$ and any k with m and k not both zero, $f(2^m 3^k) = 2m + 3k$.*

Moreover, we will present a general method that can prove this for larger m with more computation.

It is not too hard to prove this for smaller m; from Selfridge's bound we see that this holds for $m \le 2$. Indeed, Selfridge's result is in fact strong enough to show:

Proposition 1.2. *For $m \le 10$ and any k with m and k not both zero, $f(2^m 3^k) = 2m + 3k$.*

Proof. It suffices to show it for $m = 10$. Note that $E(19+3k) = 4\cdot 3^{5+k} = 972\cdot 3^k < 1024 \cdot 3^k$; hence $2^{10}3^k$ cannot be made with less than $20 + 3k$ ones. ☐

Define $E_r(k)$ to mean the r-th highest number writable with k ones (where we are zero-indexing, so $E_0(k) = E(k)$). Rawsthorne proved in 1989 in [7] that for $k \ge 6$, we have $E_1(k) = (8/9)E(k)$. From this we can conclude:

Proposition 1.3. *For $m \le 13$ and any k with m and k not both zero, $f(2^m 3^k) = 2m + 3k$.*

Proof. It suffices to show it for $m = 13$. Note that $E_1(25 + 3k) = 32 \cdot 3^{5+k} = 7776 \cdot 3^k < 8192 \cdot 3^k$. Hence $2^{13}3^k$ cannot be written with less than $26 + 3k$ ones, unless it is equal to $E(25 + 3k) = 4 \cdot 3^{7+k}$, which it is not. ☐

One might ask if $f(3n) = f(n)+3$ for $n > 1$, but this is false; 107 is the smallest counterexample, with $f(107) = 16$ but $f(321) = 18$. However, in Section 5 we will show that this is true if $n > E(f(n))/3$. Furthermore in the next section we will prove the following, earlier conjectured by Juan Arias de Reyna:

Theorem 1.4. *For any m, there exists a K such that for any $k \ge K$, we have $f(3^k m) = 3(k - K) + f(3^K m)$.*

Now one approach to proving $f(2^m 3^k) = 2m + 3k$ for larger m might be to directly extend Rawsthorne's result – if one had similar formulae for $E_r(k)$ (at least for k sufficiently large relative to r), one could apply this same technique to push m up further. In fact such formulae exist – in Section 6 we will prove:

Theorem 1.5. *Given $r \ge 0$, and k_0 a congruence class modulo 3, there exists K and h such that for $k \ge K$ with $k \equiv k_0 \pmod 3$, we have $E_r(k) = hE(k)$, where K and h are given by tables that can be found in Section 6.*

(It is also possible to prove these formulae by other methods than the ones in this paper, but we do not have the space here to discuss such techniques.) But as we will later see, regardless of the size of r, these cannot get us past $m = 18$. We would like to consider $E_\omega(k)$, but this is meaningless. A change of viewpoint is needed.

Juan Arias de Reyna in [1] found the correct way of considering these formulae to extend them beyond the finite case. Rather than write $E_r(k) = hE(k)$, instead write $E_r(k)/E(k) = h$. Replace $E_r(k)$ with an arbitrary n, and $E(k)$ with $E(f(n))$. So instead of talking about the highest numbers that can be made with k ones, we can talk about the highest values of $R(n) := n/E(f(n))$ that can occur. It is not obvious that this replacement is justified, i.e. that $f(E_r(k)) = k$ when k is large enough for the appropriate formulae to hold; we will prove this when we derive the formulae for $E_r(k)$. We will then classify the highest values of these that can occur, and use this to prove $f(2^m 3^k) = 2m + 3k$ for higher values of $m \leq 31$ and there is no obvious obstacle to using this for any given m.

For this to really "extend the formulae for E_r into the transfinite range", we would want the set of all values of $R(n)$ to be reverse well-ordered. This turns out to be true:

Theorem 1.6. *The set of all values of $R(n)$ is reverse well-ordered, with order type ω^ω.*

Arias conjectured this in [1] and we will prove this in Sections 8 and 9.

2. Introduction part 2 – defect and integral defect

While E is easy to compute, it is still not the easiest to work with, so instead of $n/E(f(n))$ we will prefer to consider $n/3^{f(n)/3}$. Nothing is lost if we switch to this formulation; $3^{k/3} = cE(k)$ for a constant c depending only on the residue class of k modulo 3 (and whether or not $k = 1$), so if we know what k is modulo 3 it is then easy to convert back and forth between the two notions. Indeed, $n/3^{f(n)/3}$ actually carries more information, as due to the irrationality of $3^{1/3}$ and $3^{2/3}$ we can from this number determine the residue of $f(n)$ modulo 3 and thus determine $R(n)$, whereas the reverse is not possible. (The case $n = 1$ puts a slight snag in this, but it easy to see that if $n/3^{f(n)/3} = 3^{-1/3}$ then we must have $n = 1$, as otherwise we would have $f(n) = 4 \cdot 3^k$ but $n = 4 \cdot 3^{k-1}$, an impossibility. More on this in a moment.)

We will make one more change of viewpoint – instead of working with $n/3^{f(n)/3}$, we will work with $f(n) - 3\log_3(n)$, which we will denote $d(n)$ and call the *defect* of n. This encodes the same information, as they're related by $n/3^{f(n)/3} = 3^{-d(n)/3}$; in particular from it we can deduce both $R(n)$ the residue of $f(n)$ modulo 3. The problem now becomes to classify numbers of small defect – numbers with complexity close to the lower bound.

We will prove in Section 7 is that there are not very many numbers with complexity close to the lower bound:

Theorem 2.1. *Let A_r denote the set of all n with $d(n) < r$, and let $A_r(x)$ denote the number of elements of A_r which are less than x. Then for any $r \geq 0$, we have $A_r(x) = \Theta((\log x)^{\lfloor r \rfloor + 1})$.*

To be concrete about the equivalence between d and R,

$$d(n) = \begin{cases} -3\log_3(R(n)) & \text{if } \quad f(n) \equiv 0 \pmod 3 \\ -3\log_3(R(n)) + d(2) & \text{if } \quad f(n) \equiv 2 \pmod 3 \\ -3\log_3(R(n)) + 2d(2) & \text{if } \quad f(n) \equiv 1 \pmod 3 \text{ and } n \neq 1 \\ 0 & \text{if } \quad n = 1 \end{cases}$$

In fact $d(n)$ encodes enough information that we have:

Lemma 2.2. *If $d(n) = d(m)$ for two numbers n and m, then either $n = m3^k$ for some k or vice versa. Indeed if $d(n) - d(m)$ is any rational number, in particular any integer, the same conclusion holds.*

Proof. If $d(n) - d(m)$ is rational, then $\log_3(n/m)$ is rational; since n/m is rational, the only way this can occur is if $\log_3(n/m)$ is an integer. □

Of course, there is a better lower bound than $3\log_3(n)$, namely $L(n)$, so define $D(n) = f(n) - L(n)$, the *integral defect* of n. We will not typically work with this directly, however, as it is much less convenient and is a lossy notion; for instance, $D(n) = 0$ for all n from 1 to 10. To say $D(n) = 0$ is to say $f(n) = L(n)$, that is, $f(n)$ is the smallest k such that $E(k) \geq n$, or in other words, that $E(f(n))$ is the smallest value of E which is at least n. So more generally to say $D(n) = l$, or $L(n) = f(n) - l$, then, is to say that $E(f(n))$ is the l-th value of E past the smallest one which is at least n. Thus what D actually tells is how many values $E(k)$ there are such that $n \leq E(k) \leq E(f(n))$, it being one less than this number. Since (barring the initial 2/1) the ratios of successive values of E repeat with a period of 3, it follows that $D(n)$ is actually telling us the broad range in which $R(n)$ falls, though the changeover points will depend on the residue of $f(n)$ modulo 3. (Hence if we know $d(n)$, since we can determine $R(n)$ and the residue of $f(n)$ modulo 3, we can determine $D(n)$.)

Equivalently, to say that $D(n) \leq l$, is the same as to say that, n is greater than all numbers m with $f(m) < f(n) - l$. Thus, $D(n) = 0$ is the same as to say that n is greater than all m with $f(m) < f(n)$. Note that D is numerically close to d, since $L(n)$ is close to $3\log_3 n$; for any n, by using the formulae for E and L we can see that $3\log_3 n \leq L(n) < 3\log_3 n + 1 + 2d(2)$, and therefore $D(n) \leq d(n) < D(n) + 1 + 2d(2)$.

Note that if $n = ab$ with $f(n) = f(a) + f(b) - k$, then $d(n) = d(a) + d(b) - k$, and in particular $d(3n) = d(n)$ if and only if $f(3n) = 3 + f(n)$. Similarly note for $n > 1$, if $f(3n) = f(n) + 3 - k$, then $D(3n) = D(n) - k$.

The notion of defect gives us an easy proof of the following conjecture of Arias:

Theorem 2.3. *For any m, there exists a K such that for any $k \geq K$, we have $f(3^k m) = 3(k - K) + f(3^K m)$.*

Proof. For any n, note that $d(3n) \leq d(n)$, with equality if and only if $f(3n) = f(n) + 3$. So the sequence $d(m), d(3m), d(9m), \ldots$ is non-increasing, nonnegative, and can only decrease in integral amounts, hence it must eventually stabilize, which proves the theorem. (The proof could be done with D instead with little modification.) □

Note that this still leaves open

Question 2.4. *In Theorem 2.3, is it possible to get a bound on K based on m?*

3. Main Lemma

Here, we state and prove what will be our primary tool for the rest of the paper. Given any $r > 0$, we can, by applying this lemma repeatedly, put restrictions on what n can satisfy $d(n) < r$.

For any real $r \geq 0$, define A_r to be $\{n \in \mathbb{N} : d(n) < r\}$; define D_r to be $\{d(n) : n \in \mathbb{N}, d(n) < r\}$; and define B_r to consist of those elements of A_r that cannot be written most efficiently as $3m$ for any m (so that A_r consists precisely of elements of B_r times powers of 3). Define $\overline{A}_r$, $\overline{D}_r$, and $\overline{B}_r$ to be like A_r, D_r, and B_r but with nonstrict inequalities.

We will use the main lemma to inductively build up a superset of A_r. Note that as a consequence of Rawsthorne's result that $E_1(k) = (8/9)E(k)$ for $k \geq 6$, we have that for any n, if $R(n) > 8/9$ then $R(n) = 1$. Considering what this yields for $d(n)$ (depending on n modulo 3), we see that $d(2) = 2 - 3\log_3 2 = 0.107\ldots$ is the smallest non-zero defect, i.e. $A_{d(2)}$ consists only of powers of 3. This is the base case to which our lemma is the inductive step.

(Though we will later derive the previously mentioned formulae for $E_r(k)$ as a consequence of this base case and main lemma, it is actually possible to prove them by an entirely different method which does not use Rawsthorne's result as a base case. Again, however, we do not have the space to discuss such here.)

Finally define T_α to consist of those natural numbers $n < \frac{1}{3^{\frac{1-\alpha}{3}}-1} + 1$ whose only shortest representation is as either 1 or as $(n-1)+1$. Note that for $\alpha < 1$, we have that T_α is a finite set. Then:

Proposition 3.1. *For any $\alpha < 1$ and $k \geq 2$, any element of $B_{(k+1)\alpha}$ can be written most efficiently in one of the following forms:*

(1) A product of an element of $B_{i\alpha}$ and an element of $B_{j\alpha}$ with $i+j = k+2$ and $2 \leq i, j \leq k$, with defects totalling less than $(k+1)\alpha$;

(2) An element a of $A_{k\alpha}$ plus a number $b \leq a$ such that

$$d(a) + f(b) < (k+1)\alpha + 3\log_3 2,$$

possibly times an element of B_α;

(3) Or an element of T_α, possibly times an element of B_α;

And any element of $B_{2\alpha}$ can be written in one of the following forms:

(1) A product of up to three elements of B_α, with defects totalling less than 2α;

(2) An element a of A_α plus a number $b \leq a$ such that

$$d(a) + f(b) < 2\alpha + 3\log_3 2,$$

possibly times an element of B_α;

(3) Or an element of T_α, possibly times an element of B_α.

Proof. Suppose $m \in B_{(k+1)\alpha}$; take a most efficient representation of m, which is either ab, $a+b$, or 1. If $m = 1$, we are done.

Suppose m can be most efficiently written as a product, say as $\prod_{i=1}^{r} m_i$ (where $r \geq 2$ and each m_i cannot be written most efficiently as a product); so $\sum_{i=1}^{r} d(m_i) = d(m) < (k+1)\alpha$. Note that no product of a subset of the m_i can be written most efficiently as 3 times another number, as else we could choose one of the m_i to be 3, contrary to the assumption $m \in B_{(k+1)\alpha}$. If $k \geq 2$, then either there exists an i with $d(m_i) \geq k\alpha$, or else we can partition the $d(m_i)$ into two nonempty sets each with sum less than $k\alpha$; in this case, call their products a and b. Then $d(a) + d(b) < (k+1)\alpha$, so if we let $(i-1)\alpha$ be the largest multiple of α at most $d(a)$, then $d(b) < j\alpha$ if we let $j = k+2-i$. If such a partition is not possible, then letting a be m_i with $d(m_i) \geq k\alpha$, and b the product of the others, then m is the product of an element of B_α and an element of $B_{(k+1)\alpha}$ which cannot be written as a product; see next paragraph for how to handle this. In the case $k = 1$, either there exists an i with $d(m_i) \geq k\alpha$, or else we can partition the $d(m_i)$ into either two or three nonempty sets each with sum less than α, so either we have the "see next paragraph case", or we have a product of three elements of B_α.

So now we have to determine what elements of $B_{(k+1)\alpha}$ cannot be written in this way, both to complete the multiplication case and to handle the addition case. Suppose $m = a + b$ with $f(m) = f(a) + f(b)$ and b minimal, so $a \geq b$. Then

$$f(a) + f(b) = f(m) < 3\log_3(a+b) + (k+1)\alpha \leq 3\log_3(2a) + (k+1)\alpha,$$

so $d(a) + f(b) < (k+1)\alpha + 3\log_3 2$. Furthermore, b cannot be written most efficiently as a sum $c + d$ or else we could regroup $a + (c+d)$ as $(a+c)+d$ which would contradict the minimality of b.

If $a \in A_{k\alpha}$, we are done. Otherwise, we have

$$3\log_3 a + k\alpha + f(b) < f(a) + f(b) = f(m) <$$
$$3\log_3(a+b) + (k+1)\alpha \le 3\log_3(2a) + (k+1)\alpha,$$

so $f(b) < 3\log_3 2 + \alpha$; since $\alpha < 1$, we have $f(b) \le 2$ and thus $b \le 2$; by the assumption that b cannot be written most efficiently as a sum, we have $b = 1$. Hence $3\log_3 a + k\alpha + 1 < 3\log_3(a+1) + (k+1)\alpha$; if we solve for a, we find that $m = a + 1 \in T_\alpha$; and we may assume that this is the only most efficient way to write m as otherwise it would be covered in one of the other cases. □

Note that while we specified that the representation is most efficient, and included constraints based on the defects of the numbers in the already-known A_r, we don't really need either of these for most purposes, when we are just interested in computing some superset of $A_{(k+1)\alpha}$. We don't even have any easy way of checking whether the resulting reprentations are most-efficient, after all. In particular, in the addition case, note that the requirement that $d(a)+f(b) < (k+1)\alpha+3\log_3 2$ implies the weaker requirement that just $f(b) < (k+1)\alpha + 3\log_3 2$, or just $L(b) < (k+1)\alpha + 3\log_3 2$. We'll define the set of b satisfying $f(b) < x + 3\log_3 2$ to be S_x; note that S_x is always finite.

For $\overline{A}_k$ and $\overline{B}_k$ we have a similar theorem – just use $\overline{B}_r$'s instead of B_r's in the pure product case and use a nonstrict inequality – and the proof is the same except for the strictness of the inequalities. Whether we state theorems in terms of A_r or $\overline{A}_r$ will depend on convenience, but typically the distinction will not matter since $\overline{A}_r \setminus A_r$ can consist of at most a single number times powers of three.

4. Computation and bootstrapping

We will explicitly computing the set A_1. We do this for two reasons. Firstly, we will later compute A_r for larger r, so A_1 is on the way. Secondly, later results in Section 7 and onward will depend on one the structure of A_1. This section accomplishes a sort of bootstrapping – we compute A_1 using $A_{d(2)}$ as our base case, but once we have A_1 it will become our new base case.

Since we want to compute A_1, we will here discuss how we can use the main lemma to compute A_r for small r. The main lemma allows us to inductively construct a superset of A_r, but if we want to determine A_r itself – and if we don't want this computation to blow up very quickly – we'll need some way of then determining the actual complexities of the resulting candidates.

Upper bounds are easy. To find lower bounds on complexities, we will typically use the following technique: Say we want to show that $f(n) \ge k$; since $f(n)$ is always an integer, it suffices to show $f(n) > k - 1$, which we do by noting that if $f(n) \le k - 1$ held, it would put n in some A_l which we have already determined and know it's not in. In particular, we have:

Lemma 4.1. *Take $\alpha \le 1/2$. Say $d(a) < i\alpha$ and $d(b) < j\alpha$, and let $k + 2 = i + j$. Then $f(ab) = f(a) + f(b)$ unless $d(ab) < k\alpha$.*

Proof. Note

$$f(ab) \ge 3\log_3(ab) + k\alpha = 3\log_3 a + 3\log_3 b + (i + j - 2)\alpha > f(a) + f(b) - 1$$

so $f(ab) \ge f(a) + f(b)$, done. □

Lemma 4.2. *Take $\alpha \le d(2)$. Say $d(a) < k\alpha$, then $f(3^m(a+1)) = 3m + f(a) + 1$ unless $d(3^m(a+1)) < k\alpha$.*

Proof. Note

$$f(3^m(a+1)) \ge 3\log_3(a+1) + 3m + k\alpha > f(a) + 3m$$

so $f(3^m(a+1)) \ge 3m + f(a) + 1$, done. □

With these two lemmas in hand, and the base case knowledge of A_α for $\alpha \le d(2)$, we can pick a step size $\alpha \le d(2)$ and inductively compute $A_{k\alpha}$. If we know $A_{k\alpha}$, and the complexities of the numbers therein, first we use the main lemma to generate candidates for $A_{(k+1)\alpha}$; then we check which of these are already in $A_{k\alpha}$. For those numbers that are already in, we are done. For those that are not, we can use the above two lemmas to determine their complexities and hence whether they are actually in. This works so long as $k\alpha < 5 - 3\log_3 2 = 3 + d(2)$ (or $k \le 28$), as then $S_{k\alpha} = \{1\}$.

So for A_r for $r < 3 + d(2)$, we have a method to determine A_r, which is almost an algorithm – unfortunately to turn it into an actual algorithm, we would need some way to represent the data that would allow the required operations, and we have not figured this out. Hence all our computations have been done by hand. This method works for $r \ge 3 + d(2)$ as well, by specially tweaking Lemma 4.2 and the additive case of the main lemma to handle additive constants other than 1, but we will not discuss this here.

It isn't obvious that expressions of the form $a+b$ are ever relevant when $a, b \ne 1$; Rawsthorne's computations in [7] failed to uncover any most-efficient representations $a+b$ with $a, b \ne 1$, for instance. However, such numbers do exist, as Arias and Van de Lune found [4]; they even give a prime p which is most efficiently represented as $6 + (p-6)$ but not $1 + (p-1)$, namely $353942783 = 6 + 353942777$.

In any case, we can then compute that the numbers with defect less than 1 are as follows:

- 3^k of complexity $3k$ for $k \ge 1$
- $2^a 3^k$ of complexity $2a + 3k$ for $a \le 9$
- $5 \cdot 2^a 3^k$ of complexity $5 + 2a + 3k$ for $a \le 3$
- $7 \cdot 2^a 3^k$ of complexity $6 + 2a + 3k$ for $a \le 2$
- $19 \cdot 3^k$ of complexity $9 + 3k$
- $13 \cdot 3^k$ of complexity $8 + 3k$
- $(3^n + 1)3^k$ of complexity $1 + 3n + 3k$ (unless $n = 0$)

(Note also 1 is the only number of defect exactly 1.) The real significance of this list, however, that makes everything in Section 7 and onward work, is:

Corollary 4.3. *For every $\alpha < 1$, we have that B_α is a finite set.*

Note this fails for B_1; the defects of $(3^n + 1)3^k$ approach 1 as n approaches infinity.

5. Further results of computation

Using the above method we have in fact classified all numbers with defect less than $22d(2) = 2.3586\ldots$. This is greater than $2 + 2d(2)$ so this in particular allows us to determine all numbers with integral defect at most one.

Due to its length, we have attached a table of the results separately.

Since determining A_r allows us to put lower bounds on the complexities of any numbers not in it. we have the following:

Lemma 5.1. *Suppose that none of $2^{n+9}3^k$ lie in $A_{nd(2)}$ for any k. Then for any $m \le n+9$ and any k (with m and k not both zero), $f(2^m 3^k) = 2m + 3k$.*

Proof. It suffices to show that $f(2^{n+9}3^k) > 2n + 3k + 17$, but by assumption,

$$f(2^{n+9}3^k) > (n+9)3\log_3 2 + 3k + nd(2) = 2n + 3k + 27\log_3 2 > 2n + 3k + 17,$$

and we are done. □

From our classification, it is straightforward to check that $2^{31}3^k$ does not lie in $A_{22d(2)}$ for any k, so we can conclude

Corollary 5.2. *For $m \le 31$ and any k with m and k not both zero, $f(2^m 3^k) = 2m + 3k$.*

as we claimed above. We can also classify the set of n with $D(n) = 0$ as follows:

Corollary 5.3. *The n with $D(n) = 0$ are precisely those numbers that can be written in one of the following forms:*

- $2^m 3^k$ *with* $m \le 10$
- $2^a(2^b 3^l + 1)3^k$ *with* $a + b \le 2$.

We will show in Section 6 that from this statement we can deduce formulae for the $E_r(k)$ (for k sufficiently large relative to r), and vice versa; it will turn out that the n satisfying $D(n) = 0$ are almost exactly the numbers that can be written as $E_r(k)$ for such k and r. Meanwhile we leave it to the reader to see how Corollary 5.3 implies that $f(2^m 3^k) = 2m + 3k$ for $m \le 18$ with m and k not both zero. Obviously a similar classification could be made for $D(n) \le 1$, but it is ugly and unenlightening so we have not listed it here.

Our computations also tell us a little about the question of when $f(3n)$ is equal to $f(n) + 3$. We see that aside from 1, the number of smallest defect for which $f(3n) \ne f(n) + 3$ is 683. We also see that

Corollary 5.4. *If $D(3n) = 0$, then $D(n) = 0$, and if $D(3n) = 1$, then $D(n) = 1$. Hence if $D(n) \le 2$ and $n \ne 1$, $f(3n) = f(n) + 3$.*

If we rewrite the condition $D(n) \le 2$ as $D(n) < 3$, and observe that $D(n) < 3$ if and only if $R(n) > 1/3$, then can rewrite this conclusion as

Corollary 5.5. *For $n > 1$, if $n > E(f(n))/3$, then $f(3n) = f(n) + 3$.*

Knowing A_2 makes the "α to 2α" case of the main lemma unnecessary, but we haven't computed the full ordering on the set of defects less than 2 so it's easier to use the main lemma.

6. Formulae for $E_r(n)$

As mentioned in in the previous section, Corollary 5.3 is equivalent to a series of formulae for $E_r(k)$ that work so long as k is sufficiently large depending on r, and the $E_r(k)$ obtained this way are almost exactly the n satisfying $D(n) = 0$. In this section we prove this. (Though we do not include here our original proof of Corollary 5.3, it is worth noting that our original proof went the other way – it proved Corollary 5.3 by way of these formulae. Hence why we formulate this as an equivalence rather than just stating that the formulae are a corollary.)

Fix k modulo 3. The numbers n with $D(n) = 0$ and $f(n) \equiv k \pmod 3$ are the numbers of lowest defect (or highest R) with $f(n) \equiv k \pmod 3$. If we look at the numbers n with integral defect of 0 and split them up by their complexity modulo 3, we get that n is one of:

- For $f(n) \equiv 0 \pmod 3$:
  - $(2 \cdot 3^m + 1)3^k$ with $R(n) = \frac{2 \cdot 3^m+1}{3^{m+1}}$
  - $2(3^m + 1)3^k$ (for $m \geq 1$) with $R(n) = \frac{2(3^m+1)}{3^{m+1}}$
  - $64 \cdot 3^k$ with $R(n) = 64/81$
  - $512 \cdot 3^k$ with $R(n) = 512/729$
- For $f(n) \equiv 2 \pmod 3$:
  - $(4 \cdot 3^m + 1)3^k$ with $R(n) = \frac{(4 \cdot 3^m+1)}{2 \cdot 3^{m+1}}$
  - $2(2 \cdot 3^m + 1)3^k$ with $R(n) = \frac{2(2 \cdot 3^m+1)}{2 \cdot 3^{m+1}}$
  - $4(3^m + 1)3^k$ (for $m \geq 1$) with $R(n) = \frac{4(3^m+1)}{2 \cdot 3^{m+1}}$
  - $2 \cdot 3^k$ with $R(n) = 1$
  - $128 \cdot 3^k$ with $R(n) = 64/81$
  - $1024 \cdot 3^k$ with $R(n) = 512/729$
- For $f(n) \equiv 1 \pmod 3$, $n \neq 1$:
  - $(3^m + 1)3^k$ (for $m \geq 1$) with $R(n) = \frac{(3^m+1)}{4 \cdot 3^{m-1}}$
  - $32 \cdot 3^k$ with $R(n) = 8/9$
  - $256 \cdot 3^k$ of $R(n) = 64/81$

Note that from the above tables, for each congruence class of k modulo 3, the numbers n with $D(n) = 0$ are precisely those with the ω largest values of R, or ω smallest defects, among m with $f(m) \equiv k \pmod 3$ – this is what will give us the connection between n with $D(n) = 0$, and the formulae for $E_r(k)$, the largest numbers writable with k ones. (Indeed, we expect that more generally, for each congruence class of k modulo 3, the numbers n with $D(n) \leq m$ are precisely those with the ω^m largest values of R; we will motivate this conjecture in Part II.)

Suppose we want to determine the r-th largest number with complexity k. From these tables, we can determine the r-th largest value of R that occurs for numbers with complexity congruent to k modulo 3. Call this value h; then $hE(k)$ is indeed the r-th largest number with complexity k if and only if all $r + 1$ values of R that are at least h and occur for numbers with complexity congruent to k modulo 3, do indeed occur for numbers with complexity k. We also know that if $D(3n) = 0$, then $D(n) = 0$, or in other words, that if $D(n) = 0$ and $k \equiv f(n) \pmod 3$, then n can be written with k ones if and only if $R(n)E(k)$ is an integer. So this take k to be large enough for all of these to be integers; hence we get the following table of resulting formulae for the r-th largest number with complexity k. When these formulae are valid, we have that they are also equal to $E_r(k)$, the r-th largest number writable with k ones, because each of them is greater than $E(k-1)$; hence any n greater than such a number which was writable with k ones would also have to have complexity exactly k (thus finally justifying our claim back in Section 1 that the expression $n/E(f(n))$ is a generalization of $E_r(k)/E(k)$).

As a corollary, we see that $D(n) = 0$ precisely when it can be written as $E_r(k)/3^l$ for some r, k as above and some l. So we have shown:

Theorem 6.1. *Given $r \geq 0$, and k_0 a congruence class modulo 3, there exists K and h such that for $k \geq K$ with $k \equiv k_0 \pmod 3$, we have $E_r(k) = hE(k)$, where K and h are given by the following tables:*

- *For $k \equiv 0 \pmod 3$:*

	r	h	K
	0	1	3
	1	8/9	6
	2	64/81	12
	3	7/9	12
	4	20/27	12
	5	19/27	12
	6	512/729	18
	7	56/81	18
	8	55/81	18
	9	164/243	18
	10	163/243	18
(for $n \geq 6$)	$2n-1$	$2/3 + 2/3^n$	$3n$
(for $n \geq 6$)	$2n$	$2/3 + 1/3^n$	$3n$

- *For $k \equiv 2 \pmod 3$:*

	r	h	K
	0	1	2
	1	8/9	8
	2	5/6	8
	3	64/81	14
	4	7/9	14
	5	20/27	14
	6	13/18	14
	7	19/27	14
	8	512/729	20
	9	56/81	20
	10	37/54	20
	11	55/81	20
	12	164/243	20
	13	109/162	20
	14	163/243	20
(for $n \geq 6$)	$3n-3$	$2/3 + 2/3^n$	$3n+2$
(for $n \geq 6$)	$3n-2$	$2/3 + 1/(2 \cdot 3^{n-1})$	$3n+2$
(for $n \geq 6$)	$3n-1$	$2/3 + 1/3^n$	$3n+2$

- *For $k \equiv 1 \pmod 3$:*

	r	h	K
	0	1	1
	1	8/9	10
	2	5/6	10
	3	64/81	16
	4	7/9	16
	5	41/54	16
(for $n \geq 4$)	$n+2$	$3/4 + 1/(4 \cdot 3^n)$	$3n+4$

Now we claim that given these formulae, it is actually possible to deduce Corollary 5.3; indeed, this is the route our original proof of Corollary 5.3 took.

Fix k modulo 3, and let $\lambda = \lim_{n\to\infty} h_r$, which is $2/3$ if k is 0 or 2 modulo 3 and $3/4$ if k is 1 modulo 3; i.e., $\lambda = E(k-1)/E(k)$ (save for when $k=1$ or $k=2$, which are easily handled). Now suppose n is of one of the forms listed in Corollary 5.3; checking that $D(n)=0$ is then straightforward. So we just need the converse; suppose $D(n)=0$ and $f(n)=k$. Then $n > E(k-1)$ and so $R(n) > \lambda$. Hence there exists some r with $R(n) \ge h_r > \lambda$. Thus if we let $N = n3^l$ be such that $k+3l \ge K_r$, we must have $R(N) = R(n) \in \{h_0, \ldots, h_{r-1}\}$. (We have $R(N) = R(n)$ because $D(n)=0$.) And given this fact it then follows from the above tables that n is of one of the forms listed in Corollary 5.3.

Of note is that the original formulae for $E_0(k)$ and $E_1(k)$ that served as our base cases were both originally proven directly by induction on k, which raises the question of whether the same can be done for general $E_r(k)$ now that the formulae for them are known.

7. Abstract results – Ternary families

In our discussion of concrete computations above, we used a small step size $\alpha \le d(2)$, and kept our superset of A_r small by using a filtering step. In what follows, we will use a different trick to keep our supersets of A_r from getting too large; we will use step sizes arbitrarily close to 1, and ignore any filtering step.

Define the set of *ternary families* to be the smallest set of functions $\mathbb{Z}_{\ge 0}^k \to \mathbb{N}$ (k varying) satisfying:

- Every "singleton" function $\mathbb{Z}_{\ge 0}^0 \to \mathbb{N}$ is a ternary family.
- Given ternary families $F : \mathbb{Z}_{\ge 0}^k \to \mathbb{N}$ and $G : \mathbb{Z}_{\ge 0}^l \to \mathbb{N}$, the function $(x,y) \mapsto F(x)G(y) : \mathbb{Z}_{\ge 0}^{k+l} \to \mathbb{N}$ is a ternary family. We'll denote this map $F \otimes G$.
- Given a ternary family F and a positive integer c, the function $(x,n) \mapsto F(x)3^n + c$ is a ternary family. We'll denote this map F_c.

We'll refer to the number of arguments of a ternary family F as its *rank* and denote it rk F. Also for a ternary family F we'll define the corresponding *expanded ternary family* to be the function $\tilde{F}(x,n) = F(x)3^n$; we'll use the rank of $\tilde{F}$ synonymously with the rank of F. We'll also refer to the image of an (expanded) ternary family as an (expanded) ternary family when there is no ambiguity.

Proposition 7.1. *Given any $0 < \alpha < 1$, and any $k \ge 1$, we have that $B_{k\alpha}$ is contained in a union of finitely many ternary families of rank at most $k-1$. (As a corollary, we get the same for $A_{k\alpha}$ but with expanded families.)*

Proof. Induct on k using the main lemma, Corollary 4.3, and the finiteness of S_r for any r and T_α. □

And hence:

Corollary 7.2. *For any r, B_r is contained in a union of finitely many ternary families of rank at most $\lfloor r \rfloor$. (Expanded ones for A_r.)*

Proof. Note that $r = (\lfloor r \rfloor + 1)(r/(\lfloor r \rfloor + 1))$ and $r/(\lfloor r \rfloor + 1) < 1$, which proves the claim. □

Now we will prove some bounds on the size of ternary families.

Proposition 7.3. *Let S be the image of a ternary family of rank k. Then $S(x) = O((\log x)^k)$. ($O((\log x)^{k+1}$ for the corresponding expanded family.)*

This is an easy induction from the definition of ternary family. With this we can show:

Theorem 7.4. *For any r, let $k = \lfloor r \rfloor$; then $B_r(x) = \Theta((\log x)^k)$, $A_r(x) = \Theta((\log x)^{k+1})$.*

Proof. The upper bound follows immediately from the above lemmas. For the lower bound, note that B_k contains the ternary family

$$F(n_1, \ldots, n_k) = (\ldots((3 \cdot 3^{n_1} + 1)3^{n_2} + 1)\ldots)3^{n_k} + 1.$$

(None of these are multiples of three, and each have complexity at most $3(1 + n_1 + \ldots + n_k) + k$ and so have defect less than k.) If we let S be the image of F, and $\tilde{S}$ the expanded version, it is easy to see that $S(x) = \Theta((\log x)^k)$ and $\tilde{S}(x) = \Theta((\log x)^{k+1})$, which proves the claim. □

(Note that barring $r = 0$, for which B_r is empty but $\overline{B}_r = \{3\}$, the same holds for $\overline{B}_r$ and $\overline{A}_r$, as $\overline{B}_r \setminus B_r$ is finite by Theorem 2.3.)

So while it's a long way from proving $f(n) \not\sim 3 \log_3 n$, at least we can prove

Corollary 7.5. *There exist numbers of arbitrarily large defect.*

In fact it is true that:

Theorem 7.6. *Let S be the image of a ternary family $F(n_1, \ldots, n_k)$ of rank k. Then $S(x) = \Theta((\log x)^k)$; more specifically,*

$$\frac{1}{k!^2}(\log_3 x)^k \lesssim S(x) \lesssim \frac{1}{k!}(\log_3 x)^k.$$

And if $\tilde{S}$ is the image of $\tilde{F}$,

$$\frac{1}{k!(k+1)!}(\log_3 x)^{k+1} \lesssim \tilde{S}(x) \lesssim \frac{1}{(k+1)!}(\log_3 x)^{k+1}.$$

But we do not have the space to prove this here and will do so in Part II. We also conjecture:

Conjecture 7.7. *For $k \geq 0$ an integer,*

$$\overline{B}_k(x) \sim \frac{(k+1)^{k-1}}{k!^2}(\log_3 x)^k,$$

$$\overline{A}_k(x) \sim \frac{(k+1)^{k-2}}{k!^2}(\log_3 x)^{k+1}.$$

In Part II we will motivate this conjecture, extend it to non-integers, and show that more uniform version of it would imply that $f(n) \not\sim 3 \log_3 n$.

8. Well-ordering of defects

In this section we now prove that, as Juan Arias de Reyna previously conjectured, the set of defects is well-ordered, with order type ω^ω. (His original conjecture actually took a slightly different form; more on that in the next section.)

Define the *leading coefficient* (denoted *LC*) of a ternary family $F(n_1, \ldots, n_k)$ to be the limit $\lim_{n_1,\ldots,n_k \to \infty} F(n_1, \ldots, n_k)/3^{\sum n_i}$. Note also that this can also be determined recursively; leading coefficient of a constant is itself, $LC(F \otimes G) = LC(F)LC(G)$, and $LC(F_c) = LC(G)$. (Hence in particular $LC(F)$ is always finite and nonzero.) From this recursion it follows that $F(n_1, \ldots, n_k)$ is always at least $3^{\sum n_i} LC(F)$. Define the *base complexity* (denoted BC) of a ternary family by a similar recursion; base complexity of a singleton n is $f(n)$, $BC(F \otimes G) = BC(F) + BC(G)$, and $BC(F_c) = BC(F) + f(c)$. As written this is actually not well-defined, but we can make it so by simply taking the smallest possible value if there's any ambiguity. Note that for any ternary family F and any $n_1, \ldots, n_k$, we have $f(F(n_1, \ldots, n_k)) \le BC(F) + 3\sum n_i$. Finally define the *obvious defect upper bound*, UB, of a ternary family by $UB(F) = BC(F) - 3\log_3 LC(F)$. Note that this is, in fact, an upper bound on the defect of any number in the image of F. Also note $BC(F) \ge \text{rk } F + f(LC(F))$ and hence $UB(F) \ge \text{rk } F + d(LC(F))$; this can be proven by induction, using a decomposition of F such that BC always adds.

Proposition 8.1. *Let S be the image of an (expanded) ternary family of rank k. Then the set of defects of S is well-ordered, with order type less than ω^{k+1}.*

Proof. Take a ternary family F of rank k. For any $n_1, \ldots, n_k$, consider the difference between the actual complexity $f(F(n_1, \ldots, n_k))$ and the upper bound of $BC(F) + 3\sum n_i$. This difference can take on only finitely many values, as it cannot be more than $UB(F)$, since otherwise the defect would be less than 0; the same applies to the expanded version. Hence we can split up $\mathbb{Z}_{\ge 0}^k$ into a union of finitely many sets, on each of which $d(F(n_1, \ldots, n_k))$ is given by $BC(F) + 3\sum n_i - 3\log_3(F(n_1, \ldots, n_k))$ minus some constant. It suffices to show that each of these sets of defects is well-ordered with order type at most ω^k, as the natural sum of finitely many ω^k's is certainly less than ω^{k+1}. And for this it suffices to show that the image of $d_F : (n_1, \ldots, n_k) \mapsto 3\sum n_i - 3\log_3(F(n_1, \ldots, n_k))$ is well-ordered, as adding constants doesn't change the order. (Note, by the way, that $\lim_{x\to\infty} d_F(x) = -3\log_3 LC(F)$ by definition of LC.)

In fact, it suffices to show that d_F is monotonic; $\mathbb{Z}_{\ge 0}^k$ is a well-partial order (in the strong sense of that term), so any totally-ordered image of it under a monotonic function is well-ordered. Futhermore the resulting well-order must have order type at most ω^k, as if we actually pull it back to a total order on $\mathbb{Z}_{\ge 0}^k$ (breaking ties between points with equal d_F by looking at lexicographic order), it extends the original partial order, and the natural product of k ω's is in fact ω^k.

To prove d_F monotonic, we induct on F. For singleton functions it is trivial, and as $d_{F\otimes G} = d_F + d_G$, if it is true for two families it is true for their product. Finally, say $G(x, n) = F_c(x, n) = F(x)3^n + c$; if d_F is monotonic, clearly d_G is monotonic in x, so we need only check monotonicity in n. This holds because the inequality $3n - 3\log_3(F(x)3^n + c) < 3(n+1) - 3\log_3(F(x)3^{n+1} + c)$ is equivalent to the inequality $c < 3c$. Hence d_F is monotonic – in fact, strictly so – and we are done. □

With this we can show:

Theorem 8.2. *For any r, let $k = \lfloor r \rfloor$; then D_r is well-ordered, with order type at least ω^k and less than ω^{k+1}.*

Proof. Well-ordering and the upper bound follows immediately from the above lemmas. For the lower bound, once again note that B_k contains the ternary family

$$F(n_1, \ldots, n_k) = (\ldots((3 \cdot 3^{n_1} + 1)3^{n_2} + 1)\ldots)3^{n_k} + 1.$$

Now if we let S the set of defects of F, we need to check that S has order type at least ω^k. But it suffices to show that the image of d_F has order type ω^k, since if ω^k is partitioned into finitely many parts, at least one must have order type ω^k. And this is easily checked, which proves the claim. □

Corollary 8.3. *The set of all defects is well-ordered with order type ω^ω.*

In fact it is also true that:

Proposition 8.4. *Let S be the image of any ternary family of rank k. Then the set of defects of S has order type at least ω^k. (Hence also for expanded families.)*

But once again we will save the proof of this for Part II.

When r is an integer, we can actually pin things down a bit more:

Theorem 8.5. *Let k be an integer; then the order type of $\overline{D}_k$ is exactly ω^k, unless $k = 1$, in which case it's $\omega + 1$.*

Proof. If it were any larger than $\omega^k + 1$, it would contain a copy of ω^k, strictly bounded above by a non-maximum element of $\overline{D}_k$, hence an element of $\overline{D}_k$ less than k. So there would be a copy of ω^k bounded away from k, contradicting Lemma 8.2. Similarly, if it has order type equal to $\omega^k + 1$, the maximum element must be k itself, or else we again get a copy of ω^k bounded away from k. So the order type is ω^k unless k is itself a defect, which happens only when $k = 1$. □

9. Alternate forms of well-ordering and some conjectures of Arias

Corollary 8.3 was previously conjectured by Juan Arias de Reyna in [1] in a slightly different form. What we have proved is that the set $\{n/3^{f(n)/3} : n \in \mathbb{N}\}$ is reverse well-ordered, with reverse order type ω^ω. It may be more natural to discuss $R(n)$ rather than $n/3^{f(n)/3}$, but it is easy to translate between these if we know $f(n)$ modulo 3. Rather than consider the set of all defects, we will separately consider the sets $\{n/3^{f(n)/3} : n \equiv 0 \pmod 3\}$, $\{n/3^{f(n)/3} : n \equiv 2 \pmod 3\}$, $\{n/3^{f(n)/3} : n \equiv 1 \pmod 3, n \neq 1\}$. (We exclude 1 for simplicity because it does not follow the pattern of other n congruent to 1 modulo 3.)

These three sets are all reverse well-ordered, with reverse order type at most ω^ω. To see that each has reverse order type exactly ω^ω, consider the ternary families

$$F_k(n_1, \ldots, n_k) = (\ldots((3 \cdot 3^{n_1} + 1)3^{n_2} + 1)\ldots)3^{n_k} + 1$$
$$G_k(n_1, \ldots, n_k) = (\ldots((2 \cdot 3^{n_1} + 1)3^{n_2} + 1)\ldots)3^{n_k} + 1$$
$$H_k(n_1, \ldots, n_k) = (\ldots((4 \cdot 3^{n_1} + 1)3^{n_2} + 1)\ldots)3^{n_k} + 1.$$

By the same reasoning as in the proof of Theorem 8.2, each of these can be seen to have defects with order type ω^k. Furthermore, there must be an ω^k's worth meeting the obvious upper bound, because for these families the set of defects not meeting the obvious upper bound is bounded above by $k+2d(2)-1 < k$ and hence has order type less than ω^k. Thus the family F_k contributes ω^k's worth to $\{d(n) : f(n) \equiv k \pmod 3\}$, the family G_k to $\{d(n) : f(n) \equiv k + 2 \pmod 3\}$, and the family H_k to $\{d(n) : f(n) \equiv k + 1 \pmod 3\}$.

We can now multiply each of the three sets $\{n/3^{f(n)/3} : n \equiv 0 \pmod 3\}$, $\{n/3^{f(n)/3} : n \equiv 2 \pmod 3\}$, and $\{n/3^{f(n)/3} : n \equiv 1 \pmod 3, n \neq 1\}$by the appropriate constants to see that each of the three sets $\{R(n) : n \equiv 0 \pmod 3\}$, $\{R(n) : n \equiv 2 \pmod 3\}$, $\{R(n) : n \equiv 1 \pmod 3, n \neq 1\}$ are reverse well-ordered with reverse order type ω^ω. Or by using different constants, we could put $3^{\lfloor f(n)/3 \rfloor}$ in the denominator instead of $E(f(n))$, which is the form Arias originally conjectured it in. (Although his actual original conjecture was slightly stronger and would have implied that $f(3n) = f(n) + 3$ for all $n > 1$, which is false; removing that aspect leaves what we have proved here.)

Indeed, knowing this we can even recombine the three (together with 1) to say that $\{R(n) : n \in \mathbb{N}\}$ (or $\{n/3^{\lfloor f(n)/3 \rfloor} : n \in \mathbb{N}\}$) is reverse well-ordered with reverse order type ω^ω. Since it's the union of finitely many reverse well-ordered sets, it too is reverse well-ordered, with reverse order type at least ω^ω. To see that it is exactly ω^ω, observe that if it were any larger, then some proper final segment of it would have reverse order type ω^ω; but this would decompose into a union of proper final segments of the four sets making it up (all of them are coinitial in it as they all get arbitrarily close to 0), implying that ω^ω was at most the natural sum of finitely many ordinals less than it, which is false.

Now let $A = \{R(n) : n \equiv 0 \pmod 3\}$, let $B = \{R(n) : n \equiv 1 \pmod 3), n \neq 1\}$, let $C = \{R(n) : n \equiv 2 \pmod 3\}$, and let a_α, b_α, c_α denote the α'th element from the top (0-indexed, so $a_0 = b_0 = c_0 = 1$) in A, B, C respectively for $\alpha < \omega^\omega$. Arias also made a conjecture which, reformulated slightly, states that for any $\beta < \omega^\omega$,

$$\lim_{n\to\infty} a_{\omega\beta+n} = (2/3)c_\beta$$

$$\lim_{n\to\infty} b_{\omega\beta+n} = (3/4)a_\beta$$

$$\lim_{n\to\infty} c_{\omega\beta+n} = (2/3)b_\beta$$

Looking at our actual data for this, it is easy to see the reason this holds for small defect; take the first of these statements as an example. For every $c_\beta = R(n)$, we get an infinite family of numbers $n3^k + 1$, as well as a collection of infinite families of numbers $p(q3^k + 1)$ for $pq = n$, $f(p) + f(q) = f(n)$; since these families occur in the same order as the original n, it is the limit of $R(n3^k + 1)$ and the other families – together with finitely many other things that don't affect the limit – that appear on the left hand side, and (for examples so far, anyway) this is always eventually the obvious upper bound for sufficiently large k, and everything matches up, with a factor of $2/3$ or $3/4$ to account for the change of modulo 3 complexity. However, it is not at all clear that this pattern will hold up once we leave the realm of products and +1's, which we know eventually happens. So the conjecture is somewhat explained, but far from proved.

10. Acknowledgements

The authors would like to acknowledge their debt to Jānis Iraids and Karlis Podnieks for supplying a wealth of numerical data, to Jeffrey Lagarias for looking over an early draft of this paper and elucidating just what it was we were doing, and to Mike Zieve for providing other assistance with early drafts of the paper.

References

[1] J. Arias de Reyna, Complejidad de los números naturales, *Gaceta R. Soc. Mat. Esp.*, **3**(2000) 230–250.

[2] Lenore Blum, Felipe Cucker, Mike Shub, & Steve Smale, Algebraic Settings for the problem "$P \neq NP$?", *Lectures in Applied Mathematics* vol 32, ed J. Renegar, M. Shub and S. Smale, Amer. Math. Soc. 1996, pp. 125–144

[3] Richard K. Guy, Some suspiciously simple sequences, *Amer. Math. Monthly*, **93**(1986) 186–190; and see **94**(1987) 965 & **96**(1989) 905.

[4] Richard K. Guy, *Unsolved Problems in Number Theory*, Third Edition, Springer-Verlag, New York, 2004, pp. 399–400.

[5] Jānis Iraids, Complexity, URL (version: 2011-05-08): `http://wiki.oranzais.lumii.lv/janiswiki/index.php?title=Special:Complexity`

[6] K. Mahler & J. Popken, On a maximum problem in arithmetic (Dutch), *Nieuw Arch. Wiskunde*, (3) **1**(1953) 1–15; *MR* **14**, 852e.

[7] Daniel A. Rawsthorne, How many 1's are needed? *Fibonacci Quart.*, **27**(1989) 14–17; *MR* **90b**:11008.

[8] Srinivas Vivek V. & Shankar B. R., Integer Complexity: Breaking the $\Theta(n^2)$ barrier, *World Academy of Science*, **41**(2008) 690–691

[9] M. V. Subbarao, Addition Chains – Some Results and Problems, *Number Theory and Applications*, Editor R. A. Mollin, Kluwer Academic Publisher Group, Dordrecht, 1989, pp. 555–574.