Rev. Mat. Iberoamericana **22** (2006), no. 2, 369–412

Arithmetic properties of positive integers with fixed digit sum

Florian Luca

Abstract

In this paper, we look at various arithmetic properties of the set of those positive integers n whose sum of digits in a fixed base $b > 1$ is a fixed positive integers s. For example, we prove that such integers can have many prime factors, that they are not very smooth, and that most such integers have a large prime factor dividing the value of their Euler ϕ function.

1. Introduction

Let $b > 1$ and $s > 1$ be fixed integers. Let $A_{b,s}$ be the set of all positive integers n which are not multiples of b and whose sum of digits in base b is precisely s. That is, $A_{b,s}$ consists of all the positive integers which can be written under the form

$$n = a_0 + a_1 b + \cdots + a_t b^t$$

with $a_0 a_t \neq 0$, $a_i \in \{0,\ 1, \ldots,\ b-1\}$ for $i = 1,\ \ldots,\ t$ and $\sum_{i=0}^{t} a_i = s$.

In this paper, we study the arithmetic properties of the positive integers n belonging to $A_{b,s}$. Notice that the arithmetic properties of these integers reflect the arithmetic properties of *all* the positive integers n (multiples of b or not) whose sum of digits in base b is precisely s, because every such positive integer n can be written in a unique way as $n = b^u m$ with $u \geq 0$ and $m \in A_{b,s}$.

A result of Senge and Strauss (see [21]) says that if $b,\ s,\ b_1$, and s_1 are positive integers (larger than 1) with b and b_1 multiplicatively independent, then $A_{b,s} \cap A_{b_1,s_1}$ is finite. An effective version of this result was given by

2000 Mathematics Subject Classification: Primary: 11A63; Secondary: 11N64.
Keywords: Sum of digits, smooth numbers, subspace theorem, linear forms in logarithms.

Stewart (see [23]). That is, Stewart used Baker's theory of lower bounds for linear forms in logarithms to show that there exists an effectively computable constant $c_1 = c_1(b, s, b_1, s_1)$ depending on b, s, b_1, and s_1, such that if $n \in A_{b,s} \cap A_{b_1,s_1}$, then $n < c_1$. Throughout this paper, we shall use $c_1, c_2, \ldots$ for computable positive constants which are either absolute or depend on our initial data (usually b and s). We also use the Vinogradov symbols $\gg$ and $\ll$ as well as the Landau symbols O and o with their usual meanings.

There are many papers in the literature dealing with arithmetic properties of the set of positive integers characterized by some digital property (see, for example, [3], [7], [10], [11], [12], [15], [16]). In fact, the paper [16] whose title is very similar to the title of our present paper deals indeed with the arithmetic properties of the set of positive integers n whose sum in base b is s, but for the purpose of the results of [16] the number s is not fixed (uniformly in n) but is taken to tend to infinity is a rather special way with respect to the number n (either a lot slower than $\log n$ or is taken to be precisely $s = \lfloor \frac{(b-1)\log n}{2\log b} \rfloor$ which is, roughly speaking, the expected value of s for an arbitrary positive integer n).

The type of questions which are addressed in the above papers is the following. Let $\mathcal{P}$ be some digital property related to the base b (such as the fact that the sum of digits is a fixed number s, or the fact that all the digits are allowed to belong to some fixed proper subset of all the possible digits $\{0, 1, \ldots, b-1\}$ satisfying some mild technical assumptions which are meant to avoid trivialities, etc.). Then in the above papers, the authors investigate the average value of the functions $\omega(n)$ and $\Omega(n)$ over the set of positive integers n characterized by the property $\mathcal{P}$. They also show that there exist such n which have many prime factors, or few, or that such sets contain smooth numbers, etc. A key tool that has been used in every single one of the above papers is an appropriate extension of a Theorem of Gelfond asserting the fact that such numbers are uniformly distributed in residue classes modulo a fixed number m. That is, let x be a large real number and let $W_{\mathcal{P}}(x)$ denote the set of all positive integers $n < x$ having the digital property $\mathcal{P}$. Then there exists a function $f(x)$ defined on the set of positive real numbers x and which tends to infinity with x (and which is made explicit in the above papers for every one of the digital properties $\mathcal{P}$) such that for every positive integer m which is, say coprime to $b(b-1)$ and every non negative integer $a \leq m$ there are *roughly* about $\frac{|W_{\mathcal{P}}(x)|}{m}$ numbers $n < x$ having the property $\mathcal{P}$ and $n \equiv a \pmod m$ and this property holds uniformly in m and a when $m < f(x)$.

We start with an elementary result which, in particular, shows that such an argument cannot be applied in order to investigate the arithmetic properties of the sets $A_{b,s}$.

Proposition 1.1. *Let $k > 1$ be any fixed integer. There exists an infinite set of positive integers S having the following properties:*

(1) *If $m \in S$, then m is coprime to b.*

(2) *If $m_1 \neq m_2$ are distinct members of S, then m_1 and m_2 are coprime.*

(3) *If $m \in S$ and n is any non zero multiple of m, then n has at least k non zero digits when written in base b.*

The set S appearing in the statement of the above Proposition 1.1 will be chosen to be a subset of all numbers of the form $(b^n-1)/(b-1)$ for $n \geq 1$.

Taking $k = s+1$ in the above Proposition 1.1, it follows that there exists an infinite set of positive integers m which are coprime to b and for which the equation $x \equiv 0 \pmod{m}$ has no positive integer solution $x \in A_{b,s}$ and moreover this infinite set of positive integers can be chosen in such a way that any two members of it are coprime (this is in order to avoid trivialities; in fact, if the equation $x \equiv 0 \pmod{m}$ has no positive integer solution $x \in A_{b,s}$, then obviously the equation $x \equiv 0 \pmod{m_1}$ will have no positive integer solution $x \in A_{b,s}$ either with m_1 being any non zero multiple of m). The proof of the above Proposition 1.1 is elementary and is based on an idea employed previously in [13].

We now address the question of smooth numbers. For a positive integer n we write $P(n)$ for the largest prime factor of n with the convention that $P(1) = 1$ and we recall that a number n is called smooth is $P(n) < n^{\varepsilon}$.

Proposition 1.2. *For every $\varepsilon > 0$ there exist infinitely many positive integers $n \in A_{b,s}$ with $P(n) < n^{\varepsilon}$.*

Once one knows that a certain infinite set of positive integers contains infinitely many smooth numbers one can ask two types of questions. One of them is, say how many smooth numbers are there in our infinite set? Here, we can show that for a fixed ε there exists a subset of $n \in A_{b,s}$ of *positive lower logarithmic density* such that the inequality $P(n) < n^{\varepsilon}$ holds for n in this subset. Unlike other authors, by *positive lower logarithmic density* here we mean that there exists a number $\delta > 0$, which can be computed in terms of s alone and another computable constant c_2 depending on b, s and ε, such that for every large positive real number x the set of numbers $n < x$ in $A_{b,s}$ for which $P(n) < n^{\varepsilon}$ holds is of cardinality at least $c_2|A_{b,s}(x)|^{\delta}$. Here and in what follows, we use $A_{b,s}(x)$ to denote the set of all $n < x$ which belong to $A_{b,s}$. If one wants *very smooth numbers* in $A_{b,s}$, then we can show that there exists a computable constant c_3 depending only on b and s such that the inequality $P(n) < n^{\varepsilon(n)}$ holds for infinitely many positive integers $n \in A_{b,s}$ with $\varepsilon(n) = \frac{c_3}{\log_3 n}$. Here and in what follows, for any positive

integer k and any positive real number x we define recursively the function $\log_k x$ as being $\max\{\log(\log_{k-1} x),\ 1\}$, where $\log_1(x) = \max\{\log x, 1\}$ and log denotes the natural logarithm function. Clearly, if x is large with respect to k, then $\log_k x$ is nothing else but the composition of the natural logarithm function log with itself k-times evaluated in x.

The proof of Proposition 1.2 above is elementary in nature and is based on the properties of the cyclotomic polynomials. This idea has been employed previously in [2] and [9] in the context of constructing long strings of smooth consecutive integers.

While Proposition 1.2 and the remarks following it assert that one can find infinitely many smooth numbers in $A_{b,s}$ the next proposition points out that there is a limit to how smooth one can make the numbers from $A_{b,s}$.

Proposition 1.3. *There exists a positive computable constant c_4 depending only on b and s such that the inequality*

$$P(n) > \frac{c_4 \log_2 n \log_3 n}{\log_4 n} \tag{1.1}$$

holds for all $n \in A_{b,s}$.

The proof of the above Proposition 1.3 uses Baker's method of lower bounds for linear forms in logarithms.

We now turn our attention to the functions $\omega(n)$ and $\Omega(n)$ when $n \in A_{b,s}$. The next two propositions show that there exist integers $n \in A_{b,s}$ having many distinct prime factors but that there are not too many positive integers $n \in A_{b,s}$ having a very large $\Omega(n)$.

Proposition 1.4. *There exists a computable constant c_5 depending only on s and b such that there exist infinitely many positive integers $n \in A_{b,s}$ having*

$$\omega(n) > \exp\left(\frac{c_5 \log_2 n}{\log_3 n}\right). \tag{1.2}$$

We point out that the inequality

$$\omega\left(\prod_{n \in A_{b,s}(x)} n\right) \gg \frac{\log^{2-\alpha} x}{\log_2 x}$$

has been proved to hold for large values of the positive real number x in the recent paper [22] provided that $s > 4$, where α is any constant in the interval $(0, 2)$ such that the inequality $s > 4\alpha^{-1} - 3$ holds.

Proposition 1.5. *The estimate*

$$\Omega(n) = o(\log n) \tag{1.3}$$

holds when n tends to infinity through positive integers in the set $A_{b,s}$.

Proposition 1.5 tells us something very interesting about the numbers $n \in A_{b,s}$. Indeed, without any digital restrictions we know that the inequality $\Omega(n) \gg \log n$ holds infinitely often and in [16] it is shown that a similar type of inequality holds infinitely often when n is allowed to run only over all the integers which are not multiples of b and whose sum of digits lies in a fixed congruence class r modulo a fixed positive integer m. So, we see that Proposition 1.5 tells us that such an inequality cannot happen infinitely often when $n \in A_{b,s}$. The proof of this result uses the Subspace Theorem of W. M. Schmidt. While Proposition 1.4 above tells us that $A_{b,s}$ contains numbers with many prime factors, what is in doubt is the maximal order of $\omega(n)$ when $n \in A_{b,s}$. That is, without any digital restrictions we know that the inequality

$$\omega(n) \gg \frac{\log n}{\log_2 n}$$

holds for infinitely many n but our inequality (1.2) is only a logarithmic version of the above inequality. So, we would like to propose the following problem:

Problem 1. *Prove or disprove that*

$$\lim_{\substack{n\to\infty \\ n\in A_{b,s}}} \frac{\omega(n)\log_2 n}{\log n} = 0. \tag{1.4}$$

We conjecture that formula (1.4) does hold but have no idea how to attack this problem.

Another question that is usually of interest for *thin* subsets of positive integers is to ask how many of them are powers.

Proposition 1.6. *Let x be a large positive real number. Then the number of positive integers $n < x$ with $n \in A_{b,s}$ which are perfect powers is $o(|A_{b,s}(x)|)$.*

Notice that there is no reason why $A_{b,s}$ should not contain infinitely many perfect powers if $s \geq 4$. Indeed, $b^{2t} + 2b^t + 1$ is a perfect square for all t and the sum of its digits is precisely 4 for all $t > 0$ and $b > 2$. It would be interesting to encounter a sharp (close to the truth) estimate for the dependence $o(|A_{b,s}(x)|)$ appearing in Proposition 1.6. In fact, it is very likely that the intersection between $A_{b,s}$ and the set of all perfect powers

consists of a union between finitely many positive integers n and finitely many parametric families of integers of the form y^k where $y = a_0 + a_1 b^{i_1} + \cdots + a_t b^{i_t}$ with $k,\ t,\ a_0,\ a_1,\ \ldots,\ a_t$ fixed and $i_1,\ \ldots,\ i_t$ variables in such a way that the number $y^k = (a_0 + a_1 b^{i_1} + \cdots + a_t b^{i_t})^k$ has, in a "canonical way", the sum of its digits s in base b. For a fixed exponent k proving this might be even doable using the technique employed by Corvaja and Zannier in [6] and particular manifestations of this are already known (see, for example, the recent paper [24] of Szalay, where it is shown, among other things, that the equation $2^m + 2^n + 1 = x^2$ has only finitely many positive integer solutions (m, n, x) aside from the infinite family $n = t$ and $m = 2t - 2$ for which $2^m + 2^n + 1 = 2^{2(t-1)} + 2^t + 1 = (2^{t-1} + 1)^2$), but it seems hard to get an absolute bound on k (depending only on b and s) although the existence of such a bound k is implied by a Generalized ABC Conjecture. Note also that for fixed k, b and s there are some simple congruence conditions which, if not fulfilled, guarantee that there are no perfect kth powers in $A_{b,s}$. For example, if $s \equiv 2 \pmod 3$, then an immediate congruence modulo 3 shows that $A_{10,s}$ contains no perfect squares. On the other hand, if say $k > 1$ is fixed and $s = s_0^k$ holds with some positive integer s_0, then it is not hard to show that there exists a positive constant δ (depending on s and k) such that for large values of the positive real number x the set $A_{b,s}(x)$ contains at least $|A_{b,s}(x)|^\delta$ positive integers n which are kth powers.

Finally, we look at the numbers $\phi(n)$ when $n \in A_{b,s}$ where $\phi(n)$ is the Euler function of n and we show that most of these are not very smooth.

Proposition 1.7. *There exists a positive computable constant c_6 depending on b and s such that the inequality*

$$P(\phi(n)) > c_6 \log_2^{1/6} n \log_3^{1/3} n \tag{1.5}$$

holds for almost all positive integers $n \in A_{s,b}$.

By "almost all" in the statement of the above Proposition 1.7 we mean that for a large positive real number x the number of numbers $n < x$ in $A_{b,s}$ for which inequality (1.5) fails is $o(|A_{b,s}(x)|)$. Notice that it is probably impossible to replace "almost all" by "all but finitely many" in the above statement. Indeed, there is no reason why there shouldn't be infinitely many primes of the form $n = 3 \cdot 2^t + 1 = 2^{t+1} + 2^t + 1$ and if there are infinitely such, then $P(\phi(n)) = 3$ will hold for all such integers n and in particular an inequality like (1.5) will fail for infinitely many positive integers $n \in A_{2,3}$. The proof of the above Proposition 1.7 is rather technical and uses, among other things, Baker's theory of lower bounds for linear forms in logarithms of algebraic numbers as well as Schlickewei's quantitative version

of W. M. Schmidt's Subspace Theorem (see [20]). This idea has been employed before in the context of giving a lower bound for $P(\phi(|u_n|))$ which is valid for almost all positive integers n, where $(u_n)_{n\geq 0}$ is a binary recurrent sequence of integers satisfying certain technical conditions.

Acknowledgments. I thank the referee for valuable suggestions and Eugenio Balanzario for an enlightening discussion concerning the behavior of the tail of the series (8.30).

2. The Proof of Proposition 1.1

We fix k and b and let ℓ be a large positive integer. Take the number

$$n_\ell = b^{\ell-1} + b^{\ell-2} + \cdots + b + 1 = \frac{b^\ell - 1}{b-1}.$$

Observe that n_ℓ is coprime to b. Now assume that

$$m = a_0 + a_1 b^{i_1} + \cdots + a_t b^{i_t}, \quad a_0 a_t \neq 0,\ t \leq k-1 \text{ and } a_i \in \{0,\ 1,\ \ldots,\ b-1\}$$

is a non zero multiple of n_ℓ which is not a multiple of b and which has at most k non zero digits in base b. For every exponent i_j with $j = 1, \ldots, t$ we let $\alpha_j \in \{0,\ 1,\ \ldots,\ \ell-1\}$ be such that $i_j \equiv \alpha_j \pmod{\ell}$. Since $b^\ell \equiv 1 \pmod{n_\ell}$ and $n_\ell \mid m$, we get that $n_\ell \mid m'$ where

$$m' = a_0 + a_1 b^{\alpha_1} + \cdots + a_t b^{\alpha_t}.$$

Thus, $m' = cn_\ell$ and since $0 < m' < (b-1)(t+1)b^{\ell-1} \leq (b-1)kb^{\ell-1}$ and $n_\ell > b^{\ell-1}$, we get that $c < c_7$ where $c_7 = (b-1)k$ is independent on ℓ. We may now assume that c is any fixed positive integer below c_7 and rewrite the equation $m' = cn_\ell$ as

$$\begin{aligned}(b-1)m' &= a_0(b-1) + a_1(b-1)b^{\alpha_1} + \cdots + a_t(b-1)b^{\alpha_t} \\ &= c(b-1)n_\ell = c(b^\ell - 1).\end{aligned} \tag{2.1}$$

By looking at the number appearing in the left hand side of equation (2.1), we see immediately that the number of its non zero digits in base b is at most $2(t+1) \leq 2k$. However, by looking at the number appearing in the right hand side of (2.1) and writing it as

$$c(b^\ell - 1) = cb^\ell - c = c(b-1)b^{\ell-1} + (b^{\ell-1} - c),$$

we recognize that this number has at least $\ell - c_8$ non zero digits in base b where one can choose $c_8 = \left\lfloor \frac{c_7}{\log b} \right\rfloor + 1$. Thus, if $\ell > c_9$ where $c_9 = 2k + c_8$, then equation (2.1) is impossible.

The set S claimed by Proposition 1.1 can be chosen to be simply $S = \{n_p \mid p > c_9 \text{ and prime}\}$ and it is clear from what we have said that all the numbers $n \in S$ satisfy 1 and 3 of Proposition 1.1. The fact that they also satisfy 2 follows from the well known fact that the relation

$$\gcd(b^u - 1, b^v - 1) = b^{\gcd(u,v)} - 1$$

holds for all positive integers u and v. In particular, when p and q are distinct primes, we then have

$$\gcd(n_p, n_q) = \gcd\left(\frac{b^p - 1}{b-1}, \frac{b^q - 1}{b-1}\right) = 1.$$

■

3. The Proof of Proposition 1.2

Let z be a large positive parameter to be fixed later and let $p_1 < p_2 < \cdots < p_t \le z$ be all the prime numbers less then or equal to z. Clearly, $t = \pi(z)$. Let λ be any positive integer. Write $\mu = \lambda p_1 \dots p_t$ and consider the number

$$n = 1 + b^\mu + b^{2\mu} + \cdots + b^{(s-1)\mu} = \frac{b^{s\mu} - 1}{b^\mu - 1}.$$

It is clear that $n \in A_{b,s}$. For any positive integer d let

$$\Phi_d(X) = \prod_{\substack{1 \le k \le d \\ \gcd(k,d)=1}} \left(X - e^{\frac{2i\pi k}{d}}\right) \in \mathbb{Z}[X]$$

be the dth cyclotomic polynomial. Since

$$X^m - 1 = \prod_{d \mid m} \Phi_d(X),$$

we get that

$$n = \frac{b^{s\mu} - 1}{b^\mu - 1} = \prod_{\substack{d \mid s\mu \\ d \nmid \mu}} \Phi_d(b). \tag{3.1}$$

From (3.1), we get immediately that for all $d \ge 1$ we have

$$\Phi_d(b) = \prod_{\substack{1 \le k \le d \\ \gcd(k,d)=1}} \left|b - e^{\frac{2i\pi k}{d}}\right| \le (b+1)^{\phi(d)}. \tag{3.2}$$

From (3.1) and (3.2), we certainly get that

$$P(n) \le \max_{\substack{d \mid s\mu \\ d \not\mid \mu}} P(\Phi_d(b)) \le (b+1)^{\phi(s\mu)}. \tag{3.3}$$

The above inequality (3.3) is the key in order to get smooth numbers $n \in A_{b,s}$. To get very smooth such numbers n, set $\lambda = 1$ and take a large z (for example, such that $z > P(s)$). Then obviously

$$\frac{\phi(s\mu)}{s\mu} = \frac{\phi(\mu)}{\mu} < \frac{c_{10}}{\log_2 \mu} \tag{3.4}$$

and the right most inequality (3.4) holds with any constant c_{10} strictly larger than $c_{11} = e^{\gamma}$ provided that z is large (see [17]) where γ is the Euler constant. In particular, we get that the inequality

$$\phi(s\mu) < \frac{c_{10} s\mu}{\log_2 \mu} < \frac{c_{12} s\mu}{\log_2(s\mu)} \tag{3.5}$$

holds with any constant c_{12} strictly larger than c_{10} provided that z is large. Since

$$b^{(s-1)\mu} < 1 + b^{\mu} + \cdots + b^{(s-1)\mu} < n,$$

we get that

$$(s-1)\mu < \frac{\log n}{\log b},$$

therefore

$$s\mu < c_{13} \log n \tag{3.6}$$

where we can take $c_{13} = \frac{s}{(s-1)\log b}$.

Since the function $y \to \frac{y}{\log_2 y}$ is increasing for $y > e^e$, it follows, by (3.5) and (3.6), that the inequality

$$\phi(s\mu) < \frac{c_{10} s\mu}{\log_2 s\mu} < \frac{c_{14} \log n}{\log_3 n} \tag{3.7}$$

holds where one can take c_{14} to be any constant larger than $c_{10}c_{13}$ provided that z is large. Finally, from (3.3) and (3.7), we get

$$P(n) < (b+1)^{\frac{c_{14}\log n}{\log_3 n}} = n^{\varepsilon(n)}$$

where

$$\varepsilon(n) = \frac{c_{15}}{\log_3 n}$$

with $c_{15} = c_{14} \log(b+1)$.

This is what concerns very smooth numbers. To see that for a fixed value of ε we can get a set of positive lower logarithmic density of numbers $n \in A_{b,s}$ we argue as follows. Write $\mu_1 = p_1 \dots p_t$ and choose a value of z large enough such that the inequality

$$\phi(s\mu_1) < (s-1)\mu_1 \varepsilon \frac{\log b}{\log(b+1)}$$

holds for this value of z. The reason that such a value of z exists for a fixed value of ε is a consequence of inequality (3.5). We now get that

$$\phi(s\mu) = \phi(s\lambda\mu_1) \le \lambda\phi(s\mu_1) \le \lambda(s-1)\mu_1\varepsilon\frac{\log b}{\log(b+1)} < \varepsilon\frac{\log n}{\log(b+1)} \tag{3.8}$$

where the last inequality in (3.8) is an immediate consequence of the fact that

$$b^{(s-1)\mu} = b^{\lambda(s-1)\mu_1} < n.$$

With (3.3) and (3.8) we get

$$P(n) < (b+1)^{\phi(s\mu)} < n^{\varepsilon} \tag{3.9}$$

and the above inequality (3.9) holds uniformly in λ.

All is left is to do the count. Let x be a large positive real number. It is not important for our purposes to understand the exact value of $|A_{b,s}(x)|$ but only the precise order of magnitude. We claim that there exist two computable constants c_{16} and c_{17} such that the inequality

$$c_{16}\log^{s-1} x < |A_{b,s}(x)| < c_{17}\log^{s-1} x \tag{3.10}$$

holds. This is almost obvious. Indeed, let us get an upper bound on the number of numbers

$$n = a_0 + a_1 b^{i_1} + \cdots + a_t b^{i_t} < x$$

such that $a_i \in \{1, \dots, b-1\}$ for $i = 0,\ 1,\ \dots,\ t$, $0 < i_1 < \cdots < i_t$ and for which $\sum_{i=0}^{t} a_i = s$. Clearly, $t \le s-1$ and there are only finitely many choices for the number $t \le s-1$ and the $t+1$-uples $(a_0, a_1, \dots, a_t)$ of positive integers less than b for which $\sum_{i=0}^{t} a_i = s$. For each one of these fixed choices, we get that $0 < i_1 < \cdots < i_t$ and $i_t < \frac{\log x}{\log b}$. Hence, there are only $O(\log^t x)$ choices for the t-uple of exponents $(i_1, \dots, i_t)$ and since $t \le s-1$, we get the right half of (3.10). To prove the other half of (3.10), notice that $A_{b,s}$ contains all the numbers

$$n = 1 + b^{i_1} + \cdots + b^{i_{s-1}}$$

where

$$0 < i_1 < i_2 < \cdots < i_{s-1} < \left\lfloor \frac{\log x}{\log b} \right\rfloor - 1 \tag{3.11}$$

and the number of $s-1$-uples of positive integers satisfying (3.11) is

$$\binom{\lfloor \frac{\log x}{\log b} \rfloor - 1}{s-1} \gg \log^{s-1} x$$

which proves the other half of (3.10).

We now get a lower bound on the number of positive integers $n < x$ in $A_{b,s}$ satisfying (3.9). From the above arguments, it follows that if z is a fixed number which is sufficiently large such that inequality (3.8) holds, then any number of the form

$$1 + b^{\lambda\mu_1} + b^{2\lambda\mu_1} + \cdots + b^{(s-1)\lambda\mu_1} \tag{3.12}$$

where λ is an arbitrary positive integer satisfies (3.9). But there are at least

$$\left\lfloor \frac{\log x}{(s-1)\mu_1 \log b} \right\rfloor - 1 \tag{3.13}$$

such numbers λ for which the number shown at (3.12) is $< x$ and it is clear that the number shown at (3.13) is $\gg \log x \gg |A_{b,s}(x)|^{\delta}$ where $\delta = \frac{1}{s-1}$ (see (3.10)). ■

Remarks In the above proof, we have showed that if ε is fixed, then there are a number $\gg |A_{b,s}(x)|^{1/(s-1)}$ numbers $n < x$ in $A_{b,s}$ having $P(n) < n^{\varepsilon}$ where the constant understood in the symbol $\gg$ above depends on ε. When s is not prime, we can do slightly better by a similar argument. Namely, we can show that the number of numbers $n < x$ in $A_{b,s}$ satisfying $P(n) < n^{\varepsilon}$ is at least $\gg |A_{b,s}|^{(\Omega(s)-1)/(s-1)}$ where the constant understood in $\gg$ above depends again on ε. Of course, this is better than the previous argument only when s is not a prime. To get such a better inequality, notice that the number $u = \Omega(s) - 1$ is the maximum positive integer k for which there exists a representation $s = d_1 d_2 \dots d_k$ with integers $d_i > 1$. Pick such a representation $s = d_1 d_2 \dots d_u$ and for $i \in \{1, \dots, u\}$ let

$$n_i = 1 + b^{\lambda_i \mu} + b^{2\lambda_i \mu} + \cdots + b^{(d_i - 1)\lambda_i \mu}$$

where λ_i are some arbitrary positive integers and μ is again of the form $\mu = \prod_{p<z} p$. With fixed ε we can find again a value of z such that the inequality $P(n_i) < n_i^{\varepsilon}$ holds independently on λ_i. Finally, set

$$n = \prod_{i=1}^{u} n_i. \tag{3.14}$$

It is clear that, generically, the sum of digits of n in base b is precisely $\prod_{i=1}^{u} d_i$ and that all its non zero digits are 1. Indeed, the only case when this might not be so is when there exist two disjoint non empty subsets J_1 and J_2 of $\{1, \ldots, u\}$ and two functions $j_1 : J_1 \to \mathbb{N}$ and $j_2 : J_2 \to \mathbb{N}$ such that $j_1(i) \in \{1, \ldots, d_i - 1\}$ and $j_2(i) \in \{1, \ldots, d_i - 1\}$ hold for all $i \in J_1$ or $i \in J_2$, respectively, and such that furthermore

$$\sum_{i \in J_1} j_1(i)\lambda_i = \sum_{i \in J_2} j_2(i)\lambda_i. \tag{3.15}$$

It is now easy to see, by arguments similar to the previous ones, that the number of u-uples $(\lambda_1, \ldots, \lambda_u)$ of positive integers for which the number n given by (3.14) is $< x$ is

$$\gg \log^u x = \log^{\Omega(s)-1} x \gg |A_{b,s}(x)|^{\frac{\Omega(s)-1}{s-1}}$$

and it is also easy to see that most such u-uples of positive integers do not satisfy any one of the finitely many linear relations (3.15) (mainly because anyone of the finitely many linear relations (3.15) will "cut down" on the degree of freedom u of the generic u-uple $(\lambda_1, \ldots, \lambda_u)$), which finishes the argument.

4. The Proof of Proposition 1.3

Let $p_1 < p_2 < \cdots < p_k$ be the first k prime numbers and assume that

$$n = p_1^{\alpha_1} \ldots p_k^{\alpha_k}$$

holds with some $\alpha_i \geq 0$ for $i = 1, \ldots, t$ where

$$n = a_0 + a_1 b^{i_1} + \cdots + a_t b^{i_t}$$

is such that $a_i \in \{1, \ \ldots, \ b-1\}$ for $i = 1, \ldots, t$ and $\sum_{i=0}^{t} a_i = s$. We assume that $i_1 < i_2 < \cdots < i_t$ and we write

$$X = i_t \quad \text{and} \quad Y = \max\{e, \ \alpha_i \mid i = 1, \ \ldots, \ k\}.$$

It is easy to see that the inequality

$$b^{X+1} > n \geq 2^Y$$

holds, therefore $Y < c_{18}X$ holds for large X with $c_{18} = 2 \log b$ (notice that $2 > \log 2$). In what follows, we shall assume that X is large enough. We use the following lower bound for linear forms in p-adic logarithms due to K. Yu (see [25]).

Lemma 4.1. *Let p be a prime number, $r_1, \ldots, r_k$ be rational numbers having both numerators and denominators coprime to p and $n_1, \ldots, n_k$ be positive integers. Write $A_i \geq e$ for the maximum between the height of the rational number r_i (i.e., the maximum between the absolute values of its numerator and denominator when written in reduced form) and e. Write also $N = \max\{e, |n_i| \mid i = 1, \ldots, k\}$. Then there exists an effectively computable constant c_{19} which is absolute such that if the rational number*

$$r_1^{n_1} \ldots r_k^{n_k} - 1$$

is non zero, then

$$\mathrm{ord}_p\left(r_1^{n_1} \ldots r_k^{n_k} - 1\right) < c_{19}^k \log A_1 \ldots \log A_k \log N$$

where for a rational number r we write $\mathrm{ord}_p(r)$ for the exponent at which the prime number p divides the numerator of r.

We use the above Lemma 4.1 to bound recursively $i_1, i_2, \ldots, i_t$ in terms of t and X. Write

$$\begin{aligned} b^{i_1}(a_1 + a_2 b^{i_2 - i_1} + \cdots + a_t b^{i_t - i_1}) &= p_1^{\alpha_1} \ldots p_k^{\alpha_k} - a_0 \\ &= a_0(a_0^{-1} p_1^{\alpha_1} \ldots p_k^{\alpha_k} - 1). \end{aligned} \tag{4.1}$$

Let p be an arbitrary prime number dividing b. The above equation (4.1) implies

$$i_1 \leq \mathrm{ord}_p(b^{i_1}) = \mathrm{ord}_p(a_0) + \mathrm{ord}_p\left(a_0^{-1} p_1^{\alpha_1} \ldots p_k^{\alpha_k} - 1\right). \tag{4.2}$$

Let c_{20} be such that $c_{20} > \mathrm{ord}_p(a_0)$. We use Lemma 4.1 to bound the second term appearing in (4.2). Clearly,

$$\mathrm{ord}_p\left(a_0^{-1} p_1^{\alpha_1} \ldots p_k^{\alpha_k} - 1\right) < c_{19}^k \log A_0 \log A_1 \log p_2 \ldots \log p_k \log Y \tag{4.3}$$

where $A_0 = \max\{|a_0|, e\}$ and $A_1 = e = \max\{p_1, e\}$. Notice that we are entitled to apply Lemma 4.1 because with the index j such that $p_j = p$ we may replace a_0^{-1} by $a_0^{-1} p_j^{\alpha_j}$ and if this rational number does not have both its denominator and numerator coprime to p, then inequality (4.2) is simply $i_1 \leq c_{20}$ (i.e., the second term in the right hand side of (4.2) is zero) while if $a_0^{-1} p^{\alpha_j}$ has both its numerator and denominator coprime to p, then $\alpha_j = \mathrm{ord}_p(a_0)$ and the height of $a_0^{-1} p^{\alpha_j}$ is less than or equal to the height of a_0 (and in this case the factor depending on p_j shouldn't even appear in the right hand side of (4.3)). Set

$$\Omega = \log A_1 \ldots \log p_k < \exp\left(\sum_{i=1}^{k} \log_2 p_k\right) < \exp\left(c_{21} k \log_2 k\right)$$

where c_{21} some absolute constant.

Inequality (4.2) becomes

$$i_1 < c_{20} + c_{19}^k \log A_0 \exp\left(c_{21} k \log_2 k\right) \log Y < \exp\left(c_{22} k \log_2 k\right) \log Y \tag{4.4}$$

where c_{22} can be chosen to be such that

$$c_{22} > 4\max\{\log c_{20}, \log c_{19}, \log_2 A_0, c_{21}\}.$$

We now use induction on the parameter j to show that the inequality

$$i_j + 1 < 3^j \exp\left(c_{22} j k \log_2 k\right) (\log B \log Y)^j \tag{4.5}$$

holds for all $j = 1, \ldots, t$ where $B = \max\{b, e\}$. The step $j = 1$ follows immediately from inequality (4.4). Assume that inequality (4.5) holds for some $j < t$ with $j \geq 1$ and write

$$c_j = a_0 + a_1 b^{i_1} + \cdots + a_j b^{i_j}.$$

Obviously,

$$c_j < b^{i_j+1},$$

therefore

$$\log c_j < (i_j + 1) \log B. \tag{4.6}$$

Notice also that since $j \geq 1$ we have $c_j \geq b + 1 > e$. Write

$$\begin{aligned} b^{i_{j+1}} \left(a_{j+1} + a_{j+2} b^{i_{j+2} - i_{j+1}} + \cdots + a_t b^{i_t - i_{j+1}}\right) &= p_1^{\alpha_1} \ldots p_k^{\alpha_k} - c_j \\ &= c_j \left(c_j^{-1} p_1^{\alpha_1} \ldots p_k^{\alpha_k} - 1\right). \end{aligned}$$

We get

$$i_{j+1} + 1 \leq \mathrm{ord}_p\left(b^{i_{j+1}}\right) + 1 = 1 + \mathrm{ord}_p(c_j) + \mathrm{ord}_p\left(c_j^{-1} p_1^{\alpha_1} \ldots p_k^{\alpha_k} - 1\right).$$

Clearly,

$$\mathrm{ord}_p(c_j) \leq \frac{\log c_j}{\log p}$$

and by Lemma 4.1 we also get

$$\mathrm{ord}_p\left(c_j^{-1} p_1^{\alpha_1} \ldots p_k^{\alpha_k} - 1\right) < c_{19}^k \Omega(\log c_j) \log Y,$$

therefore

$$i_{j+1} + 1 < 1 + (\log c_j) \left(\frac{1}{\log p} + c_{19}^k \Omega \log Y\right). \tag{4.7}$$

So, with (4.6), (4.7) and the induction hypothesis, we get

$$\begin{aligned} i_{j+1}+1 &< 1+(i_j+1)\log B\left(\frac{1}{\log p}+c_{19}^k\Omega\log Y\right)\\ &< 1+3^j(\log B)^{j+1}(\log Y)^j\exp\left(c_{22}jk\log_2 k\right)\left(\frac{1}{\log p}+c_{19}^k\Omega\log Y\right)\\ &< 3^{j+1}\left(\log B\log Y\right)^{j+1}\exp\left(c_{22}(j+1)k\log_2 k\right) \end{aligned}$$

where the last inequality above is obvious. This finishes the induction step. Evaluating (4.5) at $j=t$ we get

$$X < i_t+1 < 3^t\exp\left(c_{22}tk\log_2 k\right)\left(\log B\log Y\right)^t,$$

therefore

$$\log X < c_{23}+c_{24}k\log_2 k+c_{25}\log_2 Y$$

where $c_{23}=t(\log 3+\log_2 B)$, $c_{24}=tc_{22}$ and $c_{25}=t$ (recall that $t\le s-1$). Since $Y<c_{18}X$ we get that the inequality

$$c_{24}k\log_2 k > \log X-c_{23}-c_{25}\log_2(c_{18}X) > \frac{\log X}{2}$$

holds when $X>c_{26}$ where c_{26} is computable and depends on b and s. Thus, for $X>c_{26}$ we get

$$k\log_2 k > c_{27}\log X$$

with $c_{27}=\dfrac{1}{2c_{24}}$, therefore the inequality

$$k > c_{28}\frac{\log X}{\log_3 X}$$

holds with some computable constant c_{28} provided that $X>c_{26}$. Since for large k we also have $p_k>k\log k$ (see [18]), we get that the inequality

$$p_k > k\log k > c_{29}\frac{\log X\log_2 X}{\log_3 X} \tag{4.8}$$

holds for all large enough values of X. All it remains to notice is that $X>c_{30}\log n$ holds for all large enough values of n where c_{30} can be taken to be any constant slightly smaller than $\frac{1}{\log b}$ provided that n is large enough, and now inequality (4.8) tells us that the inequality

$$P(n) > c_{31}\frac{\log_2 n\log_3 n}{\log_4 n}$$

holds for all $n>c_{32}$. We may now replace the constant c_{31} by a smaller constant and conclude that indeed an inequality like the one asserted at (1.1) holds for all $n\in A_{b,s}$. ■

5. The Proof of Proposition 1.4

Here, we use the argument employed in the proof of Proposition 1.2. Let z be a large real parameter and let $s < p_1 < \cdots < p_t \le z$ be all the prime numbers larger than s and smaller than or equal to z. Take $\mu = p_1 \dots p_t$ and look at the number

$$n_z = 1 + b^\mu + b^{2\mu} + \cdots + b^{(s-1)\mu} = \frac{b^{s\mu} - 1}{b^\mu - 1} = \prod_{\substack{d \mid s\mu \\ d \nmid \mu}} \Phi_d(b). \tag{5.1}$$

It is clear that the number of divisors $d \mid s\mu$ such that $d \nmid \mu$ is equal to $\tau(s\mu) - \tau(\mu) = (\tau(s)-1)\tau(\mu) \ge \tau(\mu) = 2^t$. Here, for a positive integer k we use $\tau(k)$ for the number of divisors of k and we also made use of the multiplicativity of the function τ; i.e., the fact that the formula $\tau(s\mu) = \tau(s)\tau(\mu)$ holds because s and μ are coprime. In particular, there are at least 2^t factors of the form $\Phi_d(b)$ appearing in the right hand side of formula (5.1). We now use the fact well known fact that for $k > 12$ the number $b^k - 1$ has a *primitive divisor*; i.e., the above number is divisible by a prime number p such that p does not divide anyone of the numbers $b^\ell - 1$ for any positive integer $\ell < k$ (see [5]). It is also well known that the existence of a primitive divisor for the number $b^k - 1$ is equivalent to the existence of a prime divisor p of $\Phi_k(b)$ such that p does not divide $\Phi_\ell(b)$ for any positive integer $\ell < k$. This argument together with (5.1) tells us that $\omega(n_z) \ge 2^t - 12$. It remains to get a lower bound on t in terms of n_z. But this is easy. Clearly, the inequality

$$s\mu < \exp(2t \log t)$$

holds for all sufficiently large values of z and since $n_z < b^{s\mu}$ we get, by applying the double logarithm in the above inequality, that

$$\log_2 n_z < \log(s\mu) + \log_2 b < 2t \log t + \log_2 b,$$

therefore the inequality

$$c_{33} \log_2 n_z < t \log t$$

holds for large values of z where c_{33} can be taken to be $1/3$. The above inequality implies that the inequality

$$c_{34} \frac{\log_2 n_z}{\log_3 n_z} < t$$

holds with some absolute constant c_{34} provided that z is large enough, therefore the inequality

$$\omega(n_z) > 2^t - 12 > (\sqrt{2})^t = \exp\left(\frac{\log 2}{2} t\right) > \exp\left(c_{35} \frac{\log_2 n_z}{\log_3 n_z}\right)$$

holds for all sufficiently large values of z with $c_{35} = \frac{c_{34} \log 2}{2}$ which finishes the argument. ∎

6. The Proof of Proposition 1.5

We start with the following presumably well known lemma concerning the structure of positive integers n with a large $\Omega(n)$.

Lemma 6.1. **(1)** *The inequality*

$$\Omega(n) \leq \frac{\log n}{\log 2}$$

holds for all positive integers n.

(2) *Let K be any positive real number in the interval $\left(0, \frac{1}{\log 2}\right)$ and let A_K be the set of all positive integers n such that $\Omega(n) \geq K \log n$. Then A_K is infinite and there exist two computable positive constants L and δ with $\delta < 1$ depending only on K such that if $n \in A_K$, then there exists a prime number $p < L$ such that if we write $n = p^{\alpha_p} m$ where $\gcd(p, m) = 1$, then $\log m < \delta \log n$.*

Proof. Part 1 is well known. Notice that the inequality asserted at 1 is sharp and equality is obtained for all positive integers n which are powers of 2. We shall now prove part 2. Write

$$n = \prod_{p|n} p^{\alpha_p}.$$

If $n \in A_K$, then

$$\sum_{p|n} \alpha_p = \Omega(n) > K \log n = K \sum_{p|n} \alpha_p \log p = \sum_{p|n} \alpha_p (K \log p). \tag{6.1}$$

Let $2 = p_1 < p_2 < \dots$ be all the prime numbers and let k be the maximal positive integer such that the inequality $K \log p_k \leq 1$ holds. Notice that k exists and $k \geq 1$ because $K < \frac{1}{\log 2}$. Write $q = p_k$ and rewrite (6.1) as

$$\sum_{\substack{p|n \\ p \leq q}} (1 - K \log p) \alpha_p > \sum_{\substack{p|n \\ p > q}} \alpha_p (K \log p - 1). \tag{6.2}$$

Set

$$\alpha = \max \{\alpha_p \mid p \mid n \text{ and } p \leq q\},$$
$$K_1 = \sum_{p \leq q} (1 - K \log p) > 0.$$

Inequality (6.2) implies

$$\alpha K_1 \geq \sum_{\substack{p|n \\ p\leq q}} \alpha_p(1 - K\log p) > \sum_{\substack{p|n \\ p>q}} \alpha_p(K\log p - 1). \tag{6.3}$$

Set

$$K_2 = K - \frac{1}{\log p_{k+1}}$$

and notice that $K_2 > 0$. Moreover,

$$K_2 \leq K - \frac{1}{\log p} \qquad \text{holds for all } p > q,$$

therefore

$$K\log p - 1 > K_2 \log p \qquad \text{holds for all } p > q. \tag{6.4}$$

With (6.3) and (6.4) we get

$$K_1\alpha > K_2 \sum_{\substack{p|n \\ p>q}} \alpha_p \log p. \tag{6.5}$$

Finally, set

$$K_3 = \sum_{p\leq q} \log p$$

and notice that (6.5) implies

$$\begin{aligned}(K_1 + K_2K_3)\alpha = K_1\alpha + K_2(K_3\alpha) &\geq K_2 \sum_{\substack{p|n \\ p\leq q}} \alpha \log p + K_2 \sum_{\substack{p|n \\ p>q}} \alpha_p \log p \\ &\geq K_2 \sum_{p|n} \alpha_p \log p = K_2 \log n.\end{aligned}$$

Hence, the inequality

$$\alpha > K_4 \log n$$

holds with $K_4 = \frac{K_2}{K_1+K_2K_3}$. Now let $n \in A_K$ and let $p \leq q$ be the prime number $p \mid n$ for which $\alpha_p = \alpha$. We then have

$$n = p^\alpha m$$

where

$$\begin{aligned}\log m = \log n - \alpha \log p &< \log n - K_4 \log p \log n \\ &= (1 - K_4 \log p)\log n \leq (1 - K_4 \log 2)\log n\end{aligned}$$

which proves part 2 of the above Lemma 6.1 with $L = q$ and $\delta = 1 - K_4 \log 2$ (it is clear that δ is positive because $K_3 \geq \log 2$, therefore $K_1 + K_2K_3 > K_2 \log 2$ which is equivalent to the fact that $K_4 \log 2$ is smaller than 1). Lemma 6.1 is therefore proved. ∎

We can now embark in the proof of Proposition 1.5. Assuming that (1.3) does not hold we deduce the existence of a positive constant K such that the inequality $\Omega(n) > K \log n$ holds for infinitely $n \in A_{b,s}$. We may, of course, assume that K is as small as we want and so we take it such as $K < \frac{1}{\log 2}$. We now use Lemma 6.1 to conclude that there exist two positive constants L and δ such that for every $n \in A_{b,s}$ for which $\Omega(n) > K \log n$ we have $n = p^\alpha m$ with $\gcd(p, m) = 1$ and $\log m < \delta \log n$. In particular, for every such n we get a relation of the type

$$n = a_0 + a_1 b^{i_1} + \cdots + a_t b^{i_t} = p^\alpha m \tag{6.6}$$

where $a_i \in \{1, \ \ldots, \ b-1\}$, $\sum_{i=1}^{t} a_i = s$, $p < L$, $p \not| \, m$ and the inequality $\log m < \delta \log n$ holds with some fixed value of $\delta < 1$. There are only finitely many possibilities of choosing $t \leq s-1$ and the $t+2$-uple of positive integers $(a_0, a_1, \ldots, a_t, p)$ such that $a_i \in \{1, \ \ldots, \ b-1\}$ with $\sum_{i=0}^{t} a_i = s$ and $p < L$ is a prime number and since we are assuming that we have infinitely many values of n satisfying a relation of the type (6.6), we may assume that $t, \ a_0, \ \ldots, \ a_t, \ p$ are fixed and that (6.6) holds for infinitely many values of the $t+2$-uple of positive integers $(i_1, \ldots, i_t, \alpha, m)$. In order to get a contradiction, we shall show that an equation like (6.6) has only *finitely many* solutions $(i_1, \ldots, i_t, \alpha, m)$. To achieve this, we shall prove something slightly more general, namely the following.

Let $b > 1$ be a fixed positive integer. Let $t \geq 1$ and $A_0, \ A_1, \ \ldots, \ A_t$ be fixed non zero integers and p be a fixed prime. Let $\delta < 1$ be a fixed real number and consider the Diophantine equation

$$A_0 + A_1 b^{i_1} + \cdots + A_t b^{i_t} = p^\alpha m \tag{6.7}$$

in integer unknowns $(i_1, \ldots, i_t, \alpha)$ with $i_j \geq 0$ for $j = \{1, \ \ldots, \ t\}$, $\alpha \geq 0$, $m \neq 0$, p and m coprime and $\log |m| < \delta \log |n|$ where we use n to denote the common value of the integer appearing in either side of equality (6.7). A solution $(i_1, \ldots, i_t, \alpha, m)$ will be called *non degenerate* if

$$\sum_{j \in I} A_j \, b^{i_j} \neq 0$$

holds for all proper non empty subsets $I \subset \{0, 1, \ldots, t\}$ where we use $i_0 = 0$. It suffices to show that an equation of the type (6.7) has only *finitely many* non degenerate solutions $(i_1, \ldots, i_t, \alpha, m)$. Notice that our equation (6.6) is a particular type of an equation of the type (6.7) and it is non degenerate because all the coefficients a_i for $i = \{0, \ 1, \ \ldots, \ t\}$ in (6.6) are positive. Notice that what we are about to prove is a slight generalization of the well known result concerning the finiteness of the number of non degenerate solutions of a linear $\mathcal{S}$-unit equation.

To prove this, we shall apply the following particular case of the Subspace Theorem which we recall as a lemma (for a proof see [20]).

Lemma 6.2. *Let $\mathcal{S}$ be a finite set of absolute values of $\mathbb{Q}$ including ∞ (and normalized such that $|p|_p = p^{-1}$ holds for every prime number p) and let $N \in \mathbb{N}$. For $v \in \mathcal{S}$ let $L_{1,v}, \ldots, L_{N,v}$ be linearly independent forms in N variables with rational coefficients and let $\mu > 0$. Then the solutions $\mathbf{x} = (x_1, \ldots, x_N) \in (\mathbb{Z}^*)^N$ of the inequality*

$$\prod_{v \in \mathcal{S}} \prod_{i=1}^{N} |L_{i,v}(\mathbf{x})|_v < (\max\{|x_i| \mid i = 1, \ldots, N\})^{-\mu}$$

are contained in finitely many proper subspaces of $\mathbb{Q}^N$.

To prove the statement about the finiteness of the number of non degenerate solutions of (6.7) we use induction over the parameter $t \geq 1$ and the above Lemma 6.2. To see how this works let $t = 2$. Then (6.7) reduces to

$$A_1 b^{i_1} + A_0 = p^{\alpha} m. \tag{6.8}$$

Since $\alpha \log p > (1 - \delta) \log |n|$ and $A_0 \neq 0$, it follows that the only interesting case to consider is when p does not divide b (otherwise, we obviously have only finitely many solutions (i_1, α, m) of (6.8)). We apply Lemma 6.2 with $\mathcal{S} = \{p, \ \infty, \ q \mid \text{for all prime divisors } q \text{ of } b\}$, $N = 2$, $\mathbf{x} = (x_1, x_2)$ and the linear forms $L_{1,v}(\mathbf{x}) = x_1$ for all $v \in S$, $L_{2,v} = A_1 x_1 - x_2$ for $v \in S \backslash \{p\}$ and $L_{2,v}(\mathbf{x}) = x_2$ when $v = p$. It is clear that $L_{1,v}(\mathbf{x})$ and $L_{2,v}(\mathbf{x})$ are linearly independent for all $v \in \mathcal{S}$. We assume that $x_1 = b^{i_1}$ and $x_2 = p^{\alpha} m$ is a solution of

$$A_1 x_1 - x_2 = -A_0. \tag{6.9}$$

Let us compute the double product

$$\prod_{v \in \mathcal{S}} \prod_{i=1}^{2} |L_{i,v}(\mathbf{x})|_v$$

for our particular problem. Clearly,

$$\prod_{v \in \mathcal{S}} |L_{1,v}(\mathbf{x})|_v = \prod_{v \in \mathcal{S}} |x_1|_v = 1$$

and

$$\prod_{v \in \mathcal{S} \backslash \{p\}} |L_{2,v}(\mathbf{x})|_v = \prod_{v \in \mathcal{S} \backslash \{p\}} |-A_0|_v = A$$

where A is a constant. Finally,

$$|L_{2,p}|_p = |x_2|_p = \frac{1}{p^\alpha} < |n|^{-(1-\delta)} = |A_1x_1 + A_0|^{-(1-\delta)}.$$

Thus, the inequality

$$\prod_{v\in\mathcal{S}}\prod_{i=1}^{2}|L_{i,v}(\mathbf{x})|_v < A|A_1x_1 + A_0|^{-(1-\delta)} < \max\{|x_1|, |x_2|\}^{-\mu}$$

holds with $\mu = \frac{(1-\delta)}{2} > 0$ provided that $\max\{|x_1|, |x_2|\}$ is large enough. By Lemma 6.2 it follows that all but finitely many solutions of equation (6.9) are contained in finitely many proper subspaces of $\mathbb{Q}^2$. In particular, there exist finitely many pairs of rational numbers (B_1, B_2) not both zero so that $B_1x_1 + B_2x_2 = 0$. Thus, $B_1b^{i_1} + B_2p^\alpha m = 0$ which means that α is bounded and since $\alpha \log p > (1-\delta)\log|n|$, we get that n is bounded as well. This takes care of the case in which $t = 1$.

We now treat the general case. Assume that $t > 1$ and that equation (6.7) does have only finitely many non degenerate solutions for any $t' < t$ and any choice of the non zero coefficients $A_0, \ldots, A_{t'}$. We may assume that i_j are distinct because if two of them are equal, say $i_1 = i_2$, we can then group them and rewrite equation (6.7) as

$$A_0 + (A_1 + A_2)b^{i_2} + \cdots + A_tb^{i_t} = p^\alpha m$$

which is an equation like (6.7) but with fewer (i.e., $t-1$) terms in the left hand side and it is still non degenerate (notice that $A_1 + A_2 \neq 0$ if $i_1 = i_2$ because we are treating only the case of the non degenerate solutions of (6.7)). The same argument shows that we may assume that $i_j > 0$ for all $j \in \{1, \ldots, t\}$ and we may now order them in such a way that $0 < i_1 < i_2 < \cdots < i_t$. We may also assume that i_1 is as large as we want for if not, i.e., if i_1 remains bounded, we may then assume that i_1 is fixed and replace $A_0 + A_1b^{i_1}$ by A_0' and obtain again an equation like (6.7) but with fewer terms in the left hand side. Since we are assuming that i_1 can be large, it follows that the only interesting case is again when p does not divide b. We apply Lemma 6.2 with $S = \{p, \infty, q \mid \text{for all prime divisors } q \text{ of } b\}$, $N = t+1$, $L_{i,v}(\mathbf{x}) = x_i$ for all $i = 1, \ldots, t$ and all $v \in \mathcal{S}$, $L_{t+1,v}(\mathbf{x}) = A_1x_1 + A_2x_2 + \cdots + A_tx_t - x_{t+1}$ for all $v \in \mathcal{S}\backslash\{p\}$ and $L_{t+1,p}(\mathbf{x}) = x_{t+1}$. It is clear that $L_{i,v}$ for $i = 1, \ldots, N$ are linearly independent for all $v \in \mathcal{S}$. We look at the solutions of the equation

$$A_1x_1 + A_2x_2 + \cdots + A_tx_t - x_{t+1} = -A_0 \tag{6.10}$$

with $x_j = b^{i_j}$ for $j = 1, \ldots, t$ and $x_{t+1} = p^\alpha m$ where

$$\log|m| < \delta\log|n| = \delta\log|A_0 + A_1x_1 + \cdots + A_tx_t|.$$

We now compute again the double product

$$\prod_{v\in\mathcal{S}}\prod_{i=1}^{N}|L_{i,v}(\mathbf{x})|_v.$$

Let $j\le t$. Then

$$\prod_{v\in\mathcal{S}}|L_j(\mathbf{x})|_v=\prod_{v\in\mathcal{S}}|x_j|_v=\prod_{v\in\mathcal{S}}|b^{i_j}|_v=1. \tag{6.11}$$

Let now $j=t+1$ and notice that

$$\prod_{v\in\mathcal{S}\setminus\{p\}}|L_{t+1,v}(\mathbf{x})|_v=\prod_{v\in\mathcal{S}\setminus\{p\}}|-c_0|_v=A' \tag{6.12}$$

where A' is a constant while

$$|L_{t+1,p}(\mathbf{x})|_p=|x_{t+1}|_p=|p^\alpha m|_p=\frac{1}{p^\alpha}<|A_1x_1+\cdots+A_tx_t|^{-(1-\delta)}. \tag{6.13}$$

Multiplying (6.11)–(6.13) we get that the inequality

$$\begin{aligned}\prod_{i\in\mathcal{S}}\prod_{i=1}^{N}|L_{i,v}(\mathbf{x})|_v &< A'|A_1x_1+\cdots+A_tx_t|^{-(1-\delta)}\\ &< \max\{\,|x_j|\mid j=1,\ \ldots,t+1\}^{-\mu}\end{aligned}$$

holds with $\mu=\frac{(1-\delta)}{2}$ provided that i_t-i_{t-1} is large enough (we may assume that this is so for otherwise, if i_t-i_{t-1} is bounded, we may then assume that i_t-i_{t-1} is fixed and then equation (6.7) becomes again an equation with fewer unknowns in the left hand side of it which has only finitely many non degenerate solutions by the induction hypothesis). With Lemma 6.2 it follows that there exist finitely many non zero vectors $(B_1,\ \ldots,\ B_{t+1})$ in $\mathbb{Q}^{t+1}$ such that every solution of (6.10) satisfies

$$B_1x_1+\cdots+B_tx_t+B_{t+1}x_{t+1}=0. \tag{6.14}$$

Assume first that $B_{t+1}=0$ and let j be the maximal index $\le t$ for which $B_j\ne 0$. Then equation (6.14) implies that

$$x_j=B'_1x_1+\cdots+B'_{j-1}x_{j-1}$$

where $B'_i=-\frac{B_i}{B_j}$ and the above equation is a pure $\mathcal{S}$-unit equation. It is then known that i_j-i_{j-1} is bounded and so we may assume that i_j-i_{j-1} is fixed and now equation (6.7) becomes an equation with fewer than t

unknowns in the left hand side of it which by the induction hypothesis has only finitely many solutions. Assume now that $B_{t+1} \neq 0$. Expressing x_{t+1} versus $x_1, \ldots, x_t$ both from (6.14) and from (6.10) we get an equation of the form

$$A_1x_1 + \cdots + A_tx_t + A_0 = B'_1x_1 + \cdots + B'_tx_t \tag{6.15}$$

where $B'_i = -\frac{B_i}{B_{t+1}}$ for $i = 1, \ldots, t$. The above equation (6.15) is an equation of the form

$$C_1x_1 + \cdots + C_tx_t + C_0 = 0 \tag{6.16}$$

with $C_i = A_i - B'_i$ for all $i = 1, \ldots, t$. If (6.16) is non degenerate, then it has only finitely many solutions with $x_j = b^{i_j}$ for $j = 1, \ldots, t$ (notice that $A_0 \neq 0$ is fixed). If it is degenerate, let I be a set of minimal cardinality of $\{1, \ldots, t\}$ for which

$$A_0 + \sum_{i \in I} C_ix_i = 0. \tag{6.17}$$

Then equation (6.17) is non degenerate and as such it has only finitely many solutions $x_j = b^{i_j}$ for which $j \in I$. In particular, since I is non empty, every one of the exponents i_j for $j \in I$ is bounded which means that we can reduce equation (6.7) again to an equation with fewer unknown terms in the left hand side of it. The induction is therefore complete which finishes the proof of Proposition 1.5. ■

Remarks With a known value of δ it is an immediate application of lower bounds for p-adic logarithms that an equation like (6.7) with $t = 1$ has even finitely many *effectively computable* solutions (i_1, α, m) and these may be computed from knowledge of c_0, c_1, b, δ and p. However, the value of δ comes from the Lemma 6.1 and it is computable only if we assume that we do know a constant K such that $\Omega(n) > K \log n$ holds with some $n \in A_{b,s}$ which has exactly two non zero digits in base b. We have avoided to mention the words lower bounds for linear forms in logarithms at this stage mainly because we wanted only to prove an estimate of the type (1.3) which means that we wanted to show that for every $K > 0$ the inequality $\Omega(n) > K \log n$ can hold only for finitely many positive integers $n \in A_{b,s}$. However, with our method one can employ a lower bound for a linear form in p-adic logarithms to find an effective upper bound on $\frac{\Omega(n)}{\log n}$ which tends to zero with n when n tends to infinity in the set of all positive integers which are not multiples of b and which have only two non zero digits when written in base b.

7. The Proof of Proposition 1.6

By the arguments used in the proof of Proposition 1.2, it follows that for a large positive real number x we have $c_{16}\log^{s-1} x < |A_{b,s}(x)| < c_{17}\log^{s-1} x$. We thus find it easier to measure the cardinality of $|A_{b,s}(x)|$ in the *logarithmic scale*. That is, for a large positive real number x we shall count the number of numbers $n \in A_{b,s}$ of the form

$$n = a_0 + a_1 b^{i_1} + \cdots + a_t b^{i_t} \tag{7.1}$$

where $a_i \in \{1,\ \ldots,\ b-1\}$, $\sum_{i=0}^{t} a_i = s$ and $0 < i_1 < \cdots < i_t < x$. This is "almost" $A_{b,s}(b^x)$ and therefore the number of such numbers n is $\asymp x^{s-1}$. We have to show that the number of such numbers n which are also perfect powers is $o(x^{s-1})$. First of all let us notice that most numbers of the form (7.1) have $t = s-1$ and $a_i = 1$ for all $i = 0,\ \ldots,\ t$. Indeed, suppose that $t < s-1$. For each such t there are only finitely many choices for the coefficients $a_i \in \{1,\ \ldots,\ b-1\}$ such that $\sum_{i=0}^{t} a_i = s$ and for each one of these choices of the coefficients the exponents $i_1 < i_2 < \cdots < i_t < x$ can be chosen in at most $x^t \le x^{s-2}$ ways. In particular, the number of numbers n given by (7.1) with $t < s-1$ is $O(x^{s-2}) = o(x^{s-1})$. Thus, most numbers shown at (7.1) have $t = s-1$ and therefore $a_i = 1$ for all $i = 0,\ \ldots,\ t$. Let us assume now that

$$1 + b^{i_1} + \cdots + b^{i_{s-1}} = m^y \tag{7.2}$$

holds for some positive integers $m > 1$ and $y > 1$. We may assume that y is prime and since $m \ge 2$, we get that $y \le c_{36}x$. Thus, the number of ways of choosing y is $\pi(c_{36}x) < c_{37}\frac{x}{\log x}$. Since $0 < i_1 < \cdots < i_{s-1}$, we also observe that we may assume that $i_1 > \frac{x}{\log x}$ for if not, then the number of numbers n of the form (7.1) with $i_1 \le \frac{x}{\log x}$ and $i_2 < i_3 < \cdots < i_{s-1} < x$ is, of course, $O\big(\frac{x^{s-1}}{\log x}\big) = o(x^{s-1})$. In particular, since $i_1 > 0$, we read that the number shown at (7.2) is coprime to b, therefore m is coprime to b. We now fix the exponents $i_1 < i_2 \cdots < i_{s-2} < x$. The number of such choices is, of course, $O(x^{s-2})$. We write

$$c = 1 + b^{i_1} + \cdots + b^{i_{s-1}}.$$

And it thus suffices to count the number of exponents $j = i_{s-1} < x$ with $j > \frac{x}{\log x}$ and $j = i_{s-1} \notin \{i_1,\ \ldots,\ i_{s-2}\}$ and for which the relation

$$b^j + c = m^y \tag{7.3}$$

holds with some positive integer $m > 1$ coprime to b and some prime number $y > 1$. We need to show that the number of such exponents j is $o(x)$. For this

we start with the exponent y. Since y is prime, we write it as $y = p$. For any fixed $p < c_{36}x$ equation (7.3) can have no solutions (j, m) with $j < x$ or one such solution or more than one such solution. Thus, the total number of solutions (j, m, y) of equation (7.3) over all those values of $y = p < c_{36}x$ for which equation (7.3) has at most one solution (j, m) when p is fixed is at most $\pi(c_{36}x) = O\big(\frac{x}{\log x}\big) = o(x)$. We now fix a value of $p < c_{36}x$ and we assume that equation (7.3) has more than one solution (j, m) with this fixed value of $y = p$.

We distinguish three instances.

Case 1. $p > \max\{b, 3\}$.

Let (j, m) and (j', m') be two solutions of equation (7.3) with $j' > j$ corresponding to the same value $y = p$. Taking the difference between these two equations (7.3) we get

$$b^j(b^{j'-j} - 1) = m'^y - m^y = (m' - m)\frac{m'^p - m^p}{m' - m}. \tag{7.4}$$

Let $d = \gcd(m', m)$. Notice that d is coprime to b. Write $m' = dm'_1$ and $m = dm_1$ with coprime positive integers m_1 and m'_1. Equation (7.4) is of the form

$$b^j(b^{j'-j} - 1) = d^p(m'_1 - m_1)\frac{m'^p_1 - m^p_1}{m'_1 - m_1}. \tag{7.5}$$

It is now clear that every prime divisor of $\frac{m'^p_1 - m^p_1}{m'_1 - m_1}$ is either p (and this happens if $p \mid (m'_1 - m_1)$), or is congruent to 1 modulo p. Since $p > b$, we get that $\frac{m'^p_1 - m^p_1}{m'_1 - m_1}$ is coprime to b^j and therefore

$$\frac{m'^p_1 - m^p_1}{m'_1 - m_1} \le b^{j'-j} - 1.$$

If $d > 1$, then d is also coprime to b and therefore d^p divides $b^{j'-j} - 1$. In this case, it follows, by (7.5) and the above remarks, that $m'_1 - m_1$ must be a multiple of b^j and therefore

$$m'_1 - m_1 \ge b^j.$$

We thus get

$$b^{j(p-1)} \le (m'_1 - m_1)^{p-1} \le \frac{m'^p_1 - m^p_1}{m'_1 - m_1} \le b^{j'-j} - 1 < b^{j'-j},$$

therefore

$$j' \ge jp. \tag{7.6}$$

Since $j > \frac{x}{\log x}$ and $j' < x$, we first read that $p < \log x$. Secondly, assume that $j_1 < j_2 < \cdots < j_k$ are all the possible indices $j \in \left(\frac{x}{\log x},\ x\right)$ for which $b^j + c$ is a pth power. Inequality (7.6) tells us that $j_2 > pj_1$, then that $j_3 > pj_2 > p^2 j_1$, etc., therefore

$$j_k > p^k j_1.$$

Since $j_k < x$ and $j_1 > \frac{x}{\log x}$, we read that $p^k < \log x$, therefore $k \le c_{38} \log_2 x$ where $c_{38} = \frac{1}{\log 2}$. Thus, the number of numbers $y = p$ is at most $O(\log x)$ and for each such fixed prime number p the number of indices j for which $b^j + c$ can be a pth power is $O(\log_2 x)$. The totality of all these instances is $O(\log x \log_2 x) = o(x)$ which settles this case.

Case 2. $3 \le p \le \max\{3, b\}$.

In this case, we have only finitely many prime numbers p under consideration so we may assume that p is fixed. Let again (j, m) and (j', m') be two solutions of equation (7.3) with $j' > j$ corresponding to the same exponent $y = p$. With $d = \gcd(m', m)$, $m' = dm_1'$ and $m = dm_1$ we arrive again at equation (7.5), namely

$$b^j(b^{j'-j} - 1) = d^p(m_1' - m_1)\frac{m_1'^p - m_1^p}{m_1' - m_1}. \tag{7.7}$$

Since both m' and m are coprime to b, it follows that d is coprime to b and so b^j divides $m_1'^p - m_1^p$. In particular,

$$m_1'^p > m_1'^p - m_1^p \ge b^j > b^{x/\log x},$$

therefore

$$m_1' > \exp\left(c_{39}\frac{x}{\log x}\right) \tag{7.8}$$

with $c_{39} = \frac{\log b}{\max\{b,3\}}$. Since $p \ge 3$ is fixed and $m_1' > m_1$ are coprime, it follows by a result of Bugeaud (see [4]) that the inequality

$$P(m_1'^p - m_1^p) > c_{40} \log_2 m_1' \tag{7.9}$$

holds with some computable constant c_{40}. It then follows, from the above estimates (7.8) and (7.9), that the inequality

$$P(m_1'^p - m_1^p) > c_{40} \log\left(c_{39}\frac{x}{\log x}\right) > c_{41} \log x \tag{7.10}$$

holds with $c_{41} = \frac{c_{40}}{2}$ provided that x is large enough.

If x is so large such that $c_{41}\log x > b$ holds, we then get, by (7.7) and (7.10), that the largest prime divisor of $m_1'^p - m_1^p$ must divide $b^{j'-j}-1$ and therefore

$$b^{j'-j} > b^{j'-j} - 1 \geq P(m_1'^p - m_1^p) > c_{41}\log x.$$

Thus, the inequality

$$j' - j > c_{42}\log_2 x$$

must hold with $c_{42} = \frac{1}{2\log b}$ provided that x is large enough. The above *gap principle* tells us that if p is fixed and $\frac{x}{\log x} < j_1 < j_2 < \cdots < j_k < x$ are all the possible exponents in the interval $\left(\frac{x}{\log x}, x\right)$ for which $b^j + c$ is a pth power, then $k = O\left(\frac{x}{\log_2 x}\right) = o(x)$. Since p can take only finitely many values, it follows that the total number of exponents $j < x$ for which $b^j + c$ can be a pth power is again $o(x)$ which takes care of this case.

Case 3. $p = 2$.

None of the above techniques can deal with the case $p = 2$ so here we will employ a different argument. We let x be large and write $z = \left\lfloor \frac{x}{\log x} \right\rfloor$. With $j > \frac{x}{\log x}$ every positive integer solution (j, m) of the equation

$$b^j + c = m^2$$

leads to a solution of the congruence

$$m^2 \equiv c \pmod{b^z}. \tag{7.11}$$

Let $b = \prod_{q|b} q^{\alpha_q}$. Equation (7.11) implies that

$$m^2 \equiv c \pmod{q^{\alpha_q z}} \tag{7.12}$$

and for every fixed $q|b$ the above equation (7.12) has precisely two solutions $m \pmod{q^{\alpha_q z}}$ (when $q = 2$ the above equation has only two solutions modulo $2^{\alpha_2 z - 1}$ and so that it has at most 4 solutions modulo $2^{\alpha_2 z}$). Here, we implicitly used the fact that c is coprime to b. The above argument shows that the system of congruences (7.12) has, with varying q and by the Chinese Remainder Lemma, at most $c_{43} = 2^{\omega(b)+1}$ integer solutions m in the interval $(0, b^z)$. So, it suffices to count the number of exponents j such that $b^j + c = m^2$ holds with m in a fixed congruence class modulo b^z. Assume that this fixed congruence class is c', where c' is an integer in the interval $(0, b^z)$ which is coprime to b and suppose that (j, m) and (j', m) are both

solutions of equation (7.3) with $y = 2$ and $m' \equiv m \equiv c' \pmod{b^z}$. We write $m = ub^z + c',\ m' = vb^z + c'$ with $v > u \geq 0$. We now get

$$\begin{aligned} b^j(b^{j'-j} - 1) = m'^2 - m^2 &= (m' - m)(m' + m) \\ &= b^z(v - u)(b^z(v + u) + 2c'), \end{aligned}$$

therefore

$$b^{j-z}(b^{j'-j} - 1) = (v - u)(b^z(v + u) + 2c').$$

It is now clear that $\gcd(b,\ b^z(v + u) + 2c') \mid 2$ and therefore

$$b^{j'-j} - 1 \geq \frac{b^z(v + u) + 2c'}{2} > \frac{b^z}{2}.$$

We thus get

$$j' - j > \frac{1}{\log b}(z \log b - \log 2) > c_{44}\frac{x}{\log x}$$

where we can take $c_{44} = \frac{1}{2}$ provided that x is large enough. The above argument shows again that equation (7.3) can have at most $O(\log x)$ solutions (j, m) with $p = 2$ and $j < x$ such that m is in a fixed congruence class modulo b^z and since we have only finitely many such congruence classes, we get that the number of numbers $j < x$ for which $b^j + c$ can be a perfect square is $O(\log x) = o(x)$. This case is therefore settled as well which completes the proof of Proposition 1.6. ■

8. The Proof of Proposition 1.7

The proof of this proposition follows closely the proof of the main result in [14] although some of the details of the argument in [14] must be slightly modified for our present purposes. We work again in the logarithmic scale as in the proof of Proposition 1.6. That is, we assume that x is a large positive real number and that n is a number of the form

$$n = a_0 + a_1 b^{i_1} + \cdots + a_t b^{i_t}$$

where $a_i \in \{1,\ \ldots,\ b - 1\}$ for $i = 0,\ \ldots,\ t$, $\sum_{i=0}^{t} a_i = s$ and $0 < i_1 < \cdots < i_t < x$. We want to show that but for a set of such positive integers n of cardinality $o(x^{s-1})$ the inequality asserted at (1.5) holds. As we have seen in the proof of Proposition 1.6 most such positive integers n have $t = s - 1$ and $a_i = 1$ for all $i = 0, \ldots, t$ and therefore we may assume that the numbers n that we will work with are of this type. We may also assume that the inequality $i_{s-1} - i_s > c_{45}\frac{x}{\log_4 x}$ holds where c_{45} is a computable positive constant which we will fix later. Indeed, the reason is that the number of

$s-1$-uples of positive integers $(i_1, \dots, i_{s-1})$ for which $i_1 < \cdots < i_{s-1} < x$ but $i_{s-1} - i_{s-2} \le c_{45} \frac{x}{\log_4 x}$ where c_{45} is some fixed constant is $O\big(\frac{x^{s-1}}{\log_4 x}\big) = o(x^{s-1})$ and the total number of numbers n under consideration is $\sim x^{s-1}$. As in the proof of Proposition 1.2 we shall assume that $i_1 < \cdots < i_{s-2}$ are fixed and we write

$$c = 1 + b^{i_1} + \cdots + b^{i_{s-2}}.$$

We will again assume that $i_1 > \frac{x}{\log x}$. Thus, the number n we are looking at is of the form

$$n = b^j + c \tag{8.1}$$

where

$$\log\left(\frac{b^j}{c}\right) \gg \frac{x}{\log_4 x} \tag{8.2}$$

and where $\frac{x}{\log x} < j < x$ and we want to show that with fixed c and variable j in the above interval such that inequality (8.2) holds, inequality (1.5) also holds for all such j except, eventually, for a number of $o(x)$ of such j. From now on, we shall always use j and n with the meaning that they are related via formula (8.1). We use $f(x)$ to denote some function of x which is increasing for large values of x and tends to infinity with x. We shall try to find the best (i.e., "largest") such function f which comes out of our arguments and for which the inequality

$$P(\phi(n)) < f(x)$$

holds only on a set of positive integers j belonging to the interval $\big(\frac{x}{\log x}, x\big)$ of cardinality $o(x)$. As the conclusion of the Proposition 1.7 suggests our best f is a constant multiple of $\log^{1/6} x \log_2^{1/3} x$. In fact, up to modifying the lower bound on b^j/c suggested by formula (8.2), we can prove that a stronger inequality than (1.5) holds for almost all positive integers $n \in A_{b,s}$, namely that the inequality

$$P(\phi(n)) > \epsilon(n) \log_2^{1/6} n \log_3^{5/6} n$$

holds for almost all $n \in A_{b,s}$ where $\epsilon(n)$ is any function defined on the set of positive integers and which tends to zero arbitrarily slowly when n tends to infinity. We shall resume ourselves to give the proof of (1.5).

We also notice that we may assume right away that $s > 2$. Indeed, for if $s = 2$, then the numbers n we are looking at are of the form

$$n = b^j + 1.$$

It is well known that for $j > 12$ the number $b^j + 1$ has a primitive divisor; i.e., there exists a prime number $p \mid b^j + 1$ such that p does not divide $b^\ell + 1$ for any positive integer $\ell < j$. It is known that such a prime satisfies the congruence $p \equiv 1 \pmod{j}$. In particular, for all but finitely many values of j the inequality $P(n) = P(b^j+1) \geq P(j)$ holds and we may now use a result of de Bruijn (see [8]) which asserts that the inequality $P(j) > x^{\eta(x)}$ holds for all positive integers $j < x$ except for a set of cardinality $O\big(\frac{x}{\log^{2.5} x}\big)$ where we can take $\eta(x) = \frac{\log_4 x}{3 \log_3 x}$. We thus get an even better inequality than the one asserted at (1.5) for $P(\phi(n))$ which holds for almost all positive integers $n \in A_{b,s}$ when $s = 2$. From now on we assume that $s > 2$.

For the moment, we shall work with an unknown function f. Pick a large positive real number x_0 such that $f(x_0) > bc(b-1)(c+1)$, let $x > x_0$ be a large positive real number and let j be a positive integer in the interval $\frac{x}{\log x} < j < x$ for which the inequality $P(\phi(n)) \leq f(x)$ holds. Assume that $p_1 < p_2 < \cdots < p_t \leq f(x)$ are all the prime numbers less that $f(x)$ and set $\mathcal{S} = \{p_1^{\alpha_1} \dots p_t^{\alpha_t} \mid \alpha_i \geq 0\}$ to be the set of all positive integers m with $P(m) \leq f(x)$. Notice that for large x we have $t = \pi(f(x)) \leq \frac{2f(x)}{\log(f(x))}$. Since $\phi(n) \in \mathcal{S}$, we may write

$$n = \prod_{p^\alpha || n} p^\alpha = AB$$

where

$$A = \prod_{\substack{p^\alpha || n \\ \alpha > 1}} p^\alpha \quad \text{and} \quad B = \prod_{p || n} p.$$

Clearly,

$$\phi(n) = \frac{A}{\mathrm{rad}(A)} \phi(\mathrm{rad}(A))\phi(B)$$

where for a positive integer k we write $\mathrm{rad}(k) = \prod_{p|k} p$. Since $p|\phi(n)$ whenever $p|A$, it follows that $A \in \mathcal{S}$. We now bound the size of A. We claim that with any function $f(x)$ such that $f(x) = O(\log_2 x)$ the inequality

$$\alpha_p \leq p \log x \tag{8.3}$$

holds for all $p < f(x)$ and for all $j < x$ except for a subset of j of cardinality $o(x)$. To see why this is so notice first of all that since $i_1 > \frac{x}{\log x} > 0$, it follows that n is coprime to b. Hence, it suffices to prove that inequality (8.3) holds for most values of $j < x$ and for all $p < f(x)$ such that p is coprime to b. For every such prime p let $u(p)$ be the *order of apparition* of p^2 in the Lucas sequence of general term $(b^m - 1)_{m \geq 0}$. That is, $u(p)$ is the smallest positive integer m such that $b^m - 1 \equiv 0 \pmod{p^2}$. This number $u(p)$ exists

because p is coprime to b and by Euler's Theorem it is a divisor of $p(p-1)$. Let $v(p) = \mathrm{ord}_p(b^{u(p)} - 1)$; that is, $v(p)$ is the order at which p appears in the prime factorization of $b^{u(p)} - 1$. Clearly, $v(p) \geq 2$ and $v(p) \ll p(p-1) \leq f(x)^2$. Suppose now that $j < x$ is such that there exists some prime number $p < f(x)$ for which inequality (8.3) fails. Fix such a number p and let $j_0 < x$ be minimal such that with $n_0 = b^{j_0} + c$ we have $p^{\alpha_p} | n_0$ where α_p is larger than $p \log x$. If for every prime number p there exists at most one such positive integer j_0, then the number of such numbers is $\leq t = \pi(f(x)) = o(x)$ and we are done. So, we now fix p and we assume that there is more than one index $j < x$ for which $p^{\alpha_p} | n$ with $\alpha_p > p \log x$. In this case, with $\beta_p = \lfloor p \log x \rfloor$ we get that $p^{\beta_p} \mid b^j - b^{j_0}$. Since $j > j_0$ and p is coprime to b, we get that $p^{\beta_p} \mid b^{j-j_0} - 1$. Let us notice that $p \log x > f(x)^2 \geq p(p-1) > v(p) \geq 2$ and this inequality holds uniformly in $p < f(x)$ and for large x when $f(x) = O(\log_2 x)$. Moreover, let us also notice that

$$\beta_p - v(p) > p \log x - 1 - f(x)^2 \geq 2 \log x - 1 - f(x)^2 \geq \log x. \tag{8.4}$$

From the well known divisibility properties of the Lucas sequence $(b^m - 1)_{m \geq 0}$ it follows that

$$j \equiv j_0 \pmod{u(p) p^{\beta_p - v(p)}}. \tag{8.5}$$

Congruence (8.5) together with inequality (8.4) puts j into an arithmetic progression modulo $u(p)p^{\beta_p - v(p)}$ and the number of such numbers up to x is at most

$$\left\lfloor \frac{x}{u(p) p^{\beta_p - v(p)}} \right\rfloor + 1 \leq \frac{x}{2^{\log x}} + 1 < 2x^{c_{46}} \tag{8.6}$$

where $c_{46} = 1 - \log 2$. Summing up the above inequality (8.6) over all the values of $p < f(x)$ we get that the number of numbers $j < x$ for which inequality (8.3) fails for at least one prime number $p < f(x)$ is $O(x^{c_{46}} f(x)) = o(x)$. Thus, from now on we shall assume that all inequalities (8.3) hold for all $p < f(x)$. Hence,

$$\log A = \sum_{\substack{p < f(x) \\ p^{\alpha_p} || n}} \alpha_p \log p \ll \log x \sum_{p \leq f(x)} p \ll f(x)^2 \log x.$$

Since certainly

$$\log n \gg \frac{x}{\log x}$$

it follows that the inequality

$$\log B = \log n - \log A \gg \frac{x}{\log x} - f(x)^2 \log x \geq \frac{x}{2 \log x} \tag{8.7}$$

holds for any large enough positive real number x provided again that our function $f(x)$ satisfies $f(x) = O(\log_2 x)$. Since our function $f(x)$ will ultimately be given by what is shown in the right hand side of formula (1.5), we may assume that inequality (8.7) does hold. In particular, $B > 1$.

From now on, we write $u_j = b^j + c$ for all $j \geq 0$. Notice that $(u_j)_{j\geq 0}$ is simply a non degenerate binary recurrent sequence whose characteristic equation has roots b and 1 and the numbers n under consideration are simply the numbers which can be members u_j of the sequence $(u_j)_{j\geq 0}$ with some $j < x$. Now let $p > f(x)$ be any prime number which divides some member of the sequence $(u_k)_{k\geq 0}$. For this p we set $r = r(p)$ to be the minimal non negative integer k for which $p|u_k$ and set $d = d(p)$ to be the minimal positive integer k for which p divides the kth term of the Lucas sequence $(L_m)_{m\geq 0}$ of general term

$$L_m = \frac{b^m - 1}{b - 1} \qquad \text{for } m \geq 0.$$

We claim that d exists, that $r < d$ and that $p|u_j$ if and only if $j \equiv r \pmod{d}$. To see this, notice that since $(u_j)_{j\geq 0}$ is periodic modulo p, it follows that infinitely many positive integers j exist such that $p|u_j$. Pick $j_2 > j_1$ to be such that $p|u_{j_2}$, $p|u_{j_1}$ and the difference $j_2 - j_1 = k$ is minimal. In particular,

$$u_{j_2} = u_{j_1+k} = b^{j_1}(b-1)\frac{b^k - 1}{b-1} + (b^{j_1} + c) = b^{j_1}(b-1)L_k + u_{j_1}$$

and since p divides both u_{j_2} and u_{j_1} and $p > bc(b-1)(c+1)$, we read that $p|L_k$. From the well known divisibility properties of Lucas sequence $(L_m)_{m\geq 0}$, it follows that $d|k$ and now the same argument as above shows that if $p|u_j$, then $p|u_{j+d}$ as well. Hence, by the minimality of k, we get $d = k$. Further, by the minimality of r, we get that the number r is less than d, that it is the unique number $l < d$ for which $p|u_l$ and finally that any positive integer j for which $p|u_j$ must be congruent to r modulo d and conversely, if $j \equiv r \pmod{d}$, then $p|u_j$. The fact that $r > 0$ follows from the fact that $p > c + 1 = u_0$.

We now pick q to be the smallest divisor of B for which $d(q) > x^{1/t^2}$. We show that this q exists and we find an upper bound on it. To show that q exists, let

$$C(x) = \prod_{p,\ d(p)\leq x^{1/t^2}} p.$$

Certainly,

$$C(x) \mid \prod_{1\leq d\leq x^{1/t^2}} L_d,$$

therefore

$$\log C(x) \le \log\Bigg(\prod_{1\le d\le x^{1/t^2}} L_d\Bigg) = \sum_{1\le d\le x^{1/t^2}} \log L_d \ll \sum_{1\le d\le x^{1/t^2}} d = O(x^{2/t^2}). \tag{8.8}$$

Set

$$D = \gcd(B, C(x))$$

and write

$$B = DE$$

where obviously

$$E = \prod_{\substack{p>f(x),\ p\mid n\\ d(p)>x^{1/t^2}}} p.$$

By (8.8), we get

$$\log D \le \log C(x) \le c_{47} x^{2/t^2} \tag{8.9}$$

with some constant c_{47} and by (8.7) and (8.9), we get that the inequality

$$\log E = \log B - \log D \ge \frac{x}{2\log x} - c_{47}x^{2/t^2} > \frac{x}{3\log x}$$

holds for sufficiently large x. In particular, such a prime number q exists and we write it as $q = q(j)$ and set $d(j) = d(q(j))$. To get an upper bound on q, write

$$E = q_1 q_2 \dots q_k$$

where $q = q_1 < q_2 < \dots < q_k$ are distinct primes. Certainly, $E \le \phi(n) < n$ and therefore

$$2^k \le E \le n.$$

Thus,

$$k \log 2 \ll \log n < x$$

and hence,

$$k \ll x. \tag{8.10}$$

In fact, using the fact that E is square free one can even infer that the inequality $k \ll \frac{x}{\log x}$ holds, but inequality (8.10) suffices for our purposes.

Now write $F = AD$, therefore $n = EF$ with E and F coprime. Hence,

$$\frac{\phi(n)}{n} = \frac{\phi(F)}{F}\frac{\phi(E)}{E}$$

or

$$1 - \frac{\phi(E)}{E} = 1 - \frac{F}{\phi(F)}\frac{\phi(n)}{n} = \frac{n - \frac{F}{\phi(F)}\phi(n)}{n} = \frac{(b^j - \frac{F}{\phi(F)}\phi(n)) + c}{b^j + c}. \tag{8.11}$$

To get an upper bound on q, we use a lower bound depending on q on the left hand side of (8.11) and an upper bound depending only on x (and $f(x)$) on the right hand side of (8.11). For the left hand side of (8.11) we write

$$1 - \frac{\phi(E)}{E} = 1 - \prod_{i=1}^{k} \left(1 - \frac{1}{q_i}\right). \tag{8.12}$$

In light of the inequality

$$1 - \prod_{i=1}^{k} (1 - x_i) \le \sum_{i=1}^{k} x_i \tag{8.13}$$

which holds for all $k \ge 1$ and all real numbers $x_i \in (0,1)$ for $i = 1, \ldots, k$ and which can be immediately proved by induction on k we get, from (8.12), (8.13) and (8.10) that

$$1 - \frac{\phi(E)}{E} = 1 - \prod_{i=1}^{k} \left(1 - \frac{1}{q_i}\right) \le \sum_{i=1}^{k} \frac{1}{q_i} \ll \frac{x}{q}. \tag{8.14}$$

We now need a lower bound for the right hand side of (8.11). We look at the expression

$$b^j - \frac{F}{\phi(F)} \phi(n). \tag{8.15}$$

Assume first that the expression appearing at (8.15) is zero. In this case, we get

$$b^j = \frac{F}{\phi(F)}\, \phi(n) = F\phi(E)$$

therefore $F|c$. Since $F|(b^j + c)$, we read that $F|b^j$ and since $\gcd(b^j, c) = 1$, we get $F = 1$. Thus, equation (8.15) becomes

$$c = \phi(E)$$

and formula (8.11) becomes

$$1 - \frac{\phi(E)}{E} = \frac{b^j}{n} \ll \frac{c}{b^j}.$$

On the other hand, since

$$1 - \frac{\phi(E)}{E} = 1 - \prod_{i=1}^{k} \left(1 - \frac{1}{q_i}\right) \ge \frac{1}{q_1},$$

we get

$$\frac{1}{q_1} \ll \frac{c}{b^j}.$$

Using (8.2), we get that the inequality

$$q_1 > \exp\left(\frac{c_{48}x}{\log_4 x}\right)$$

holds for all x sufficiently large where the constant c_{48} can be chosen to be $\frac{c_{45}}{2}$. So,

$$2b^x > n = E \geq q_1^k > \exp\left(\frac{kc_{48}x}{\log_4 x}\right)$$

which implies that $k < c_{49}\log_4 x$. We now write

$$c = E - \phi(E) = q_1 q_2 \dots q_k - \phi(E),$$

or

$$c = \prod_{i=1}^{k}((q_i - 1) + 1) - \phi(E) = \sum_{\substack{I\subset\{1,2,\dots,k\}\\ I\neq\{1,2,\dots,k\}}} \prod_{i\in I}(q_i - 1). \tag{8.16}$$

Assume that $k = 1$. In this case, equation (8.16) becomes $c = 1$ meaning $s = 2$ which is a case already treated.

Assume now that $k > 1$. We fix k such that $k < c_{49}\log_4 x$. Recalling that $q_i - 1 \in \mathcal{S}$, equation (8.16) is a particular case of an equation of the type

$$c = \sum_{\substack{I\subset\{1,2,\dots,k\}\\ I\neq\{1,2,\dots,k\}}} x_I$$

in $2^k - 1 \geq 3$ indeterminates $x_I = \prod_{i\in I}(q_i - 1)$ for a proper subset I of $\{1, 2, \dots, k\}$ (including the subset $I = \emptyset$ for which $x_\emptyset = 1$). Since $x_I > 0$ for all I, it follows that the above equation is *non degenerate* in the sense that no proper sub sum of the form $x_{I_1} + \cdots + x_{I_j}$ vanishes. We now use a result of Schlickewei (see [19]) on the number of non degenerate solutions of $\mathcal{S}$-unit equations. That is, if $\gamma_1, \dots, \gamma_m$ are fixed non zero rational numbers, then there exists at number of at most

$$\ell \leq \exp\left(2^{37m}t^6\log(8t)\right) \tag{8.17}$$

non degenerate solutions $(x_1^{(j)}, \dots, x_n^{(j)})$ with $j = 1, 2, \dots, \ell$ of the equation

$$\sum_{i=1}^{m}\gamma_i x_i = 0 \quad \text{with } x_i \in \mathcal{S} \text{ for } i = 1, \dots, m \tag{8.18}$$

such that for any other non degenerate solution $(x_1,\ \ldots,\ x_m)$ of equation (8.18) there exists a number $\rho \in \mathcal{S}$ and a number $\lambda \leq \ell$ such that $(x_1,\ \ldots,\ x_m) = \rho(x_1^{(\lambda)},\ \ldots,\ x_n^{(\lambda)})$. Let us notice that from the above result it follows that equation (8.16) can have at most ℓ solutions j where ℓ is bounded above as in (8.17) with $m = 2^k \leq 2^{c_{49}\log_4 x}$. Indeed, we can label the indeterminates x_I for $I \subset \{1,\ \ldots,\ k\}$ such that $1 = x_\emptyset = x_1$, so that if equation (8.16) has more than ℓ solutions, then there must exist two solutions $(x_1,\ \ldots,\ x_{2^k})$ and $(x'_1,\ \ldots,\ x'_{2^k})$ and a rational number $\rho \neq 1$ composed only from the primes $p_1,\ \ldots,\ p_t$ such that $(x'_1,\ \ldots,\ x'_{2^k}) = \rho(x_1,\ \ldots,\ x_{2^k})$. In particular, $1 = x'_1 = \rho x_1 = \rho$ forcing $\rho = 1$ which is a contradiction. This is for a fixed k and now letting k run from 2 to $c_{49}\log_4 x$ we get that the number of solutions of (8.16) with $j < x$ is at most

$$c_{49}(\log_4 x)\exp\big(g(x)t^6\log(8t)\big) \tag{8.19}$$

with $g(x) = 2^{37\cdot 2^{c_{49}\log_4 x}}$ and it is enough for our purposes to check that the number appearing at (8.19) is smaller than $\frac{x}{\log x}$. But this will be so provided that

$$g(x)t^6\log(8t) < \log x - \log_2 x - \log_5 x - \log c_{49}$$

holds and this last inequality will hold provided that

$$t^6\log(8t) < \frac{\log x}{2g(x)} < \frac{\log x}{\log_2 x}. \tag{8.20}$$

The right most inequality asserted at (8.20) will hold provided that c_{49} (i.e., c_{45}) is chosen in such a way that

$$2^{37\cdot 2^{c_{49}\log_4 x}} < \log_2 x$$

which will hold provided that c_{49} (i.e., c_{45}) is chosen to be small enough. Now the remaining inequality (8.20) is fulfilled provided that

$$t < c_{50}\Big(\frac{\log x}{\log_2^2 x}\Big)^{1/6} \tag{8.21}$$

and in order for (8.21) to hold for large x it suffices that the inequality

$$\frac{f(x)}{\log f(x)} < c_{51}\left(\frac{\log x}{\log_2^2 x}\right)^{1/6} \tag{8.22}$$

holds with, say $c_{51} = c_{50}/2$. Clearly, inequality (8.22) holds provided that one chooses

$$f(x) = c_{52}\log^{1/6} x\log_2^{1/3} x \tag{8.23}$$

as stated in the conclusion of Proposition 1.7.

From now on, we may therefore assume that the expression (8.15) is non zero. In this case, we can find a lower bound on the expression appearing at (8.15) by using a linear form in logarithms (see [1]). That is, write

$$\phi(u_j) = p_1^{\beta_1} \dots p_t^{\beta_t}$$

and

$$\left| b^j - \frac{F}{\phi(F)}\phi(n) \right| = b^j \left| 1 - \left(\frac{F}{\phi(F)} \right) p_1^{\beta_1} \dots p_t^{\beta_t} b^{-j} \right|.$$

Since

$$\phi(n) < n \ll b^x,$$

we get that

$$\max_{i=1}^{t} \{\beta_i\} \ll x.$$

For any rational number w let $H(w)$ be the maximum of the absolute values of its numerator and denominator when written in reduced form. Since $F/\phi(F)$ is a rational number which written in reduced form has its numerator greater than its denominator and the numerator is square free and composed of primes less than x^{1/t^2}, we get that

$$\log\left(H\left(\frac{F}{\phi(F)} \right) \right) \ll \sum_{p < x^{1/t^2}} \log p \ll x^{1/t^2}. \tag{8.24}$$

Let

$$\Omega = \prod_{i=1}^{t} \log p_i$$

and notice the following upper bound

$$\Omega \leq \log^t f(x) = \exp(t \log_2(f(x)) < \exp(2t \log_2 t) \tag{8.25}$$

which is valid for large values of x.

With the estimates (8.24), (8.25) and a classical lower bound for linear forms in complex logarithms (like in [1]), we deduce the existence of a constant $c_{53} > 1$ such that

$$\begin{aligned} \left| 1 - \left(\frac{F}{\phi(F)} \right) p_1^{\beta_1} \dots p_t^{\beta_t} b^{-j} \right| &> \exp\left(-t^{c_{53}t} \log\left(H\left(\frac{F}{\phi(F)} \right) \right) \Omega \log j \right) \\ &> \exp\left(-x^{1/t^2} (\log j) \exp\left(c_{53} t \log t + 2t \log_2 t \right) \right) \end{aligned}$$

holds for large enough values of j. Let us observe that

$$\exp\left(c_{53} t \log t + 2t \log_2 t \right) < x^{1/t^2} \tag{8.26}$$

holds for large enough values of x.

Indeed, inequality (8.26) is implied by

$$c_{53}t^2\,(t\log t + 2t\log_2 t) < \log x \tag{8.27}$$

which obviously holds for large values of x because $t = \pi(f(x)) < f(x)$ and $f(x)$ is given by (8.23). Thus, with (8.26) and (8.27) we get that

$$\left|1 - \left(\frac{F}{\phi(F)}\right) p_1^{\beta_1} \dots p_t^{\beta_t} b^{-j}\right| > \exp(-x^{2/t^2}\log j).$$

We now show that the expression appearing at (8.15) is positive. Indeed, assuming that the expression appearing at (8.15) is negative, then from formula (8.11) and the fact that the left hand side of (8.11) is positive we get that

$$c > \left|b^j - \frac{F}{\phi(F)}\phi(n)\right| > b^j \exp(-x^{2/t^2}\log j).$$

The above inequality implies, after rearranging it, taking logarithms and recalling (8.2) that

$$x^{2/t^2}\log x \gg \log\left(\frac{b^j}{c}\right) \gg \frac{x}{\log_4 x}$$

which is impossible for large enough values of x. This shows that for large enough values of x the expression appearing at (8.15) is positive. In particular, the numerator of the expression appearing in the right hand side of (8.11) is at least as large as

$$b^j \exp(-x^{2/t^2}\log j) + c > b^j \exp(-x^{2/t^2}\log j).$$

Since $u_j \ll b^j$, it follows that

$$\frac{\left(b^j - \frac{F}{\phi(F)}\,\phi(n)\right) + c}{b^j + c} \gg \exp(-x^{2/t^2}\log j). \tag{8.28}$$

Combining estimate (8.28) with (8.14) we get $q = q_1 \ll x\exp(x^{2/t^2}\log x)$, so the inequality

$$q = q_1 < \exp(x^{3/t^2}) \tag{8.29}$$

holds for large enough values of x. Since $q - 1 \in \mathcal{S}$, it follows the number of numbers q that can fulfill (8.29) is certainly no more than

$$O(x^{3/t}) = o(x).$$

We now return to the values of j. From what we have said but for $o(x)$ positive integers j in the interval $\left(\frac{x}{\log x}, x\right)$ for which $\phi(n) \in \mathcal{S}$ a prime number $q = q(j)$ exists such that $q > f(x)$, $d(j) = d(q(j)) > x^{1/t^2}$ and q is minimal with this property. Moreover, this number q satisfies inequality (8.29) and the number of such numbers q is $o(x)$. Fix such a number q. Since $q|u_j$, this means that j is in the arithmetical progression $r(q) \pmod{d(q)}$. The number of such numbers $j < x$ is certainly at most $\frac{x}{d(q)} + 1$. So the total contributions when $d(q) > x$ are at most twice the number of such numbers q which as we have seen is $o(x)$. Thus, it remains to find an upper bound for

$$x \sum_{\substack{q-1 \in \mathcal{S} \\ x^{1/t^2} < d(q) < x}} \frac{1}{d(q)}.$$

In particular, Proposition 1.7 will be proved provided that we can show that

$$\sum_{\substack{q-1 \in \mathcal{S} \\ x^{1/t^2} < d(q) < x}} \frac{1}{d(q)} = o(1). \tag{8.30}$$

Set

$$\mathcal{D} = \{d \mid d = d(q) \text{ for some } q \text{ with } q - 1 \in \mathcal{S}\}.$$

In order to prove (8.30), we first need to understand an upper bound for the *multiplicity* of an element $d \in \mathcal{D}$. That is, given $d \in \mathcal{D}$ how many primes q with $q - 1 \in \mathcal{S}$ are there such that $d = d(q)$? Denote this number by $T(d)$. We shall later show that the inequality

$$T(d) \ll (3t)! d^{1-1/(t+1)} \tag{8.31}$$

holds for large enough values of x and uniformly in d. Assume, for the moment, that we have proved (8.31). Then we can bound the expression appearing in the left hand side of (8.30) by saying that

$$\sum_{\substack{q-1 \in \mathcal{S} \\ x^{1/t^2} < d(q) < x}} \frac{1}{d(q)} \ll (3t)! \sum_{\substack{d \in \mathcal{D} \\ x^{1/t^2} < d}} \frac{1}{d^{1/(t+1)}}.$$

Finally, let us notice that from the way $d = d(q)$ was defined we get that for every prime number q which divides u_j for some j the number q is a *primitive divisor* of $L_{d(q)}$. That is, $q|L_{d(q)}$ but q does not divide L_m for any positive integer $m < q$. From the well known properties of the primitive divisors we get that $d(q) \mid q - 1$. So, in particular, when $q - 1 \in \mathcal{S}$, we get that $d(q) \in \mathcal{S}$.

We thus have

$$\sum_{\substack{q-1\in\mathcal{S}\\ x^{1/t^2}<d(q)<x}} \frac{1}{d(q)} \ll \sum_{\substack{d\in\mathcal{S}\\ x^{1/t^2}<d}} \frac{1}{d^{1/(t+1)}}.$$

It remains to show that

$$\sum_{\substack{d\in\mathcal{S}\\ x^{1/t^2}<d}} \frac{1}{d^{1/(t+1)}} = o\left(\frac{1}{(3t)!}\right). \tag{8.32}$$

But obviously the series

$$\sum_{d\in\mathcal{S}} \frac{1}{d^{1/(t+1)}} \tag{8.33}$$

is convergent and the sum of the above series is precisely

$$h(t) = \prod_{i=1}^{t} \left(1 - \frac{1}{p_i^{1/(t+1)}}\right)^{-1}.$$

We now find an upper bound on $h(t)$. Notice that with fixed p we have

$$\begin{aligned}\left(1 - \frac{1}{p^{1/(t+1)}}\right)^{-1} &= \sum_{i\ge 0} \frac{1}{p^{i/(t+1)}} \\ &= \left(\sum_{i=0}^{t} \frac{1}{p^{i/(t+1)}}\right)\left(\sum_{j\ge 0} \frac{1}{p^j}\right) \le (t+1)\left(1 - \frac{1}{p}\right)^{-1},\end{aligned}$$

so that the inequality

$$\begin{aligned}h(t) &\le (t+1)^t \prod_{i=1}^{t} \left(1 - \frac{1}{p}\right) \\ &< \exp\left(t\log(t+1) + c_{54}\log_2 t\right) < \exp(2t\log t)\end{aligned} \tag{8.34}$$

holds for large enough values of x. To estimate the tail of the series (8.33) appearing in left hand side of formula (8.32) we let $d \in \mathcal{S}$ be such that $d > x^{1/t^2}$ and assume that α denotes the maximum of the exponents at which the prime numbers dividing d can appear in the prime factor factorization of d. Then obviously

$$t\log f(x)\alpha \ge \log d \ge \frac{\log x}{t^2},$$

so that the inequality

$$\alpha \ge \frac{\log x}{t^3 \log f(x)} \ge \frac{\log x}{2t^3 \log t} \tag{8.35}$$

holds for all large enough values of x.

By separating the prime power of maximal exponent α from d and then summing up over all the primes $p \in \mathcal{S}$ and over all the powers α which are at least as large as shown in (8.35) we get that the sum appearing in the left hand side of (8.32) is bounded above by

(8.36)
$$\begin{aligned}
&\sum_{i=1}^{t} \sum_{j \ge \frac{\log x}{2t^3 \log t}} \frac{1}{p^{j/(t+1)}} \sum_{d \in S} \frac{1}{d^{1/(t+1)}} \le h(t) \sum_{i=1}^{t} \sum_{j \ge \frac{\log x}{2t^3 \log t}} \frac{1}{p^{j/(t+1)}} \\
&\le h(t) \sum_{i=1}^{t} \exp\left(-\frac{\log x \log p_i}{2t^3(t+1)\log t}\right) \left(1 - \frac{1}{p_i^{1/(t+1)}}\right)^{-1} \\
&\ll h(t) t \sum_{i=1}^{t} \exp\left(-\frac{\log x \log p_i}{2t^3(t+1)\log t}\right) \le h(t) t^2 \exp\left(-\frac{\log x \log 2}{2t^3(t+1)\log t}\right) \\
&< \exp\left(2t \log t + 2 \log t - \frac{\log x}{2t^3(t+1)\log t}\right).
\end{aligned}$$

Since $1/(3t)^{3t} = o(1/(3t)!)$, it follows, with (8.36) that in order for (8.32) to hold it suffices that

$$\exp\left(2t \log t + 2 \log t - \frac{\log x}{2t^3(t+1)\log t}\right) < \frac{1}{(3t)^{3t}}$$

which is implied by

$$6t \log(3t) < \frac{\log x}{2t^3(t+1)\log t}$$

which is fulfilled provided that

$$12t^4(t+1)\log^2(3t) < \log x \tag{8.37}$$

and (8.37) obviously holds for large values of x because $t = \pi(f(x) < f(x)$ and $f(x)$ is given by formula (8.23).

Thus, Proposition 1.7 is proved once we are able to show that inequality (8.31) holds. To prove (8.31) assume that t is large, pick a number $d \in \mathcal{S}$, set $T = T(d)$ and write

$$b^d \gg L_d = \prod_{i=1}^{T} q_i$$

where $q_1 < q_2 < \cdots < q_T$ are distinct primes with $q_i - 1 \in \mathcal{S}$. Then certainly

$$b^d \gg \prod_{i=1}^{T} (q_i - 1). \tag{8.38}$$

To get a large T we have to assume that all the q_i's are as small as possible. But how small can we make the product on the left? Well, discarding the fact that q_i have to be primes we will certainly want to first put $q-1=p_i$ for $i=1,2,\ldots,t$, then for the next numbers we will want to put $q-1=p_ip_j$ for $1\le i\le j\le t$ and so on. The above argument shows that in order to get an upper bound on T we should write

$$T=\binom{t}{t-1}+\binom{t+1}{t-1}+\cdots+\binom{t+u}{t-1}+N$$

where u is the unique positive integer such that $0\le N<\binom{t+u+1}{t-1}$ and by (8.38) the maximal value of T will certainly be bounded above by those u and N for which the inequality

$$\binom{t}{1}+2\binom{t+1}{2}+\cdots+(u+1)\binom{t+u}{t-1}+(u+2)N\ll d \tag{8.39}$$

holds. Inequality (8.39) together with the obvious lower bound

$$\binom{t+i}{t-1}\ge\frac{(i+1)^{t-1}}{(t-1)!}$$

uniformly in i, shows that

$$\sum_{i=1}^{u+1}i^t\le(t-1)!d$$

and since

$$\sum_{i=1}^{u+1}i^t\gg\frac{u^{t+1}}{t+1}$$

holds uniformly in i and u, we get

$$u\ll(t+1)!^{1/t+1}d^{1/(t+1)}.$$

In particular, using now the fact that the inequality

$$\binom{t+i}{t-1}\le(i+1)^{t-1}$$

holds uniformly in i, we get

$$\begin{aligned}T&\le\sum_{i=0}^{u+1}\binom{t+i}{t-1}\le\sum_{i=1}^{u+2}i^{t-1}\le t(u+2)^{t-1}\\&\ll t(t+1)!^{(t-1)/(t+1)}d^{t/(t+1)}\left(1+\frac{2}{u}\right)^{t-1}<(3t)!d^{1-\frac{1}{(t+1)}},\end{aligned} \tag{8.40}$$

where in the last step of (8.40) we used the fact that

$$t((t+1)!)^{(t-1)/(t+1)}\left(1+\frac{2}{u}\right)^{t-1} \le t((t+1)!)^{(t-1)/(t+1)}3^{t-1} = o((3t)!)$$

which holds for large t (by Stirling's formula, for example).

Proposition 1.7 is therefore proved. ■

References

[1] BAKER, A. AND WÜSTHOLZ, G.: Logarithmic forms and group varieties. *J. Reine Angew. Math.* **442** (1993), 19–62.

[2] BALOG, A. AND WOOLEY, T.: On strings of consecutive integers with no large prime factors. *J. Austral Math. Soc. Ser. A* **64** (1998), 266–276.

[3] BANKS, W. AND SHPARLINSKI, I. E.: Arithmetic properties of numbers with restricted digits. *Acta Arith.* **112** (2004), 313–332.

[4] BUGEAUD, Y.: Lower bounds for the greatest prime factor of $ax^m + by^n$. *Acta. Math. Inform. Univ. Ostraviensis* **6** (1998), 53–57.

[5] CARMICHAEL, R. D.: On the numerical factors of the arithmetic forms $\alpha^n \pm \beta^n$. *Ann. of Math. (2)* **15** (1913), 30–70.

[6] CORVAJA, P. AND ZANNIER, U.: Diophantine equations with power sums and universal Hilbert sets. *Indag. Math. (N.S.)* **9** (1998), 317–332.

[7] DARTYGE, C. AND MAUDUIT, C.: Nombres presque premiers dont l'écriture en base r ne comporte pas certaines chiffres. *J. Number Theory* **81** (2000), 270–291.

[8] DE BRUIJN, N. G.: On the number of positive integers $\le x$ and free of prime factors $> y$. *Nederl. Akad. Wetensch. Proc. Ser. A* **54** (1951), 50–60.

[9] EGGLETON, R. B. AND SELFRIDGE, J. L.: Consecutive integers with no large prime factors. *J. Austral. Math. Soc. Ser. A* **22** (1976), 1–11.

[10] FOUVRY, E. AND MAUDUIT, C.: Methódes de crible et fonctions sommes des chiffres. *Acta Arith.* **77** (1996), 339–351.

[11] FOUVRY, E. AND MAUDUIT, C.: Sommes des chiffres et nombres presque premiers. *Math. Ann.* **305** (1996), 571–599.

[12] KONYAGIN, S., MAUDUIT, C. AND SÁRKÖZY, A.: On the number of prime factors of integers characterized by digit properties. *Period. Math. Hungar.* **40** (2000), 37–52.

[13] LUCA, F.: The number of non zero digits of $n!$. *Canad. Math. Bull.* **45** (2002), 115–118.

[14] LUCA, F.: How smooth is $\phi(2^n+3)$? *Rocky Mountain J. Math.* **34** (2004), 1367–1389.

[15] Mauduit, C. and Sárközy, A.: On the arithmetic structure of sets characterized by sum of digits properties. *J. Number Theory* **61** (1996), 25–38.

[16] Mauduit, C. and Sárközy, A.: On the arithmetic structure of the integers whose sum of digits is fixed. *Acta Arith.* **81** (1997), 145–173.

[17] Nicolas, J.-L.: Petites valeurs de la fonction d'Euler. *J. Number Theory* **17** (1983), 375–388.

[18] Rosser, J. B. and Schoenfeld, L.: Approximate formulas for some functions of prime numbers. *Illinois J. Math.* **6** (1962) 64–94.

[19] Schlickewei, H. P.: S-unit equations over number fields. *Invent. Math.* **102** (1990), no. 1, 95–107.

[20] Schmidt, W. M.: *Diophantine approximations and Diophantine equations*. Lecture Notes in Mathematics **1467**. Springer-Verlag, Berlin, 1991.

[21] Senge, H. G. and Strauss, E. G.: PV-numbers and sets of multiplicity. *Period. Math. Hungar.* **3** (1973), 93–100.

[22] Shparlinski, I. E.: Prime divisors of sparse integers. *Period. Math. Hungar.* **46** (2003), no. 2, 215–222.

[23] Stewart, C. L.: On the representation of an integer in two different bases. *J. Reine Angew. Math.* **319** (1980), 63–72.

[24] Szalay, L.: The equations $2^n \pm 2^m \pm 2^l = z^2$. *Indag. Math. (N.S.)* **13** (2002), 131–142.

[25] Yu, K.: p-adic logarithmic forms and group varieties. II. *Acta Arith.* **89** (1999), no. 4, 337–378.

Recibido: 1 de octubre de 2003

Florian Luca
Instituto de Matemáticas
Universidad Nacional Autónoma de México
C.P. 58089, Morelia, Michoacán, México
fluca@matmor.unam.mx

This work was partly supported by Grant SEP-CONACYT 37259-E.