The Journal of Symbolic Logic
Volume 00, Number 0, XXX 0000

ON THE ADDITIVE THEORY OF PRIME NUMBERS

PATRICK CÉGIELSKI, DENIS RICHARD, AND MAXIM VSEMIRNOV

Abstract. The undecidability of the additive theory of prime numbers (with identity) as well as the theory $\mathrm{Th}(\mathbb{N},\ +,\ n \mapsto p_n)$, where p_n denotes the $(n+1)$-th prime, are open questions. In a first part, we show the undecidability of $\mathrm{Th}(\mathbb{N},\ +,\ n \mapsto nf(n))$ where f is a good approximation of the enumeration $n \mapsto p_n/n$. In a second part, as a possible approach, we extend the former theory by adding some extra function. In this direction we show the undecidability of the existential part of the theory $\mathrm{Th}(\mathbb{N},\ +,\ n \mapsto p_n, n \mapsto r_n)$, where r_n is the remainder of p_n divided by n in the euclidian division.

L'indécidabilité de la théorie additive des nombres premiers ainsi que de la théorie $\mathrm{Th}(\mathbb{N},\ +,\ n \mapsto p_n)$, où p_n désigne le $(n+1)$-ième premier, sont deux questions ouvertes. Dans une première partie, nous montrons l'indécidabilité de $\mathrm{Th}(\mathbb{N},\ +,\ n \mapsto nf(n))$ où f est une bonne approximation de la fonction $n \mapsto p_n/n$ des nombres premiers. Dans une seconde partie, nous étendons la première théorie en lui ajoutant une fonction supplémentaire et nous montrons l'indécidabilité de la théorie $\mathrm{Th}(\mathbb{N},\ +,\ n \mapsto p_n, n \mapsto r_n)$, où r_n désigne le reste de p_n dans la division euclidienne de p_n par n, et même de sa seule partie existentielle.

Introduction. The questions of decidability and of arithmetical definability raised by the *multiplicative theories of prime numbers* such as $\mathrm{Th}(\mathbb{N},\ \bullet,\ \mathbb{P})$ and $\mathrm{Th}(\mathbb{N},\ |,\ \mathbb{P})$, where $\mathbb{P}$ denotes the set of prime numbers, was solved in 1930 by Skolem since P is $(\bullet)$-definable and $(|)$-definable. Moreover, the theories $\mathrm{Th}(\mathbb{N},\ \bullet,\ n \mapsto p_n)$ and $\mathrm{Th}(\mathbb{N},\ |,\ n \mapsto p_n)$, where p_n denotes the n-th prime number ($p_0 = 2$, $p_1 = 3$, etc.), have been shown to be undecidable in [CMR]. On the other hand the *additive theory of prime numbers* is a well-developed theory which enables us to express such well known problems as Goldbach's conjecture or Polignac's conjecture ([RIB], p.250) (the infinity of twin primes is a special case here), or many other classical questions or results like the Schnirelman Theorem. From the logical point of view, the additive theory of prime numbers consists in investigating the first-order structure $\langle \mathbb{N},\ +,\ \mathbb{P} \rangle$. Some authors provided [BJW, BOF, LM] conditional proofs (through the linear case in Schinzel's Hypothesis [SS]) of the undecidability of the additive theory of primes $\mathrm{Th}(\mathbb{N},\ +,\ \mathbb{P})$, where $\mathbb{P}$ is the set of all primes.

In order to further the study of this theory without assuming any conjecture such as Schinzel's Hypothesis, we solve the decision problems for some first-order theories close in a sense to $\mathrm{Th}(\mathbb{N},\ +,\ \mathbb{P})$. Instead of studying $\langle \mathbb{N},\ +,\ \mathbb{P} \rangle$

0022-4812/00/0000-0000/$00.00

itself, our investigation concerns $\mathrm{Th}(\mathbb{N},\ +,\ n \mapsto p_n)$. This leads us immediately to the following.

Open problem 1: *Is the natural enumeration of prime numbers* $\mathrm{n} \mapsto \mathrm{p_n}$ *definable in the structure* $\langle \mathbb{N},\ +,\ \mathbb{P} \rangle$?

According to the prime numbers theorem (Hadamard-de la Vallée Poussin, 1896) we know $p_n \sim n\log(n)$, so that we begin by considering $\mathrm{Th}(\mathbb{N},\ +,\ n \mapsto n\lfloor \log(n) \rfloor)$ and $\mathrm{Th}(\mathbb{N},\ +,\ n \mapsto n\lfloor \log_2(n) \rfloor)$, where $\log(n)$ and $\log_2(x)$ respectively denote neperian and binary logarithms of x, putting $0.\log(0) = 0$ and $0.\log_2(0) = 0$.

The first approach is approximation. Another approach consists of extending the language $\{+,\ n \mapsto p_n\}$ to $\{+,\ n \mapsto p_n,\ n \mapsto r_n\}$, where r_n is the remainder of p_n divided by n. The main result of the second part is the following:

THEOREM 0.1. *Multiplication is existentially* $\langle \mathbb{N},\ +,\ \mathrm{n} \mapsto \mathrm{p_n},\ \mathrm{n} \mapsto \mathrm{r_n} \rangle$*-definable at first-order.*

This leads to the following (without use of conjectures) result:

THEOREM 0.2. $\mathrm{Th}_\exists(\mathbb{N},\ +,\ \mathrm{n} \mapsto \mathrm{p_n},\ \mathrm{n} \mapsto \mathrm{r_n})$ *is undecidable.*

The general framework of definability can be found, for instance, in [END] or is presented in a more detailed way in the survey carried out by the first author [CEG]. For a structure M, we denote by $\mathrm{DEF}(M)$ the set of constants, functions and relations which are first-order definable within M. Since $\mathrm{Th}(\mathbb{N},\ +,\ \bullet)$ is undecidable (TURING, 1936), a method for proving the undecidability of the theory of a structure M consists in showing $\mathrm{DEF}(M) = \mathrm{DEF}(\mathbb{N},\ +,\ \bullet)$. The set $\mathrm{DEF}(\mathbb{N},\ +,\ \bullet)$ is well-known and the inclusion $\mathrm{DEF}(M) \subset \mathrm{DEF}(\mathbb{N},\ +,\ \bullet)$ is very often trivial, which is the case in the present paper. In fact, the only problem is to know whether the converse inclusion holds.

§1. Approximation theories.

PROPOSITION 1.1. *The equality*

$$\mathrm{DEF}(\mathbb{N},\ +,\ \mathrm{n} \mapsto \mathrm{n}.\ \lfloor \log_2(\mathrm{n}) \rfloor) = \mathrm{DEF}(\mathbb{N},\ +,\ \bullet)$$

holds hence $\mathrm{Th}(\mathbb{N},\ +,\ \mathrm{n} \mapsto \mathrm{n}.\lfloor \log_2(\mathrm{n}) \rfloor)$ *is undecidable.*

This result is a corollary of Proposition 1.3 below which we shall prove after the introduction of the mappings used.

DEFINITION 1.2. For any positive integer k, we define f_k and $\exp_k$ as mappings respectively from a final segment of $\mathbb{N}$ into $\mathbb{N}$ which are such that

$$f_k(n) = n.\lfloor \log_2(\log_2(\ldots \log_2(n) \ldots)) \rfloor$$

$$\exp_k(n) = 2^{2^{\cdot^{\cdot^{\cdot^{2^n}}}}}$$

with k occurrences of $\log_2$ for f_k and k occurrences of 2 for $\exp_k(n)$. The domain of f_k will be $\{n \in \mathbb{N} \ / \ n \geq n_k\}$, where n_k is the smallest integer n satisfying $f_k(n) \geq 0$.

PROPOSITION 1.3. *The equality* $\mathrm{DEF}(\mathbb{N},\ +,\ f_k) = \mathrm{DEF}(\mathbb{N},\ +,\ \bullet)$ *holds, hence* $Th(\mathbb{N},\ +,\ f_k)$ *is undecidable for any integer* k.

PROOF. For any sufficiently large integer x there exists an integer n such that $\exp_k(n) \leq x < \exp_k(n+1)$, then we have $f_k(x) = x.n$,

$$f_k(x-1) = \begin{cases} (x-1)(n-1) & \text{if } x = \exp_k(n), \\ (x-1)n & \text{otherwise,} \end{cases}$$

and consequently

$$f_k(x) - f_k(x-1) = \begin{cases} n + x - 1 > x & \text{if } x = \exp_k(n), \\ n < x & \text{otherwise.} \end{cases}$$

Therefore the set $A_k = \{\exp_k(n) \ / \ n \in \mathbb{N}\}$ is definable within $\langle \mathbb{N},\ +,\ f_k \rangle$ since $x \in A_k$ if and only if $[f_k(x) - f_k(x-1) > x]$.

The mapping $\exp_k$ is consequently definable within $\langle \mathbb{N},\ +,\ f_k \rangle$ since $y = \exp_k(x)$ is equivalent to

$$[y \in A_k \wedge f_k(y+1) - f_k(y) = x].$$

The function $n \mapsto n^2$ from $\mathrm{Dom}(f_k)$ into $\mathbb{N}$ is definable within $\langle \mathbb{N},\ +,\ f_k \rangle$ since we have

$$f_k(\exp_k(n) + n) = n.(\exp_k(n) + n) = f_k(\exp_k(n)) + n^2.$$

The classical result of PUTNAM insuring

$$\mathrm{DEF}(\mathbb{N},\ +,\ n \mapsto n^2) = \mathrm{DEF}(\mathbb{N},\ +,\ \bullet)$$

allows us to conclude. ⊣

Proposition 1.3 is in a sense an improvement of the result of BATEMAN-JOCKUSCH-WOODS [BJW] expressing the fact $\mathrm{DEF}(\mathbb{N},\ +,\ f) = \mathrm{DEF}(\mathbb{N},\ +,\ \bullet)$ for some polynomial functions f with a degree not smaller than 2. Here we have a result for functions which increase but remain as close as we want to the zero function. This proof itself suggests a more general proposition which is our fundamental lemma and permits us to generalize Proposition 1.3 to other functions approximating more precisely than $n \mapsto n \log n$ the natural enumeration of primes. For this purpose, we introduce a class of real functions containing the usual approximations of $n \mapsto p_n/n$.

DEFINITION 1.4. The class C is the set of all (invertible) real functions $f : [a_0,\ +\infty[\to \mathbb{R}$ satisfying the following conditions:

1. f is continuous;

2. f is strictly increasing;
3. $\lim_{x\to+\infty} f(x) = +\infty$;
4. for every x which is positive real, $f(x) < x$;
5. There exists $x_0 \in (\mathbb{R}^*)^+$ such that for all reals $x \geq x_0$ the inequality $\mathrm{f(x+1)} < \mathrm{f(x)} + 1/2$ holds.

EXAMPLE 1.5. The functions $x \mapsto \log_2(x)$, $x \mapsto \log_2(x) + \log_2(\log_2 x) - 1$, $x \mapsto \log(x)$, and $x \mapsto \log(x) + \log(\log x) - 1$ belong to C.

Conditions 1), 2), 3), and 4) are obviously verified and for 5), we have, for $\log_2$ with $x_0 = 2/\log(2)$ and $0 < \theta < 1$, using the TAYLOR formula:

$$\log_2(x+1) - \log_2(x) = \frac{\log(x+1) - \log(x)}{\log 2} < \frac{1}{(x+\theta)\log 2} < \frac{1}{2}.$$

Similar arguments work for the three other examples.

The main result of this section is a straightforward corollary of the following

PROPOSITION 1.6. *(Fundamental Lemma) For any function f of the class C, we have* $\mathrm{DEF}(\mathbb{N},\ +,\ n \mapsto n\lfloor f(n)\rfloor) = \mathrm{DEF}(\mathbb{N},\ +,\ \bullet)$*, hence* $\mathrm{Th}(\mathbb{N},\ +,\ n \mapsto n\lfloor f(n)\rfloor)$ *is undecidable.*

Some notations. By definition $\tilde{f} = \lfloor f \rfloor$ is the function from $\mathbb{N}$ into $\mathbb{N}$ which associates to n the greatest integer $\lfloor f(n)\rfloor$ smaller than $f(n)$.

Let $\widetilde{f^{-1}}$ be the function from $\mathbb{N}$ into $\mathbb{N}$ determined by $\widetilde{f^{-1}}(n) = \mu\ m(f(m+1)) > n$, where μ means as usual "the smallest ... such that ...".

Let $\widehat{f^{-1}}$ be the function from $\mathbb{N}$ into $\mathbb{N}$ which associates to n the smallest integer $\lceil f^{-1}(x)\rceil$ greater than $f^{-1}(x)$.

Below, we list some useful facts about the previous mappings.

FACT 1.7. The range of $\tilde{f}$ contains $[\tilde{f}(0),\ +\infty[\cap \mathbb{N}$.

PROOF. Let p_0 be the greatest integer p such that $\lfloor f(p)\rfloor = n_0$. Then $\lfloor f(p_0+1)\rfloor \geq n_0+1$. But, following Condition 5) of Definition 1.4, $f(p_0+1) < f(p_0)+\frac{1}{2}$ so that $\lfloor f(p_0+1)\rfloor \leq \lfloor f(p_0)+\frac{1}{2}\rfloor \leq n_0+1$ and $\tilde{f}(p_0+1) = n_0+1$. ⊣

From Fact 1.7, we can deduce the existence of a function $h : \mathbb{N} \to \mathbb{N}$ such that $h(n)$ is the greatest integer q satisfying $\tilde{f}(q) = n$. It is easy to prove h is increasing.

We leave to the reader the proof that $h(n+2) - h(n) > 2n+1$ is true for each h determined by a functions f we gave as examples above. Consequently we assume from now onwards:

Condition 6. For every non negative integer n, the inequality $h(n+2) - h(n) > 2n+1$ holds.

FACT 1.8. For every positive integer, $\tilde{f}(n+1) \leq \tilde{f}(n)+1$.

PROOF. It trivially follows from Condition 5) of Definition 1.4. ⊣

FACT 1.9. For every $f \in C$ and for all positive reals we have $f^{-1}(x+1) - f^{-1}(x) \geq 1$.

PROOF. Indeed, for $y = f^{-1}(x)$, we have $f(y+1) - f(y) < 1/2 < 1$ which implies $f(f^{-1}(x)+1) < x+1$. Then $f^{-1}(f(f^{-1}(x)+1)) < f^{-1}(x+1)$ and $f^{-1}(x)+1 < f^{-1}(x+1)$. ⊣

FACT 1.10. We have $\tilde{f}(h(n)) = n$ and $h(\tilde{f}(n)) \geq n$ for every integer $n \geq \tilde{f}(0)$.

PROOF. The first equality comes from Fact 1.6 which implies $\{q \geq \tilde{f}(0)$ such that $\tilde{f}(q) = n\} \neq \emptyset$. By definition every member of the former set satisfies $\tilde{f}(h(n)) = n$. Since $h(\tilde{f}(n))$ is the greatest integer q such that $\tilde{f}(q) = \tilde{f}(n)$, we have $q \geq n$. ⊣

PROOF. (Proposition 1.6)

We begin by showing that the set $h(\mathbb{N})$ belongs to DEF$(\mathbb{N}, +, n \mapsto n\tilde{f}(n))$. Indeed, if $q_0 \in h(\mathbb{N})$, there exists $p \in \mathbb{N}$ such that $\tilde{f}(q_0) = p$ and $\tilde{f}(q_0+1) = p+1$ as a consequence of Fact 1.7. Let us put $g(n) = n\tilde{f}(n)$, so that $g(q_0) = q_0 p$ and $g(q_0+1) = (q_0+1)(p+1)$. Therefore $g(q_0+1) - g(p_0) = q_0 + p + 1 > q_0$.

Conversely if $q_0 \notin h(\mathbb{N})$, then p satisfies $\tilde{f}(q_0) = p$, and also $\tilde{f}(q_0+1) = p$, implying $g(q_0+1) - g(q_0) = (q_0+1)p - q_0 p = p < q_0$ (by Definition 1.4(4)). Finally $q \in h(\mathbb{N})$ if and only if $g(q+1) - g(q) > q$, a condition which is obviously definable in the structure $\langle \mathbb{N}, +, n \mapsto n\tilde{f}(n) \rangle$.

Now, the function h itself is definable in the former structure through the logical equivalence between $h(p) = q$ and $p + q + 1 = g(p+1) - g(p)$.

The next step in the proof of Proposition 1.6 is to show the $\langle \mathbb{N}, +, g \rangle$-definability of $\tilde{f}$. For this purpose, firstly we intend to prove that for every $n \geq \tilde{f}(0)$, we have the inequality $\tilde{f}^{-1}(n-1) \leq h(n) < \widehat{f^{-1}}(n+1)$. Since f is increasing we have $f(\lceil f^{-1}(n+1) \rceil) \geq f(f^{-1}(n+1)) = n+1$, therefore $\tilde{f}(\widehat{f^{-1}}(n+1)) = \lfloor f(\lceil f^{-1}(n+1) \rceil) \rfloor \geq n+1$, since $n+1$ is an integer. Now, assume by *reductio ad absurbum* the inequality $h(n) \geq \widehat{f^{-1}}(n+1)$. We should get from this hypothesis the inequality $\tilde{f}(h(n)) \geq \tilde{f}(\widehat{f^{-1}}(n+1)) \geq n+1$ (since $\tilde{f}$ is increasing), which contradicts the equality $\tilde{f}(h(n)) = n$ (Fact 1.10) and thereby proves the inequality $h(n) < \widehat{f^{-1}}(n+1)$.

By definition of the ceil-function, we have $\widehat{f^{-1}}(n-1) = \lceil f^{-1}(n-1) \rceil < f^{-1}(n-1)+1$. But we know (Fact 1.9) that $f^{-1}(n-1)+1 \leq f^{-1}(n)$, providing

$\widehat{f^{-1}}(n-1) \leq f^{-1}(n)$. Since f and f^{-1} are (strictly) increasing and since, for $n \geq \tilde{f}(0)$, we have $\tilde{f}(h(n)) = n$, we get $f^{-1}(n) = f^{-1}(\tilde{f}(h(n))) \leq f^{-1}(f(h(n))) = h(n)$, proving the desired inequality $\widehat{f^{-1}}(n-1) \leq h(n)$.

The $(\mathbb{N},\ +,\ g)$-definability of $x \mapsto x^2$ uses Condition 6 we have imposed on h. We can distinguish two cases.

First case. Suppose $n \in \mathbb{N}$ satisfying $h(n+1) - h(n) > n$. In this case, $\tilde{f}(h(n)+n) = n+1$ since, by definition of h, the integer $h(n)$ is the greatest q such that $\tilde{f}(q) = n$, we have on the one hand $\tilde{f}(h(n)+n) \geq n+1$. On the other hand, $h(n)+n \leq h(n+1)$ implies $\tilde{f}(h(n)+n) \leq \tilde{f}(h(n+1)) = n+1$ by definition of $h(n+1)$. It follows $g(h(n)+n) = (h(n)+n)\tilde{f}(h(n)+n) = (h(n)+n)(n+1) = h(n).n + n^2 + h(n) + n$. From $n = \tilde{f}(h(n))$, we get $g(h(n)+n) = \tilde{f}(h(n))h(n) + n^2 + h(n) + n = g(h(n)) + n^2 + h(n) + n$ and finally, $n^2 = g(h(n)) - h(n) - n$ which leads, in the first case, to an easy $\langle \mathbb{N},\ +,\ g\rangle$-definition of the square.

Second case. Suppose n does not satisfy $h(n+1) - h(n) > n$ and therefore verifies $h(n+1) - h(n) \leq n$. Using the former inequality and Condition 6, the inequality $h(n+2) - h(n+1) > n+1$ holds, and coming back to Case 1, we can $\langle \mathbb{N},\ +,\ g\rangle$-define $(n+1)^2$.

Consequently, both cases lead to a $\langle \mathbb{N},\ +,\ g\rangle$-definition of the relation $m = n^2$ which is logically equivalent to

$$\{[(h(n+1) - h(n) > n) \wedge g(h(n)+n) = g(h(n)) + h(n) + n + m] \vee (h(n+1) - h(n) \leq n) \wedge [g(h(n+1) + n + 1) = g(h(n+1)) + h(n+1) + n + 1 + m + n + n + 1]\}.$$

⊣

Proposition 1.6 has applications concerning the additive theory of primes. Indeed the best known approximation of $n \mapsto p_n$ is the following ([RIB], p.249):

$$p_n = n.\log(n) + n.(\log(\log(n)) - 1) + O\left(\frac{n.\log(\log(n))}{\log(n)}\right).$$

This approximation does not seem to be sufficient to prove the undecidability of $Th(\mathbb{N},\ +,\ n \mapsto p_n)$. However we have:

PROPOSITION 1.11. *The equality* $\mathrm{DEF}(\mathbb{N},\ +,\ \mathrm{n} \mapsto \mathrm{n}.\lfloor \log(\mathrm{n}) + \log(\log(\mathrm{n})) - 1 \rfloor) = \mathrm{DEF}(\mathbb{N},\ +,\ \bullet)$ *holds so that* $\mathrm{Th}(\mathbb{N},\ +,\ \mathrm{n}.\lfloor \log(\mathrm{n}) + \log(\log(\mathrm{n})) - 1) \rfloor)$ *is undecidable.*

COROLLARY 1.12. *For any of the restrictions to* $\mathbb{R}^{*+}$ *of the real functions* $\mathrm{f} \in \{\log,\ \log_2,\ \log + \log(\log) - 1,\ \log_2 + \log_2(\log_2) - 1\}$*, we have* $\mathrm{DEF}(\mathbb{N},\ +,\ \mathrm{n} \mapsto \mathrm{n}\lfloor \mathrm{f}(\mathrm{n}) \rfloor) = \mathrm{DEF}(\mathbb{N},\ +,\ \bullet)$ *and the theory* $\mathrm{Th}(\mathbb{N},\ +,\ \mathrm{n} \mapsto \mathrm{n}\lfloor \mathrm{f}(\mathrm{n}) \rfloor)$ *is undecidable.*

PROOF. We are going to explicit the proof for $f = \log + \log(\log) - 1$ and to show this function belongs to the class C so that we can apply our fundamental Lemma. It is clear f satisfies Conditions 1), 2), 3), 4), and 5) of Definition 1.4. We check Condition 6. Since $q_0 = h(n) = \text{Max}\{q \in \mathbb{N} \text{ such that } \exists \alpha_q (0 \leq \alpha_q < 1)$ and $f(q) = n + \alpha_q\}$, we also get $h(n) = \text{Max}\{q \in \mathbb{N} \text{ such that } \exists \alpha_q (0 \leq \alpha_q < 1)$ and $\log q + \log(\log q) - 1 = n + \alpha_q\} = \text{Max}\{q \in \mathbb{N} \text{ such that } \exists \alpha_q (0 \leq \alpha_q < 1)$ and $e^{n+\alpha_q+1} = e^{\log q} e^{\log(\log q)} = q \log q\}$. Similarly $q_2 = h(n+2)$ is equal to $\text{Max}\{q' \in \mathbb{N}$ such that $\exists \alpha_{q'} (0 \leq \alpha_{q'} < 1)$ and $e^{n+\alpha_{q'}+3} = q' \log q'\}$. Therefore

$$h(n+2) - h(n) = q_2 - q_0 = \frac{e^{n+3+\alpha_{q_2}}}{\log q_2} - \frac{e^{n+1+\alpha_{q_0}}}{\log q_0}$$

a lower bound of which is

$$\frac{e^{n+1}}{\log q_2}(e^{2+\alpha_{q_2}} - e^{\alpha_{q_0}}) \geq \frac{e^{n+1}(e^2 - e)}{\log q_2}.$$

We know $\log q_2 + \log(\log q_2) - 1 = n + 3 + \alpha_{q_2}$, hence $\log q_2 \leq n + 5$ and $h(n+2) - h(n) \geq \frac{e^{n+1}}{n+5}(e^2 - e) > 2n + 1$, for all nonnegative integers. ⊣

Remark.- Inspecting the definition of multiplication in the proof of Proposition 1.11, we see the obtained formula of definition is existential (only order appears). Consequently the Π_2-theory of $\langle \mathbb{N}, +, n \mapsto n\tilde{f}(n) \rangle$ is undecidable.

§2. Extended theories.

Introduction.- Actually all positive integer constants are existentially $\{+, \mathbb{P}\}$-definable in the following manner:

$$\begin{array}{lcl} x = 0 & \Leftrightarrow & x + x = x; \\ x = 1 & \Leftrightarrow & \exists y (y = x + x \wedge y \in \mathbb{P}); \\ x = 2 & \Leftrightarrow & \exists y (y = 1 \wedge x = y + y); \\ & \vdots & \\ x = n + 1 & \Leftrightarrow & \exists y \exists z (y = n \wedge z = 1 \wedge x = y + z). \end{array}$$

Moreover $\mathbb{P}$ is existentially definable within the language $\{+, n \mapsto p_n\}$, hence all positive integer constants are also existentially $\{+, n \mapsto p_n\}$-definable. Note, that $n \left\lfloor \frac{p_n}{n} \right\rfloor = p_n - r_n$. We intend to define $\left\lfloor \frac{p_n}{n} \right\rfloor$ from $+$ and $n \left\lfloor \frac{p_n}{n} \right\rfloor$. Then the strategy will be to define multiplication through the function $n \mapsto cn^2$ (where c is a fixed constant), which is to be proved $\{+, \left\lfloor \frac{p_n}{n} \right\rfloor, n \left\lfloor \frac{p_n}{n} \right\rfloor\}$-definable. Consequently, multiplication will be existentially $\{+, n \mapsto p_n, \quad n \mapsto r_n\}$-definable at first-order.

In the first part we have considered continuous real strictly increasing functions and their inverses. Since we work with integer parts we have to introduce pseudo-inverses of discrete functions. For such a discrete unbounded function f from $\mathbb{N}$ into $\mathbb{N}$, we define its pseudo-inverse f^{-1} from $\mathbb{N}$ into $\mathbb{N}$ by $f^{-1}(n) = \mu m[f(m + 1) > n]$, where μ means "the smallest ... such that". Due to the unboundness of f such an f^{-1} is correctly defined.

2.1. Some preliminary results in Number Theory.

Contrarily to what happens with log, the behavior of $\left\lfloor \frac{p_n}{n} \right\rfloor$ is *a priori* irregular but we shall prove it is not too much chaotic. Namely, we prove:

PROPOSITION 2.1. *Let us denote the mapping* $n \mapsto \left\lfloor \frac{p_n}{n} \right\rfloor$ *by* f.

1. *For* $m > n$, *we have* $f(m) - f(n) \geq -1$;
2. *For* $n \geq 11$, *we have* $f^{-1}(n+1) - f^{-1}(n) > n$.

PROOF. 1) We use the following estimates for p_n ([RIB], p. 249):

$$p_m \geq m \log m + m \log\log m - 1.0072629m \quad \text{for } m \geq 2;$$
$$p_m \leq m \log m + m \log\log m - 0.9385m \quad \text{for } m \geq 7022. \tag{1}$$

For $m > n \geq 7022$, we have $f(m) - f(n) = \left\lfloor \frac{p_m}{m} \right\rfloor - \left\lfloor \frac{p_n}{n} \right\rfloor$
$\geq \frac{p_m}{m} - \frac{p_n}{n} - 1 \geq log(\frac{m}{n}) - log(\frac{logm}{logn}) - 0.9385 + 1.0072629 - 1.$
Hence $f(m) - f(n) \geq -1$ because the sum of the two first terms is positive as is the sum of terms three and four.

If $n < 7022$, one may check the desired inequality by a direct computation.

2) Let m be $f^{-1}(n)$. By the very definition of f^{-1}, the equality $m = f^{-1}(n)$ is equivalent to the conjunction of the two following conditions:

$$\begin{cases} \left\lfloor \dfrac{p_{m+1}}{m+1} \right\rfloor \geq n+1; \\ \forall k \leq m \ \left\lfloor \dfrac{p_k}{k} \right\rfloor \leq n. \end{cases} \tag{2}$$

For $k \leq 7022$, the maximum of $\frac{p_k}{k}$ is attained for $k = 7012$ and equal to $\frac{p_{7012}}{7012} < 10.102824 < 11$. Consequently, we see that $m = f^{-1}(n) \geq f^{-1}(11) \geq 7022$ and this is the reason why in the hypothesis of Proposition 2.1, item 2) we assume $n \geq 11$.

To prove the inequality, it is sufficient to prove that for $k = m + n$ we have $\left\lfloor \frac{p_k}{k} \right\rfloor \leq n + 1$, or in other words,

$$\frac{p_k}{k} < n + 2. \tag{3}$$

Note that for $m \geq 7022$, we have by (2):

$$n + 1 \leq \left\lfloor \frac{p_{m+1}}{m+1} \right\rfloor + 1 \leq \frac{p_{m+1}}{m+1} + 1 \leq \log(m+1) + \log\log(m+1) - 0.07 < m.$$

Consequently it is sufficient – and actually more convenient – to prove a somehow stronger result, namely the same inequality (3) but for $m \geq 7022$ and $m + 1 \leq k \leq 2m$.

From the second estimate of (1) we have, since $k \geq m \geq 7022$, the following inequalities:

$$\begin{aligned} \tfrac{p_k}{k} &< \log k + \log\log k - 0.9385 \\ &\leq \log 2m + \log\log 2m - 0.9385 \\ &= \log m + \log\log m + \log 2 + \log(1 + \tfrac{\log 2}{\log m}) - 0.9385; \end{aligned}$$

using the first estimate of (1) and $\frac{\log 2}{\log m} \leq \frac{\log 2}{\log 7022}$, we have:

$$\log m + \log\log m - 1.0072629 \leq \frac{p_m}{m};$$

consequently:

$$\frac{p_k}{k} \leq \frac{p_m}{m} + 0.07 + \log 2 + \log(1 + \frac{\log 2}{\log 7022}) \leq \frac{p_m}{m} + 1$$

by an easy computation and finally, due to (2), we obtain $\frac{p_k}{k} < n + 2$. ⊣

Item 1) of previous proposition emphasizes on the fact that $f : n \mapsto \lfloor \frac{p_n}{n} \rfloor$ is "almost" increasing and item 2) shows that the difference $f^{-1}(n+1) - f^{-1}(n)$ is big enough with respect to n. This suggests to introduce a new class of functions, containing f, for which we prove that the existential part of the theory $\mathrm{Th}(\mathbb{N}, +, n \mapsto nf(n))$ is undecidable.

2.2. The class $C(k, d, n_0)$. Let $k \geq 0$ be a fixed nonnegative integer. We shall say f is *k-almost increasing* if and only if

$$\forall y \geq x[f(y) - f(x) \geq -k]. \tag{4}$$

In this sense 0-almost increasing means increasing (not necessarily strictly) and $n \mapsto \lfloor \frac{p_n}{n} \rfloor$ is 1-almost increasing (due to Proposition 2.1).

Still looking at $n \mapsto \lfloor \frac{p_n}{n} \rfloor$, we intend to consider functions whose pseudo-inverse is defined and asymptotically increases quickly enough with respect to its argument. Let us say that f^{-1} has at least $(1/d)$-linear difference, if

$$\exists n_0 \in \mathbb{N} \forall n \geq n_0[f^{-1}(n+1) - f^{-1}(n) > \frac{n}{d}]. \tag{5}$$

In fact, for $\lfloor \frac{p_n}{n} \rfloor$, the constant d is 1 and $n_0 = 11$, but results and proofs hold for an arbitrary (fixed) d.

Now let us definite the class $C(k, d, n_0)$ as the set of functions from $\mathbb{N}$ into $\mathbb{N}$ satisfying conditions (4) of being *k-almost increasing* and (5) of having its pseudo-inverse with an *at least $(1/d)$-linear difference*.

In order to prove the above FUNDAMENTAL LEMMA, whose Theorem 0.1 is a corollary, we show some properties of the class $C(k, d, n_0)$. Firstly, in section 2.3 we present in three lemmas these properties and comment them. Afterwards, in section 2.4, we prove them.

2.3. Properties of $C(k, d, n_0)$.

LEMMA 2.2. *For any function $f \in C(k, d, n_0)$ the following items hold:*

(i) For any $n \geq n_0$, we have $f^{-1}(n + d) - f^{-1}(n) > n$;

(ii) For any $n \geq n_0 + 1$, the set $\{x \in \mathbb{N} \mid f(x) = n\}$ is nonempty;

(iii) For any $n \geq n_0 + 1$, the equality $f(x) = n$ implies

$$x > \frac{1}{2d}[(n-1)(n-2) - n_0(n_0 - 1)].$$

LEMMA 2.3. *If $f \in C(k,d,n_0)$ and $f(x) = n \geq n_0$, then for any c such that $1 \leq c \leq n$, we have:*

$$-k \leq f(x+c) - f(x) \leq k + d. \tag{6}$$

LEMMA 2.4. *For any $f \in C(k,d,n_0)$, let $x_0 = f^{-1}(2+4d+n_0^2+k)$. Consider $\tilde{f} : [x_0+1,\ +\infty[\cap \mathbb{N} \to \mathbb{N}$ with $\tilde{f}(x) = f(x)$. Then $\tilde{f}$ is existentially definable at first-order within $\langle \mathbb{N},\ +,\ 1,\ x \mapsto xf(x) \rangle$.*

Remarks 1) Item (i) of Lemma 2.2 provides a linear lower bound of values of f^{-1} when difference of arguments is the parameter d of the considered class.

Item (ii) of the same lemma insures that f is asymptotically onto, and item (iii) gives a quadradic lower bound for solutions of the equation $f(x) = n$.

2) Actually, as the reader will see within the proof, Lemma 2.2 does not use condition (4) of being k-almost increasing.

3) Lemma 2.3 provides asymptotical bounds for the difference of two values of f with arguments taken in a short interval with respect to the values of these arguments. Refering to the previous Lemma 2.2 we see that n is at most $\sqrt{2dx + n_0^2} + 2$.

4) Lemma 2.4 generalizes the situation of the main result of the first part when $\lfloor \log n \rfloor$ was "extracted", *i.e.* defined, from $+$ and $n \lfloor \log n \rfloor$.

2.4. Proofs of the three Lemmas.

PROOF. (Lemma 2.2) (i) By condition (5):

$$\begin{array}{rcl} f^{-1}(n+d) - f^{-1}(n) & = & [f^{-1}(n+d) - f^{-1}(n+d-1)] \\ & & +[f^{-1}(n+d-1) - f^{-1}(n+d-2)] \\ & & +\ldots \\ & & +[f^{-1}(n+1) - f^{-1}(n)] \\ & > & \frac{n+d-1}{d} + \frac{n+d-2}{d} + \ldots + \frac{n}{d} \\ & > & n. \end{array}$$

(ii) If there was no x such that $f(x) = n$, we would have $f^{-1}(n) = f^{-1}(n-1)$. But $f^{-1}(n) > f^{-1}(n-1)$ according to condition (5).

(iii) By definition of f^{-1}, we have: $x > f^{-1}(n-1)$.

As in (i), we have:

$$\begin{array}{rcl} f^{-1}(n-1) - f^{-1}(n_0) & = & [f^{-1}(n-1) - f^{-1}(n-2)] \\ & & +\ldots \\ & & +[f^{-1}(n_0+1) - f^{-1}(n_0)] \\ & > & \frac{n-2}{d} + \frac{n_0}{d} + \ldots + \frac{n}{d} \\ & = & \frac{(n-2)(n-1) - n_0(n_0+1)}{2d}. \end{array}$$

and the result. ⊣

PROOF. (Lemma 2.3) The left-hand side of the inequality is an immediate consequence of the very definition of a k-almost increasing function. For proving the right-hand side, note that, using k-almost increasing property of f together with $f(x) = n$, we obtain:

$$\max_{y \leq x} f(y) \leq f(x) + k = n + k,$$

so that $f^{-1}(n+k) \geq x$, by the definition of f^{-1}. By previous Lemma 2.2, item (i) and the latter inequality, we have:

$$f^{-1}(n+k+d) > f^{-1}(n+k) + n + k \geq x + n + k \geq x + n \geq x + c$$

since $1 \leq c \leq n$. Using again the definition of f^{-1}, we see that $f(x+c) \leq n + k + d = f(x) + k + d$ and we are done. ⊣

PROOF. (Lemma 2.4) To define $\tilde{f}$ within the structure $\langle \mathbb{N},\ +,\ x \mapsto xf(x) \rangle$ we shall make use of the inequality:

$$0 \leq f(x) < x$$

together with the remainder of $f(x)$ modulo $x+1$, which we must define in the considered structure.

FACT 2.5. $f(x) < x$.

PROOF. By the definition of f^{-1}, we have $f(x_0+1) > k + 2 + 4d + n_0^2$ and by the k-almost increasing property we deduce, for $x \geq x_0 + 1$,

$$n = f(x) \geq f(x_0 + 1) - k > 2 + 4d + n_0^2. \tag{7}$$

Hence $\frac{n-2}{2d} > 2$.

From (7), we obtain $n > n_0 + 1$ so that by Lemma 2.2, item (iii), we have:

$$x > \frac{1}{2d}[(n-1)(n-2) - n_0(n_0-1)],$$

hence:

$$x > 2(n-1) - \frac{n_0(n_0-1)}{2d} > 2(n-1) - n_0^2 = n + (n - 2 - n_0^2) > n = f(x).$$

⊣

FACT 2.6. We have:

$$f(x) \equiv (x+1)f(x+1) - xf(x) \pmod{x+1}. \tag{8}$$

PROOF. It is sufficient to note that $(x+1)f(x+1) - xf(x) = f(x) + (x+1)[f(x+1) - f(x)]$. ⊣

We are still unable to define general congruences. Fortunately here the difference $|f(x+1)-f(x)|$ is bounded, namely,

$$|f(x+1)-f(x)| \leq k+d, \tag{9}$$

due to Lemme 2.3, with $c=1$. This suggests to introduce the notion of a restricted congruence, namely, for a, b, m in $\mathbb{N}$ and some fixed integer c, we define $a \equiv_c b (\mathrm{mod}\, m)$ by:

$$\bigvee_{h=0}^{c} \{[a = b + \underbrace{m + \cdots + m}_{\text{h times}}] \vee [b = a + \underbrace{m + \cdots + m}_{\text{h times}}]\}.$$

Obviously, the first-order latter formula is expressible within the structure $\langle \mathbb{N},\ + \rangle$, since c is fixed. The congruence (8) and inequality (9) provide together the following restricted congruence:

$$f(x) \equiv_{k+d} (x+1)f(x+1) - xf(x) (\mathrm{mod}\, x+1),$$

which is a definition of $f(x)$ within $\langle \mathbb{N},\ +,\ 1,\ x \mapsto xf(x) \rangle$ since $1 \leq f(x) < x$. Finally, we provide explicitely an existential first-order definition of f, namely:

$$[x > x_0 \wedge y = f(x)] \leftrightarrow$$

$$\{x > x_0 \wedge y \leq x \wedge \bigvee_{h=0}^{k+d} [(y + xf(x) = (x+1)f(x+1) + \underbrace{(x+1) + \cdots + (x+1)}_{\text{h times}})$$

$$\vee ((x+1)f(x+1) = y + xf(x) + \underbrace{(x+1) + \cdots + (x+1)}_{\text{h times}})]\}.$$

$\dashv$

2.5. Fundamental Lemma and the proof of the Main Theorem. In order to prove the undecidability of $\mathrm{Th}(\mathbb{N},\ n \mapsto p_n,\ n \mapsto r_n)$, we prove a more general result, namely:

LEMMA 2.7. *(Fundamental Lemma) For any $f \in C(k, d, n_0)$, multiplication is existentially $\{+,\ 1,\ x \mapsto xf(x)\}$-definable at first-order.*

As shown by Y. MATIYASEVICH, the existential true theory of numbers is exactly the set of arithmetical relations, which are definable by diophantine equations. Therefore the negative solution of the 10-th Hilbert's problem [MY] implies the following corollary.

COROLLARY 2.8. *The existential theory $\mathrm{Th}_{\exists}(\mathbb{N},\ +,\ 1,\ x \mapsto xf(x))$ is undecidable.*

PROOF. (Lemma 2.7) It suffices to show that for some constants c and n_1 the function $n \mapsto cn^2$ from $[n_1, +\infty[\cap \mathbb{N}$ into $\mathbb{N}$ is $\{+, 1, x \mapsto xf(x)\}$-definable. More precisely, we shall take $c = 5d$ and $n_1 = 2 + 5d + n_0^2$. Consider $n \geq n_1$. Since $n_1 > n_0 + 1$, we can apply Lemma 2.2, item (ii), proving there exists x such that $f(x) = 5dn$. Let x_0 be the same as in Lemma 2.4, namely $x_0 = f^{-1}(2 + 4d + n_0^2 + k)$. Let us show $x > x_0$. Otherwise $x \leq x_0$, so that by the k-almost increasing property $f(x) \leq f(x_0) - k$, implying, by the definitions of f^{-1} and x_0,

$$f(x) \leq 2 + 4d + n_0^2 + k - k < n_1 < 5dn_1 \leq 5dn = f(x),$$

which is impossible.

Note that $5dn$ is $\{+\}$-definable as the sum of $5d$ terms equal to n (d is a fixed constant). Now thanks to Lemma 2.4, an x such that $f(x) = 5dn$ is $\{+, 1, x \mapsto xf(x)\}$-definable.

On the other hand:

$$(x+n)f(x+n) - xf(x) = (x+n)[f(x+n) - f(x)] + nf(x)$$

$$= (x+n)[f(x+n) - f(x)] + 5dn^2.$$

By Lemma 2.3 applied to $c = n$, we have $|f(x+n) - f(x)| \leq k + d$, so that:

$$5dn^2 \equiv_{k+d} (x+n)f(x+n) - xf(x) (\mathrm{mod}\, x+n). \tag{10}$$

According to Lemma 2.2, item (iii), since $f(x) = 5dn$ and $5dn > n_1 > n_0 + 1$ the inequalities $n \geq n_1 > n_0^2$ and:

$$\begin{aligned} x + n &> \frac{(5dn-1)(5dn-2)}{2d} - \frac{n_0(n_0-1)}{2d} + n \\ &> \frac{25d^2n^2 - 15nd}{2d} > 5dn^2 \end{aligned} \tag{11}$$

hold.

Using (10) and (11), a similar argument as in Lemma 2.4 shows that the function $n \mapsto 5dn^2 = cn^2$ with domain $[n_1, +\infty[\cap \mathbb{N}$ is existentially $\{+, 1, x \mapsto xf(x)\}$-definable. By a routine argument, multiplication is clearly existentially $\{+, 1, x \mapsto xf(x)\}$-definable. ⊣

PROOF. (Main-Theorem) We remind the reader that 1 was existentially $\{+, \mathbb{P}\}$ and $\{+, n \mapsto p_n\}$-defined in the introduction.

We also noted that $n \left\lfloor \frac{p_n}{n} \right\rfloor = p_n - r_n$ and $n \mapsto n \left\lfloor \frac{p_n}{n} \right\rfloor$ belongs to $C(1,1,11)$, the latter due to Proposition 2.1. Then Fundamental Lemma can be applied and multiplication is existentially $\{+, n \mapsto p_n, n \mapsto r_n\}$-definable. ⊣

2.6. Conclusion. Our main result is absolute in the sense that does not depend on any conjecture. In order to shed more light on the considered theories $\mathrm{Th}_\exists(\mathbb{N}, +, \mathbb{P})$ and $\mathrm{Th}_\exists(\mathbb{N}, n \mapsto p_n, n \mapsto r_n)$, we recall a conditional result of A. WOODS. Let us recall that DICKSON'S CONJECTURE [DL] claims that if

$a_1, a_2, \ldots a_n, b_1, b_2, \ldots b_n$ are integers with all $a_i > 0$ and

$$\forall y \neq 1 \exists x [y \not| \prod_{1 \leq i \leq n} (a_i x + b_i)]$$

then there exist infinitely many x such that $a_i x + b_i$ are primes for all i. Let us call DC this conjecture, then A. WOODS proved [WA]:

If DC is true then the existential theory $\mathrm{Th}_\exists(\mathbb{N}, +, \mathbb{P})$ is decidable.

Now, the question is to know whether there is a gap between $\mathrm{Th}_\exists(\mathbb{N}, +, n \mapsto p_n, n \mapsto r_n)$ and this one or whether they are exactly the same. In the case of equality between these two theories, DC is false (and hence SCHINZEL's HYPOTHESIS on primes, whose DC is the linear case, is also false).

Open problem 2: *Is $\mathrm{Th}_\exists(\mathbb{N}, +, \mathbb{P})$ equal to $\mathrm{Th}_\exists(\mathbb{N}, +, n \mapsto p_n, n \mapsto r_n)$?*

REFERENCES

[BJW] P.T. BATEMAN, C.G. JOCKUSCH, and A.R. WOODS, *Decidability and Undecidability of theories with a predicate for the prime*, **The Journal of Symbolic Logic**, vol. 58, (1993) pp.672-687.

[BOF] M. BOFFA, *More on an undecidability result of* BATEMAN, JOCKUSCH *and* WOODS, **The Journal of Symbolic Logic**, vol. 63, (1998) p. 50.

[CEG] P. CEGIELSKI, *Definability, decidability, complexity*, **Annals of mathematics and Artificial Intelligence**, **Baltzer**, M. NIVAT and S. GRIGORIEFF ed., vol. 16, 1996, nos 1-4, pp.311-342.

[CMR] P. CEGIELSKI, Y. MATIIASSEVITCH, and D. RICHARD, *Definability and decidability issues in extensions of the Integers with the divisibility predicate*, **The Journal of Symbolic Logic**, vol. 61, Number 2, June 1996, pp.515-540.

[DL] L.E. DICKSON, *A new extension of* DIRICHLET*'s theorem on prime numbers*, **Messenger of Mathematics**, vol. 33 (1903–04), pp. 155–161.

[END] H.B. ENDERTON, **A Mathematical Introduction to Logic**, Academic Press, (1972), XIII + 295p.

[LM] T. LAVENDHOMME & A. MAES, *Note on the undecidability of $\langle \omega, +, P_{m,r} \rangle$*, Definability in arithmetics and computability, pp.61-68, **Cahier du Centre de logique**, Belgium, 11 (2000).

[MY] Yuri MATIYASEVICH, **Hilbert's tenth Problem**, The MIT Press, Foundations of computing, 1993, XXII+262p.

[RIB] P. RIBENBOIM, **The new book of Prime Records**, Springer, 1996, XIV+541p.

[SS] A. SCHINZEL & W. SIERPIEŃSKY, *Sur certaines hypothses concernant les nombres premiers*, **Acta Arithmetica**, vol. 4, 1958, 185–208 and 5, 1959, 259.

[WA] Alan WOODS, **Some problems in logic and number theory, and their connection**, Ph.D. thesis, University of Manchester, Manchester, 1981.

LACL, UMR–FRE 2673,
UNIVERSITÉ PARIS 12 – IUT
ROUTE FORESTIÈRE HURTAULT
F-77300 FONTAINEBLEAU
E-mail: cegielski@univ-paris12.fr

LLAIC1
UNIVERSITÉ D'AUVERGNE – IUT INFORMATIQUE
B.P. 86
F-63172 AUBIÈRE CEDEX
E-mail: richard@iut.u-clermont1.fr

STEKLOV INSTITUTE OF MATHEMATICS (POMI)
27 FONTANKA
ST PETERSBURG
191011, RUSSIA
E-mail: vsemir@pdmi.ras.ru