On the additive theory of prime numbers I

Patrick CEGIELSKI[‡], Denis RICHARD[§] & Maxim VSEMIRNOV[¶]

21 mai 2001

Abstract

The undecidability of the additive theory of prime numbers is an open question. We show the undecidability of $\mathrm{Th}(\mathbb{N}, +, n \mapsto nf(n))$ where f is a good approximation of the enumeration $n \mapsto p_n/n$ and where p_n is the $(n+1)$-th prime.

Résumé

L'indécidabilité de la théorie additive des nombres premiers est une question ouverte. 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 et où p_n est le $(n+1)$-ième premier.

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 recently 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 $\langle \mathbb{N}, +, \mathbb{P} \rangle$. Two recent articles ([BJW] and [BOF]) in the JSL have shown that the undecidability of the first-order theory $\mathrm{Th}(\mathbb{N}, +, \mathbb{P})$ can be proved under certain assumptions as, for instance, the linear case in SCHINZEL's Hypothesis. 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 $\langle \mathbb{N}, +, \mathbb{P} \rangle$ itself, our investigation concerns $\mathrm{Th}(\mathbb{N}, +, n \mapsto p_n)$. This leads us immediately to the following.

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

[‡] LACL Université PARIS 12, IUT Route Forestière Hurtault F-77300 FONTAINEBLEAU,
- Email: cep@capella.liafa.jussieu.fr

[§] LLAIC1 Université d'Auvergne, IUT Informatique, B.P. 86, F-63172 AUBIÈRE Cedex
- Email: richard@iut.u-clermont1.fr

[¶] Steklov Institute of Mathematics (POMI), 27 Fontanka St PETERSBURG, 191011, Russia
- Email: vsemir@pdmi.ras.ru

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 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 $\mathcal{M}$, we denote by $\mathrm{DEF}(\mathcal{M})$ the set of constants, functions and relations which are first-order definable within $\mathcal{M}$. Since $\mathrm{Th}(\mathbb{N},+,\bullet)$ is undecidable (Turing, 1936) a method for proving the undecidability of the theory of a structure $\mathcal{M}$ consists in showing $\mathrm{DEF}(\mathcal{M}) = \mathrm{DEF}(\mathbb{N},+,\bullet)$. The set $\mathrm{DEF}(\mathbb{N},+,\bullet)$ is well-known and the inclusion $\mathrm{DEF}(\mathcal{M}) \subset \mathrm{DEF}(\mathbb{N},+,\bullet)$ 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.

Proposition 1. *The equality* $\mathrm{DEF}(\mathbb{N},+,\mathrm{n}\mapsto \mathrm{n}.\lfloor\log_2(\mathrm{n})\rfloor) = \mathrm{DEF}(\mathbb{N},+,\bullet)$ *holds so that* $\mathrm{Th}(\mathbb{N},+,\mathrm{n}\mapsto\mathrm{n}.\lfloor\log_2(\mathrm{n})\rfloor)$ *is undecidable.*

This result is a corollary of Proposition 2 below which we shall prove after the introduction of the mappings used.

Definition 1. *For any positive integer* k, *we define* $\mathrm{f_k}$ *and* $\exp_\mathrm{k}$ *as mappings respectively from a final segment of* $\mathbb{N}$ *into* $\mathbb{N}$ *which are such that*

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

$$\exp_\mathrm{k}(\mathrm{n}) = 2^{2^{\cdot^{\cdot^{2^\mathrm{n}}}}}$$

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

Proposition 2. *The equality* $\mathrm{DEF}(\mathbb{N},+,\mathrm{f_k}) = \mathrm{DEF}(\mathbb{N},+,\bullet)$ *holds so that* $\mathrm{Th}(\mathbb{N},+,\mathrm{f_k})$ *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 2 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 2 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 2. *The class* C *is the set of all (invertible) real functions* $\mathrm{f} : [\mathrm{a}_0, +\infty[\to \mathbb{R}$ *satisfying the following conditions :*

1) f *is continuous ;*

2) f *is strictly increasing ;*

3) $\lim_{\mathrm{x} \to +\infty} \mathrm{f}(\mathrm{x}) = +\infty$ *;*

4) for every x *which is positive real,* $\mathrm{f}(\mathrm{x}) < \mathrm{x}$.

5) There exists $\mathrm{x}_0 \in (\mathbb{R}^*)^+$ *such that for all reals* $\mathrm{x} \geq \mathrm{x}_0$ *the inequality* $\mathrm{f}(\mathrm{x}+1) < \mathrm{f}(\mathrm{x}) + 1/2$ *holds.*

Examples : The functions $x \mapsto \log_2(x)$, $x \to \log_2(x) + \log_2(\log_2 x) - 1$, $x \mapsto \log(x)$ 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 paper is a straightforward corollary of the following

Proposition 3. (Fundamental Lemma). *For any function* f *of the class* C *(see Definition 2), we have* $\mathrm{DEF}(\mathbb{N}, +, \mathrm{n} \mapsto \mathrm{n}\lfloor \mathrm{f}(\mathrm{n}) \rfloor) = \mathrm{DEF}(\mathbb{N}, +, \bullet)$ *so that* $\mathrm{Th}(\mathbb{N}, +, \mathrm{n} \mapsto \mathrm{n}\lfloor \mathrm{f}(\mathrm{n}) \rfloor)$ *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. 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 2, $f(p_0+1) < f(p_0) + \frac{1}{2}$ 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, 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 2. For every positive integer, $\tilde{f}(n+1) \leq \tilde{f}(n) + 1$.
It trivially follows from Condition 5) of Definition 2.

Fact 3. For every $f \in C$ and for all positive reals we have $f^{-1}(x+1) - f^{-1}(x) \geq 1$.
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 4. We have $\tilde{f}(h(n)) = n$ and $h(\tilde{f}(n)) \geq n$ for every integer $n \geq \tilde{f}(0)$.
The first equality comes from Fact 1 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 of Proposition 3. We begin by showing that the set $h(\mathbb{N})$ belongs to $\mathrm{DEF}(\mathbb{N}, +, n \mapsto n\tilde{f}(n))$. 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. 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 2(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 3 is to show the $\langle \mathbb{N}, +, g \rangle$-definability of $\tilde{f}$. For this purpose, we intend to prove firstly that for every $n \geq \tilde{f}(0)$, 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 4) 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 3) 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 $n \in \mathbb{N}$ satisfies $h(n+1) - h(n) > n$. 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

$$\begin{aligned}&\{[(h(n+1)-h(n)>n) \wedge g(h(n)+n) =\\ &g(h(n))+h(n)+n+m] \vee (h(n+1)-h(n) \leq n) \wedge [g(h(n+1)+n+1) =\\ &g(h(n+1))+h(n+1)+n+1+m+n+n+1]\}. \qquad \square\end{aligned}$$

Proposition 3 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 4. *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 *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 2. We check Condition 6. Since $q_0 = h(n) = \mathrm{Max}\{q \in \mathbb{N}$ such that $\exists\alpha_q(0 \leq \alpha_q < 1)$ and $f(q) = n+\alpha_q\}$, we also get $h(n) = \mathrm{Max}\{q \in \mathbb{N}$ such that $\exists\alpha_q(0 \leq \alpha_q < 1)$ and $\log q + \log(\log q) - 1 = n + \alpha_q\} = \mathrm{Max}\{q \in \mathbb{N}$ 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 $\mathrm{Max}\{q' \in \mathbb{N}$ 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. $\square$

Remark - Inspecting the definition of multiplication in the proof of Proposition 3, we see the obtained formula of definition is existential (only order appears). Consequently the $\mathbb{P}i_2$-theory of $\langle \mathbb{N}, +, n \mapsto n\tilde{f}(n)\rangle$ is undecidable.

References

[BJW] P.T. Bateman, C.G. Jockusch and A.R. Woods : *Decidability and Undecidability of theories with a predicate for the prime.* This **Journal**, vol. 58, (1993) pp.672-687.

[BOF] M. Boffa : *More on an undecidability result of* Bateman, Jockusch *and* Woods. This **Journal**, vol. 63, (1998) p. 50.

[CEG] P. Cegielski : *Definability, decidability, complexity.* **Annals of mathematics and Artificial Intelligence**, **Baltzer** ed. M. Nivat, S. Grigorieff 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.* This **Journal**, vol. 61, Number 2, June 1996, pp.515-540.

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

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