Journal of Integer Sequences, Vol. 22 (2019),
Article 19.2.1

Arithmetic Progressions in the Graphs of Slightly Curved Sequences

Kota Saito and Yuuya Yoshida
Graduate School of Mathematics
Nagoya University
Furo-cho, Chikusa-ku, Nagoya, 464-8602
Japan
m17013b@math.nagoya-u.ac.jp
m17043e@math.nagoya-u.ac.jp

Abstract

A strictly increasing sequence of positive integers is called a *slightly curved sequence with small error* if the sequence can be well-approximated by a function whose second derivative goes to zero faster than or equal to $1/x^{\alpha}$ for some $\alpha > 0$. In this paper, we prove that arbitrarily long arithmetic progressions are contained in the graph of a slightly curved sequence with small error. Furthermore, we extend Szemerédi's theorem to a theorem about slightly curved sequences. As a corollary, it follows that the graph of the sequence $\{\lfloor n^a \rfloor\}_{n\in A}$ contains arbitrarily long arithmetic progressions for every $1 \le a < 2$ and every $A \subset \mathbb{N}$ with positive upper density. Using this corollary, we show that the set $\{\, \lfloor \lfloor p^{1/b} \rfloor^a \rfloor \mid p \text{ prime} \,\}$ contains arbitrarily long arithmetic progressions for every $1 \le a < 2$ and $b > 1$. We also prove that, for every $a \ge 2$, the graph of $\{\lfloor n^a \rfloor\}_{n=1}^{\infty}$ does not contain any arithmetic progressions of length 3.

1 Introduction

This paper considers problems involving arithmetic progressions. Let $k \ge 3$ and $d \ge 1$ be integers. A sequence $\{a(j)\}_{j=0}^{k-1} \subset \mathbb{N}^d$ is called an *arithmetic progression (AP) of length k* if there exists $D \in \mathbb{N}^d$ such that

$$a(j) = a(0) + jD$$

for all $j = 0, 1, \dots, k-1$. We discuss only the cases $d = 1$ and $d = 2$. Here, note the following two points. First, all components of the common difference D are positive, since $\mathbb{N} = \{1, 2, \dots\}$ does not contain zero. Second, this paper addresses strictly increasing sequences alone and then regards sequences of positive integers as subsets of $\mathbb{N}$. That is why we denote a sequence by, say, $\{a(n)\}_{n=1}^{\infty}$ instead of $(a(n))_{n=1}^{\infty}$.

APs have elicited much interest from researchers studying arithmetic combinatorics, geometric measure theory, and fractal geometry. Most studies consider the density of sets to ensure the existence of long APs. For example, we recall Szemerédi's celebrated result:

Proposition 1 (Szemerédi [1]). *For every $k \geq 3$ and $0 < \delta \leq 1$ there exists an integer $N(k, \delta) > 0$ such that if $N \geq N(k, \delta)$, then every set $A \subset \{1, 2, \dots, N\}$ with $|A| \geq \delta N$ contains an AP of length k.*

Here $|X|$ denotes the cardinality of a finite set X. Furthermore, Steinhaus showed that every set with positive Lebesgue measure contains arbitrarily long APs from Lebesgue's density theorem; for instance, see [2, Theorem 3]. These sufficient conditions are traditional and important, but it is difficult to weaken them. Hence we have found a new class of sets containing arbitrarily long APs, which is a main contribution of this paper. We call this new class *slightly curved sequences*, which are defined below.

Let $g : \mathbb{N} \to \mathbb{R}$ be an eventually positive function and let $\mathbb{R}^+ = (0, \infty)$. A strictly increasing sequence $\{a(n)\}_{n=1}^{\infty} \subset \mathbb{N}$ is called a *slightly curved sequence with error $O(g(n))$* if there exists a twice differentiable function $f : \mathbb{R}^+ \to \mathbb{R}$ such that

$$\begin{aligned} f''(x) &= O(1/x^{\alpha}), \\ a(n) &= f(n) + O(g(n)) \end{aligned} \tag{1}$$

for some $\alpha > 0$. A slightly curved sequence with error $o(g(n))$ is also defined in the same way. Here the notation $f(x) = O(g(x))$ denotes that there exist $C > 0$ and $x_0 > 0$ such that $|f(x)| \leq Cg(x)$ for all $x \geq x_0$, where $g(x)$ is an eventually positive function. In this paper, the constant C often depends on the length k or the exponent α in (1). When emphasizing the dependence on k or α, we write $f(x) = O_{k,\alpha}(g(x))$. Also, the notation $f(x) = o(g(x))$ denotes that $f(x)/g(x)$ goes to zero as $x \to \infty$. Furthermore, to address two-dimensional APs, we define the *graph* of a sequence: for every $A \subset \mathbb{N}$, the *graph* of a sequence $\{a(n)\}_{n \in A} \subset \mathbb{N}$ is defined as the set $\{(n, a(n)) \in \mathbb{N}^2 \mid n \in A\}$.

Surprisingly, we do not impose any density conditions on sets of the above new class. Instead of density conditions, we restrict the behavior of leading terms of sequences. The first goal of this paper is to prove the following theorem:

Theorem 2. *Fix an integer $k \geq 3$. The graph of every slightly curved sequence with error $o((\log \log n)^{1/c_k})$ contains an AP of length k, where $c_k = 2^{2^{k+9}}$.*

For small k, we can improve the above error term:

Theorem 3. *The graph of every slightly curved sequence with error* $o((\log n)(\log\log n)^{-4})$ *contains APs of length* 3*. The graph of every slightly curved sequence with error* $o((\log n)^c)$ *contains APs of length* 4*, where* $c > 0$ *is an absolute constant.*

Theorems 2 and 3 are proved in Section 3. Since the second derivative (1) of f must vanish as $x \to \infty$, we cannot apply Theorem 2 or 3 to the case when f is a quadratic function. To compensate this weak point, Appendix A proves another result without the assumption $f''(x) = O(1/x^\alpha)$, which can be applied to the sequence $\{\lfloor c_2 n^2 + c_1 n + c_0 \rfloor\}_{n=1}^\infty$ with a sufficiently small $c_2 > 0$.

To state the second goal of this paper, let us introduce basic notions. We say that a set $A \subset \mathbb{N}$ has *positive upper density* if the condition

$$\limsup_{N\to\infty} \frac{|A \cap [1, N]|}{N} > 0$$

holds. We also say that a set $A \subset \mathbb{N}$ has *positive upper Banach density* if the condition

$$\limsup_{N\to\infty} \frac{\max_{n\geq 0} |A \cap [n+1, n+N]|}{N} > 0$$

holds. If a set $A \subset \mathbb{N}$ has positive upper density, then *a fortiori* A has also positive upper Banach density. However, the converse does not hold in general.

The second goal of this paper is to show the following result:

Theorem 4 (An extension of Szemerédi's theorem)**.** *If* $\{a(n)\}_{n=1}^\infty$ *is a slightly curved sequence with error* $O(1)$ *and a set* $A \subset \mathbb{N}$ *has positive upper Banach density, then the graph of* $\{a(n)\}_{n\in A}$ *contains arbitrarily long APs.*

As a corollary, we also obtain the following result:

Corollary 5. *If a set* $A \subset \mathbb{N}$ *has positive upper Banach density, then the graph of* $\{\lfloor n^a \rfloor\}_{n\in A}$ *contains arbitrarily long APs for every* $1 \leq a < 2$.

Here, for every $x \in \mathbb{R}$ the notation $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x and the notation $\lceil x \rceil$ denotes the least integer greater than or equal to x. Corollary 5 with $a = 1$ is just Szemerédi's theorem (Proposition 1). By using Corollary 5, Section 2 shows that the set $\{ \lfloor \lfloor p^{1/b} \rfloor^a \rfloor \mid p \text{ prime} \}$ contains arbitrarily long APs for every $1 \leq a < 2$ and $b > 1$.

Proof of Corollary 5 assuming Theorem 4. Fix $1 \leq a < 2$ and let $f(x) = x^a$. Then $\lfloor n^a \rfloor = f(n) + O(1)$. Since $f''(x) = O(1/x^{2-a})$, Theorem 4 implies Corollary 5. □

In particular, Corollary 5 with $A = \mathbb{N}$ implies the following result immediately:

Corollary 6. *The graph of* $\{\lfloor n^a \rfloor\}_{n=1}^\infty$ *contains arbitrarily long APs for every* $1 \leq a < 2$.

When $a > 1$ and $a \notin \mathbb{N}$, the sequences $\{\lfloor n^a \rfloor\}_{n=1}^{\infty}$ are called Piatetski-Shapiro sequences. Hence Corollary 6 implies that the Piatetski-Shapiro sequence with every exponent $1 < a < 2$ contains arbitrarily long APs. Piatetski-Shapiro sequences are named after Piatetski-Shapiro [3], who proved that for every $1 < a < 12/11$ the sequence $\{\lfloor n^a \rfloor\}_{n=1}^{\infty}$ contains infinitely many primes. It is known that the range $1 < a < 12/11$ can be improved to $1 < a < 243/205$ [4]. Many preceding studies considered prime numbers contained in Piatetski-Shapiro sequences. For instance, Mirek [5] proved that for every $1 < a < 72/71$ the set of all primes of the form $\lfloor n^a \rfloor$ contains an AP of length 3. Recently, Li and Pan [6] claimed that for every integer $k \geq 3$ there exists $a_k > 1$ such that for every $1 < a < a_k$ the set of all primes of the form $\lfloor n^a \rfloor$ contains an AP of length k. Although this paper considers Piatetski-Shapiro sequences, prime numbers are not our main topic.

Remark 7. When $a > 1$, the sum of the reciprocals of $\{\lfloor n^a \rfloor\}_{n=1}^{\infty}$ converges:

$$\sum_{n=2}^{\infty} \frac{1}{\lfloor n^a \rfloor} \leq \sum_{n=2}^{\infty} \frac{1}{n^a - 1} \leq \sum_{n=2}^{\infty} \frac{2}{n^a} < \infty.$$

Thus we cannot apply the Erdős-Turán conjecture to this sequence. Here the Erdős-Turán conjecture asserts that every subset of positive integers whose sum of reciprocals diverges must contain arbitrarily long APs [7]. Since the sum of the reciprocals of all primes diverges, the Erdős-Turán conjecture implies that the set of all primes contains arbitrarily long APs. Green and Tao [8] proved that the set of all primes contains arbitrarily long APs, but the Erdős-Turán conjecture is still open even in the case when the length of an AP is three. The Erdős-Turán conjecture is a strong statement that implies the result by Green and Tao; nevertheless, it does not imply Corollary 6 directly.

Corollary 6 ensures that the graph of $\{\lfloor n^a \rfloor\}_{n=1}^{\infty}$ with $1 \leq a < 2$ contains arbitrarily long APs, but the graph of $\{\lfloor n^a \rfloor\}_{n=1}^{\infty}$ with $a \geq 2$ does not contain any APs of length 3, which is proved in Section 6. Hence the graph of $\{n^2\}_{n=1}^{\infty}$ does not contain any APs of length 3, while the sequence $\{n^2\}_{n=1}^{\infty}$ (not the graph) contains APs of length 3, e.g., $\{1, 25, 49\}$. However, there are not any APs of length 4 in $\{n^2\}_{n=1}^{\infty}$, which was proposed by Fermat in 1640 and first shown by Euler in 1780, according to Dickson's book [9, pp. 440 and 635]. Euler also showed that there are not any APs of length 3 in $\{n^3\}_{n=1}^{\infty}$, according to [9, pp. 572–573]. Moreover, Dénes [10] showed that for every integer $3 \leq a \leq 30$ there are not any APs of length 3 in $\{n^a\}_{n=1}^{\infty}$. Finally, for every integer $a \geq 3$, Darmon and Merel [11] showed that $\{n^a\}_{n=1}^{\infty}$ contains no APs of length 3. We do not know whether $\{\lfloor n^a \rfloor\}_{n=1}^{\infty}$ would contain long APs if $a > 2$ is not an integer. One might guess that such a sequence would not contain APs, but the following sequences are APs of length 4:

$$\{\lfloor 2^{2.2} \rfloor, \lfloor 11^{2.2} \rfloor, \lfloor 15^{2.2} \rfloor, \lfloor 18^{2.2} \rfloor\},$$
$$\{\lfloor 14^{2.655015} \rfloor, \lfloor 39^{2.655015} \rfloor, \lfloor 50^{2.655015} \rfloor, \lfloor 58^{2.655015} \rfloor\},$$
$$\{\lfloor 27^{2.720398} \rfloor, \lfloor 89^{2.720398} \rfloor, \lfloor 114^{2.720398} \rfloor, \lfloor 132^{2.720398} \rfloor\}.$$

In the course of this study, we proved a result in Appendix B, which is that the graph of $\{\lfloor f(n) \rfloor\}_{n=1}^{\infty}$ contains an AP of length 4 if a function $f : \mathbb{N} \to \mathbb{R}^+$ satisfies $\Delta f > 0$,

$\Delta^2 f \geq 0$, and $\liminf_{n\to\infty} f(n)/n^2 < 1/18$, where Δ denotes the difference operator which is defined in Section 3. This result can be applied to the sequence $\{\lfloor c_2n^2 + c_1n + c_0\rfloor\}_{n=1}^{\infty}$ with $0 < c_2 < 1/18$, but Theorem 2 or 3 cannot be applied.

2 Warming up

As an application of Corollary 5, this section shows that the set $\{\, \lfloor\lfloor p^{1/b}\rfloor^a\rfloor \mid p \text{ prime}\,\}$ contains arbitrarily long APs for every $1 \leq a < 2$ and $b > 1$. Once we put $A = \{\, \lfloor p^{1/b}\rfloor \mid p \text{ prime}\,\}$, the above set can be expressed as $\{\, \lfloor\lfloor p^{1/b}\rfloor^a\rfloor \mid p \text{ prime}\,\} = \{\lfloor n^a\rfloor\}_{n\in A}$. Thus, if A has positive upper density, the above statement follows from Corollary 5. Let us verify that A has positive upper density. Let $\pi(x)$ be the number of primes less than or equal to $x > 0$ and $\tilde{\pi}(S)$ be the number of primes contained in a set $S \subset \mathbb{R}^+$. Then every $x > 0$ satisfies

$$\begin{aligned}
&|A \cap [1,x]| = \left|\{\, \lfloor p^{1/b}\rfloor \leq x \mid p \text{ prime}\,\}\right| \\
&\geq \frac{\left|\{\, p \leq x^b \mid p \text{ prime}\,\}\right|}{\max_{n\leq x}\tilde{\pi}[n^b,(n+1)^b)} \geq \frac{\pi(x^b)}{\max_{n\leq x}\tilde{\pi}[n^b,(n+1)^b]}.
\end{aligned}$$

Since there exists $y_0 > 0$ such that every $x > 0$ and $y \geq y_0$ satisfy $\tilde{\pi}[x, x+y] \leq 3y/\log y$ [12, Corollary 3.4], the denominator $\tilde{\pi}[n^b,(n+1)^b]$ of the above right-hand side is upper bounded as

$$\begin{aligned}
&\tilde{\pi}[n^b,(n+1)^b] \overset{\text{(i)}}{\leq} \max_{n\leq x}\tilde{\pi}[n^b, n^b + b(n+1)^{b-1}] \\
&\leq \max_{n\leq x}\tilde{\pi}[n^b, n^b + b(x+1)^{b-1}] \overset{\text{(ii)}}{\leq} \frac{3b(x+1)^{b-1}}{\log(b(x+1)^{b-1})}
\end{aligned}$$

for every $x \geq (y_0/b)^{1/(b-1)} - 1$, where (i) and (ii) follow from the mean value theorem and [12, Corollary 3.4], respectively. We now use the prime number theorem [12, Chapter 6]: $\pi(x) \sim x/\log x$, where the notation $f(x) \sim g(x)$ denotes that $f(x)/g(x)$ goes to one as $x \to \infty$. Therefore, the prime number theorem yields

$$\begin{aligned}
&\frac{|A\cap[1,x]|}{x} \geq \pi(x^b)\Big/ x\frac{3b(x+1)^{b-1}}{\log(b(x+1)^{b-1})} \\
&\sim \frac{x^b}{b\log x}\Big/ x\frac{3b(x+1)^{b-1}}{(b-1)\log(x+1)} \sim \frac{b-1}{3b^2} > 0 \quad (x\to\infty),
\end{aligned}$$

whence A has positive upper density.

As a special case of the above result, it follows that the set $\{\, \lfloor\lfloor p^{1/b}\rfloor^b\rfloor \mid p \text{ prime}\,\}$ contains arbitrarily long APs for every $1 < b < 2$. However, the case $b = 1$, i.e., the Green-Tao theorem [8] would not be proved by the above method. This is because the set $\{\, \lfloor p^{1/b}\rfloor \mid p \text{ prime},\ |\lfloor p^{1/b}\rfloor^b - p| \leq C\,\}$ has probably upper density zero for every $1 < b < 2$ and $C > 0$.

3 More general statement than Theorems 2 and 3

In order to prove Theorems 2 and 3, this section introduces another proposition which is proved in Section 4. This proposition uses a function satisfying three properties below. Before stating them, let us define the *van der Waerden numbers* [13]. For every $k \geq 3$ and $r \geq 2$, the *van der Waerden number* $W(r,k)$ is the smallest number N such that if $\{1, 2, \ldots, N\}$ are partitioned into r different sets then there exists at least one set which contains an AP of length k. Fix an integer $k \geq 3$. This section considers a function U_k satisfying the following properties:

(U1) there exist $x_k \geq 1$ and $y_k \geq 1$ such that a function $U_k : [x_k, \infty) \to [y_k, \infty)$ is increasing and bijective;

(U2) $U_k(r^\alpha) = O_{k,\alpha}(U_k(r))$ for all $\alpha > 0$;

(U3) the inverse function U_k^{-1} of U_k satisfies $W(r,k) = O_k(U_k^{-1}(r))$.

Using such a function U_k, we obtain the following proposition:

Proposition 8. *Fix an integer $k \geq 3$. Assume that a function U_k satisfies (U1), (U2), and (U3). If $\{a(n)\}_{n=1}^\infty$ is a slightly curved sequence with error $o(U_k(n))$, then there exists an arithmetic progression P of length k such that $a(P)$ is also an arithmetic progression of length k.*

The proof of Theorem 2 uses the function $U_k(r) = (\log\log r)^{1/c_k}$ with $c_k = 2^{2^{k+9}}$. Similarly, the proof of Theorem 3 uses the functions $U_3(r) = C_3^{-1}(\log r)(\log\log r)^{-4}$ and $U_4(r) = C_4^{-1}(\log r)^c$ for some $C_3, C_4 > 0$ and some absolute constant $c > 0$. Thus Theorems 2 and 3 are special cases of Proposition 8. To prove Theorem 2 by assuming Proposition 8, we need to use upper bounds of the van der Waerden numbers. As such upper bounds, Gowers' upper bounds [14, Theorem 18.6] are known:

$$W(r,k) \leq 2^{2^{r^{c_k}}}, \quad c_k = 2^{2^{k+9}}. \tag{2}$$

For general r and k, Gowers' upper bounds are best at present.

Proof of Theorem 2 assuming Proposition 8. Fix an integer $k \geq 3$. Let $U_k(r) = (\log\log r)^{1/c_k}$ with $c_k = 2^{2^{k+9}}$. It is enough to show that the function U_k satisfies (U1), (U2), and (U3). The properties (U1) and (U2) follow from the definition of U_k. Also, Gowers' upper bound (2) implies that the function U_k satisfies $W(r,k) \leq \exp\bigl(\exp(r^{c_k})\bigr) = U_k^{-1}(r)$, which is just (U3). □

Next, to prove Theorem 3 by assuming Proposition 8, we introduce the notation $r_k(N)$ [7]. The notation $r_k(N)$ is the maximum cardinality of all subsets of $\{1, 2, \ldots, N\}$ that contain no APs of length k. For $k = 3, 4$, the following upper bounds of $r_k(N)$ are known [15, Theorem 1.1], [16, Theorem 1.1]:

$$r_3(N) = O\Big(\frac{N}{(\log N)(\log\log N)^{-4}}\Big), \quad r_4(N) = O\Big(\frac{N}{(\log N)^c}\Big) \tag{3}$$

for some absolute constant $c > 0$. As explained below, the upper bounds (3) correspond to upper bounds of the van der Waerden numbers $W(r, k)$ with $k = 3, 4$, which are better than Gowers' upper bounds (2).

Proof of Theorem 3 assuming Proposition 8. First, we show that if a function U_k satisfies (U1) and

$$r_k(N) < N/U_k(N) \quad (N \geq x_k) \tag{4}$$

then the function U_k satisfies (U3). Due to (U1), there exists $r_k > 0$ such that every $r \geq r_k$ satisfies $U_k^{-1}(r) \geq 1$. Let $r \geq r_k$ be an integer and let $N = \lceil U_k^{-1}(r) \rceil$. The definition of N yields $N = \lceil U_k^{-1}(r) \rceil \geq U_k^{-1}(r)$ and then (U1) does $U_k(N) \geq r$. Take an arbitrary partition of $\{1, 2 \ldots, N\}$ into r small sets. Then the pigeonhole principle implies that there exists a set A consisting of at least $\lceil N/r \rceil$ elements. Thus A contains an AP of length k because the inequality

$$|A| \geq \lceil N/r \rceil \geq N/r \overset{\text{(i)}}{\geq} N/U_k(N) \overset{(4)}{>} r_k(N)$$

holds, where (i) follows from $U_k(N) \geq r$. Since the partition of $\{1, 2 \ldots, N\}$ is arbitrary, the van der Waerden number $W(r, k)$ is upper bounded by N. Thus it follows that

$$W(r, k) \leq N \leq U_k^{-1}(r) + 1 \leq 2U_k^{-1}(r)$$

from the definition of N and (U1). Therefore, the function U_k satisfies (U3).

Next, using the result in the previous paragraph, we show Theorem 3. The upper bounds (3) imply that there exist constants $C_3, C_4 > 0$ such that

$$r_3(N) < C_3 \cdot \frac{N}{(\log N)(\log\log N)^{-4}}, \quad r_4(N) < C_4 \cdot \frac{N}{(\log N)^c}$$

for every sufficiently large N. If taking the functions $U_3(N) = C_3^{-1}(\log N)(\log\log N)^{-4}$ and $U_4(N) = C_4^{-1}(\log N)^c$, the inequality (4) holds for $k = 3$ and $k = 4$, respectively. Since the functions U_3 and U_4 satisfy (U1) and (U2), the result in the previous paragraph implies that the functions U_3 and U_4 also satisfy (U3). Therefore, Proposition 8 implies Theorem 3. $\square$

4 Proof of Proposition 8

This section proves Proposition 8. Before proving Proposition 8, let us define a semi-norm on the vector space $\mathcal{F} = \{f \mid f : \mathbb{R}^+ \to \mathbb{R}\}$. Let $k \geq 3$ be an integer and $P = \{b(j)\}_{j=0}^{k-1} \subset \mathbb{R}^+$

be a strictly increasing sequence. We define

$$N_P(f) = \sum_{j=0}^{k-3} \left|\Delta^2[f \circ b](j)\right|,$$

for every $f \in \mathcal{F}$, where Δ denotes the difference operator, that is,

$$\Delta f(x) = f(x+1) - f(x),$$

and $\Delta^2 := \Delta \circ \Delta$. We can find that N_P satisfies the following properties:

(N1) for every strictly increasing function $f \in \mathcal{F}$,

$$N_P(f) = 0 \text{ if and only if } f(P) \text{ is an AP of length } k;$$

(N2) $N_P(f) \geq 0$ for all $f \in \mathcal{F}$;

(N3) $N_P(f+g) \leq N_P(f) + N_P(g)$ for all $f, g \in \mathcal{F}$.

All the properties above can be easily checked from the definition.

Proof of Proposition 8. Fix an integer $k \geq 3$ and take an arbitrary slightly curved sequence $\{a(n)\}_{n=1}^{\infty}$ with error $o(U_k(n))$: there exists a twice differentiable function $f : \mathbb{R}^+ \to \mathbb{R}$ satisfying (1) and $a(n) = f(n) + o(U_k(n))$. Let r be a sufficiently large positive integer. We define $W(r) = W(r,k)$, $R(n) = a(n) - f(n)$, and

$$A(r) = \left\{\, m \in \mathbb{N} \;\middle|\; \lfloor W(r)^{3/\alpha} \rfloor \leq m < \lfloor W(r)^{3/\alpha} \rfloor + W(r) \,\right\}.$$

Then every $m \in A(r)$ satisfies

$$R(m) = o\big(U_k(2W(r)^{\max\{3/\alpha,1\}})\big) = o\big(U_k(U_k^{-1}(r))\big) = o(r)$$

thanks to (U1), (U2), and (U3). Thus there exists a positive function $\delta(r)$ such that

$$R(m) \in [-\delta(r), \delta(r))$$

for all $m \in A(r)$, and

$$\delta(r) = o(r) \tag{5}$$

as $r \to \infty$. We also define

$$I_j = \Big[-\delta(r) + \frac{2\delta(r)}{r} j,\ -\delta(r) + \frac{2\delta(r)}{r}(j+1)\Big), \quad A_j(r) = \{m \in A(r) \mid R(m) \in I_j\}$$

for all $j = 0, 1, \ldots, r-1$. Note that the union of all the small sets $A_j(r)$ equals $A(r)$. From the definition $W(r) = W(r,k)$, it follows that there exists an integer $q \in \{0, 1, \ldots, r-1\}$ such that $A_q(r)$ contains an arithmetic progression $P = \{b(j)\}_{j=0}^{k-1}$ of length k. Here $b(j)$ is expressed as $b(j) = dj + e$ with two integers $d, e > 0$. Since every $j = 0, 1, \ldots, k-1$ satisfies $b(j) \in A(r)$, we obtain

$$d \le W(r), \quad \lfloor W^{3/\alpha}(r) \rfloor \le e. \tag{6}$$

Then the triangle inequality (N3) implies $N_P(a) \le N_P(f) + N_P(R)$.

Finally, we show that $N_P(a) = o(1)$ as $r \to \infty$, which follows from the two relations $N_P(f) = o(1)$ and $N_P(R) = o(1)$. Since we have (5) and $b(j) \in A_q(r)$ for all $j = 0, 1, \ldots, k-1$, the inequality

$$N_P(R) \le \sum_{j=0}^{k-3} (|\Delta[R \circ b](j+1)| + |\Delta[R \circ b](j)|) \le \frac{4(k-2)\delta(r)}{r} = o(1)$$

holds as $r \to \infty$. Hence the remaining is to show that $N_P(f) = o(1)$ as $r \to \infty$. The mean value theorem implies that for every $j = 0, 1, \ldots, k-3$ there exist $\theta_j, \eta_j \in (0,1)$ such that

$$\begin{aligned} \Delta^2[f \circ b](j) &= \Delta[f \circ b](j+1) - \Delta[f \circ b](j) = (\Delta[f \circ b])'(j + \theta_j) \\ &= d\{f' \circ b(j + \theta_j + 1) - f' \circ b(j + \theta_j)\} = d^2 f'' \circ b(j + \theta_j + \eta_j), \end{aligned}$$

where $b(x) := dx + e$. Since (6) and the assumption $f''(x) = O(1/x^\alpha)$ hold, every $j = 0, 1, \ldots, k-3$ satisfies

$$\Delta^2[f \circ b](j) = d^2 f''(d(j + \theta_j + \eta_j) + e) = O\Big(W(r)^2 \cdot \frac{1}{W(r)^3}\Big) = o(1)$$

as $r \to \infty$, whence $N_P(f) = o(1)$ as $r \to \infty$. Therefore, a sufficiently large integer $r > 0$ satisfies $N_P(a) < 1$. Since $N_P(a)$ is a non-negative integer from the definition, the equation $N_P(a) = 0$ holds, which implies that $a(P)$ is also an AP of length k due to (N1). $\square$

5 Proof of Theorem 4

This section considers, for every $\alpha > 0$, whether the set $A \cap (h_\alpha(x), h_\alpha(x+1)]$ contains an AP or not, which is useful to prove Theorem 4. Here h_α is a function satisfying the following properties:

(H1) h_α is a differentiable function defined on the interval (x_0, ∞) for some $x_0 > 0$;

(H2) there exists a positive number $x_1 \ge x_0$ such that h_α and h'_α strictly increase on the interval (x_1, ∞);

(H3) $\lim_{x\to\infty} h'_\alpha(x+1)^2 / h_\alpha(x)^\alpha = 0$;

(H4) $\lim_{x\to\infty} h'_\alpha(x) = \infty$.

For example, the function $h_\alpha(x) = x \log x$ satisfies the above three properties. As another example, we can take the function $h_\alpha(x) = x^{1+\alpha/2}$. The latter example is used in Appendix A. In order to prove Theorem 4, we introduce the following condition for a set $A \subset \mathbb{N}$:

(C) for every integer $k \geq 3$ and every $\alpha > 0$, there exist a function h_α satisfying (H1)–(H3) and a strictly increasing sequence $\{x_n\}_{n=1}^{\infty} \subset \mathbb{R}^+$ which diverges such that for every integer $n > 0$ the set $A \cap (h_\alpha(x_n), h_\alpha(x_n+1)]$ contains an AP of length k.

Theorem 4 follows from the following proposition and preliminary lemma.

Proposition 9. *Let A be a subset of positive integers with (C), and $\{a(n)\}_{n=1}^{\infty}$ be a slightly curved sequence with error $O(1)$. Then the graph of $\{a(n)\}_{n\in A}$ contains arbitrarily long APs.*

Proof. Let $k \geq 3$ be an integer. There exists a twice differentiable function $f : \mathbb{R}^+ \to \mathbb{R}$ satisfying (1) and $a(n) = f(n) + O(1)$. Defining $R(n) = a(n) - f(n)$, we can take an integer M satisfying $|R(n)| < M$. Put $r = 4kM$. Then (C) implies that there exist a function h_α satisfying (H1)–(H3) and a strictly increasing sequence $\{x_n\}_{n=1}^{\infty} \subset \mathbb{R}^+$ which diverges such that for every integer $n > 0$ the set $A \cap (h_\alpha(x_n), h_\alpha(x_n+1)]$ contains an AP of length $W(r,k)$. Let $P(n)$ be such an AP of length $W(r,k)$. Then we define the following sets:

$$\begin{aligned} I_j &= \Big[-M + \frac{2M}{r}j, -M + \frac{2M}{r}(j+1)\Big) \quad (0 \leq j \leq r-1),\\ A_j(n) &= \{\, m \in P(n) \mid R(m) \in I_j \,\} \quad (0 \leq j \leq r-1). \end{aligned}$$

Since the union of all the small sets $A_j(n)$ equals $P(n)$, for some $0 \leq q \leq r-1$ the set $A_q(n)$ contains an arithmetic progression $\tilde{P} = \{b(j)\}_{j=0}^{k-1}$. Here $b(j)$ is expressed as $b(j) = dj + e$ with two integers $d, e > 0$. For every sufficiently large n the inequality $d \leq h_\alpha(x_n+1) - h_\alpha(x_n) \leq h'_\alpha(x_n+1)$ holds due to (H2), and the inequality $e \geq h_\alpha(x_n)$ also holds. Thus every sufficiently large n satisfies

$$\begin{aligned} N_{\tilde{P}}(f) &= \sum_{j=0}^{k-3} \left|\Delta^2[f\circ b](j)\right| = \sum_{j=0}^{k-3} d^2 \left|f''(d(j+\theta_j+\eta_j)+e)\right| \\ &= O\Big(\frac{h'_\alpha(x_n+1)^2}{h_\alpha(x_n)^\alpha}\Big) = o(1), \end{aligned} \tag{7}$$

where the last equality follows from (H3) and the values $\theta_j = \theta_j(d,e)$ and $\eta_j = \eta_j(d,e)$ are real numbers satisfying $\theta_j, \eta_j \in (0,1)$. Moreover, the relation $\tilde{P} \subset A_q(n)$ implies

$$\begin{aligned} N_{\tilde{P}}(R) &= \sum_{j=0}^{k-3} \left|\Delta^2[R\circ b](j)\right| \leq \sum_{j=0}^{k-3} (|\Delta[R\circ b](j+1)| + |\Delta[R\circ b](j)|) \\ &\leq 2(k-2)\frac{2M}{r} = 1 - 2/k. \end{aligned} \tag{8}$$

Hence a sufficiently large integer n satisfies $N_{\tilde{P}}(f) < 2/k$. The inequality $N_{\tilde{P}}(a) < 1$ follows from (7) and (8). Since $N_{\tilde{P}}(a)$ is a non-negative integer, the equation $N_{\tilde{P}}(a) = 0$ follows: $a(\tilde{P})$ is an AP of length k. Therefore, the graph of $\{a(n)\}_{n\in A}$ contains an AP of length k. $\square$

Lemma 10. *If a set $A \subset \mathbb{N}$ has positive upper Banach density, then there exists a function $h = h_\alpha$ (to be independent of α) satisfying (H1)–(H4), and*

$$\limsup_{x\to\infty} \frac{|A \cap (h(x), h(x+1)]|}{|\mathbb{N} \cap (h(x), h(x+1)]|} > 0.$$

Proof. **Step 1.** For each integer $N > 0$, we take a non-negative integer $F(N)$ satisfying

$$\max_{n\geq 0} |A \cap [n+1, n+N]| = |A \cap [F(N)+1, F(N)+N]|$$

and then define the functions $F_{\max} : [0,\infty) \to [0,\infty)$ and $\tilde{F} : [0,\infty) \to [0,\infty)$ as follows:

$$\begin{aligned} F_{\max}(x) &= (N+1-x) \max_{1\leq n\leq N} F(n) + (x-N) \max_{1\leq n\leq N+1} F(n) \\ &\qquad (N \in \mathbb{N} \cup \{0\},\ N \leq x < N+1), \\ \tilde{F}(x) &= xe^x (F_{\max}(x) + 1), \end{aligned}$$

where $\max_{1\leq n\leq x} F(n) = 0$ for all $0 \leq x < 1$. Note that $F_{\max}$ is continuous and satisfies $F_{\max}(N) = \max_{1\leq n\leq N} F(n)$. Thus $\tilde{F}$ is continuous and satisfies the following properties:

1. $F(N) + 1 \leq \tilde{F}(N)$ for all $N \in \mathbb{N}$;
2. $\tilde{F}(x) \geq e^x$ for all $x \geq 1$;
3. $\tilde{F}(2x) \geq 2\tilde{F}(x)$ for all $x \geq 0$;
4. $\tilde{F} : [0,\infty) \to [0,\infty)$ is a homeomorphism and satisfies $\tilde{F}(0) = 0$ and $\lim_{x\to\infty} \tilde{F}(x) = \infty$.

Using this function, we define the function $h : [0,\infty) \to [0,\infty)$ as

$$h(x) = \int_0^x \tilde{F}^{-2}(y)\, dy,$$

where $\tilde{F}^{-2} := \tilde{F}^{-1} \circ \tilde{F}^{-1}$, and $\tilde{F}^{-1}$ denotes the inverse function of $\tilde{F}$. Since $\tilde{F}^{-2}$ is continuous, the function h is continuously differentiable. Also, h satisfies (H1) and (H2).
Step 2. To prove that h satisfies (H3), we show the following facts:

1. $h(2x) \leq 4h(x)$ for all $x \geq 0$;
2. $h(x) \leq x \log x$ for every sufficiently large x;
3. $xh'(x) \leq 4h(x)$ for all $x \geq 0$;
4. $h^{-1}(x) \leq x$ for all $x \geq \tilde{F}^2(2)$,

where $\tilde{F}^2 := \tilde{F} \circ \tilde{F}$. First, thanks to property 3 in Step 1, every $y \geq 0$ satisfies $2\tilde{F}^{-1}(y) \geq \tilde{F}^{-1}(2y)$. Therefore, every $x \geq 0$ satisfies

$$h(2x) = \int_0^{2x} \tilde{F}^{-2}(y)\,dy = 2\int_0^x \tilde{F}^{-2}(2y)\,dy \leq 4\int_0^x \tilde{F}^{-2}(y)\,dy = 4h(x),$$

which is just fact 1. Second, thanks to property 2 in Step 1, every $y \geq \tilde{F}(1)$ satisfies $\tilde{F}^{-1}(y) \leq \log y$. Therefore, every sufficiently large x satisfies

$$h(x) = \int_{\tilde{F}^2(1)}^x \tilde{F}^{-2}(y)\,dy + O(1) \leq \int_{\tilde{F}^2(1)}^x \log\log y\,dy + O(1) \leq x\log\log x \leq x\log x,$$

which is just fact 2. Third, fact 3 is verified as follows:

$$xh'(x) \leq \int_x^{2x} h'(y)\,dy \leq h(2x) \leq 4h(x),$$

where the last inequality follows from fact 1. Fourth, every $x \geq \tilde{F}^2(2)$ satisfies

$$h(x) \geq \int_{\tilde{F}^2(2)}^x \tilde{F}^{-2}(y)\,dy \geq \int_{\tilde{F}^2(2)}^x 2\,dy = 2(x - \tilde{F}^2(2)) \geq x,$$

which implies fact 4.

Step 3. We show that h satisfies (H3) for every $\alpha > 0$. Facts 1 and 3 and (H2) imply

$$\frac{h'(x+1)^2}{h(x)^\alpha} \leq \frac{h'(2x)^2}{h(x)^\alpha} \leq \left(\frac{4h(2x)}{2x}\right)^2 \frac{1}{h(x)^\alpha} \leq \left(\frac{8h(x)}{x}\right)^2 \frac{1}{h(x)^\alpha} = \frac{8^2 h(x)^{2-\alpha}}{x^2}. \tag{9}$$

If $\alpha \geq 2$, the right-hand side in (9) goes to zero as $x \to \infty$. If $0 < \alpha < 2$, fact 2 in Step 2 implies

$$\frac{h(x)^{2-\alpha}}{x^2} \leq \frac{(x\log x)^{2-\alpha}}{x^2} = \frac{(\log x)^{2-\alpha}}{x^\alpha} \xrightarrow{x\to\infty} 0.$$

Therefore, the function h satisfies (H3).

Step 4. Assuming

$$\limsup_{x\to\infty} \frac{|A \cap (h(x), h(x+1)]|}{|\mathbb{N} \cap (h(x), h(x+1)]|} = 0,$$

we now deduce a contradiction. This assumption implies that for every $0 < \varepsilon < 1/4$ there exists $x_0 > 0$ such that every $x \geq x_0$ satisfies

$$|A \cap (h(x), h(x+1)]| \leq \varepsilon |\mathbb{N} \cap (h(x), h(x+1)]|.$$

We consider only the case when N is sufficiently large below. If $F(N)+1 \le h(x_0)$, letting $l = \lceil h^{-1}(F(N)+N) - x_0 \rceil$, we have

$$\begin{aligned}
&|A \cap [F(N)+1, F(N)+N]| \\
&\le |A \cap [F(N)+1, h(x_0)]| + \sum_{j=1}^{l} |A \cap (h(x_0+j-1), h(x_0+j)]| \\
&\le h(x_0) - F(N) + \varepsilon |\mathbb{N} \cap (h(x_0), h(x_0+l)]| \le \varepsilon h(x_0+l) - F(N) + O(1) \\
&\le \varepsilon h\big(h^{-1}(F(N)+N)+1\big) - F(N) + O(1) \le \varepsilon h\big(2h^{-1}(F(N)+N)\big) - F(N) + O(1) \\
&\overset{\text{(i)}}{\le} 4\varepsilon(F(N)+N) - F(N) + O(1) = 4\varepsilon N - (1-4\varepsilon)F(N) + O(1) \le 4\varepsilon N + O(1),
\end{aligned} \tag{10}$$

where (i) follows from fact 1 in Step 2. If $F(N)+1 > h(x_0)$, taking a positive number x_1 and a positive integer l with $h(x_1) = F(N)+N$ and $h(x_1-l) < F(N)+1 \le h(x_1-l+1)$, we have

$$\begin{aligned}
&|A \cap [F(N)+1, F(N)+N]| \le |A \cap (h(x_1-l), h(x_1)]| \\
&\le \sum_{j=1}^{l} |A \cap (h(x_1-j), h(x_1-j+1)]| \\
&\le \varepsilon |\mathbb{N} \cap (h(x_1-l), h(x_1)]| \le \varepsilon \{h(x_1) - h(x_1-l)\} \\
&\le \varepsilon \big\{F(N)+N - h\big(h^{-1}(F(N)+1)-1\big)\big\} \\
&\overset{\text{(ii)}}{\le} \varepsilon \big\{F(N)+N - (F(N)+1) + h'\big(h^{-1}(F(N)+1)\big)\big\} \\
&\le \varepsilon \big\{N + h'\big(h^{-1}(F(N)+1)\big)\big\} \overset{\text{(iii)}}{=} \varepsilon N + o(N).
\end{aligned} \tag{11}$$

We verify the above (ii) and (iii). First, the mean value theorem and (H2) imply

$$h(x-1) \ge h(x) - h'(x)$$

for every $x \ge 1$. Putting $x = h^{-1}(F(N)+1)$, we obtain the inequality (ii). Next, in order to prove (iii), we show

$$\lim_{x\to\infty} \frac{(h' \circ h^{-1} \circ \tilde{F} \circ h')(x)}{h'(x)} = 0. \tag{12}$$

Using $h'(x) = \tilde{F}^{-2}(x)$, we obtain

$$\begin{aligned}
\frac{(h' \circ h^{-1} \circ \tilde{F} \circ h')(x)}{h'(x)} &= \frac{(h' \circ h^{-1} \circ \tilde{F}^{-1})(x)}{h'(x)} \\
\overset{\text{(iv)}}{\le} \frac{(h' \circ \tilde{F}^{-1})(x)}{h'(x)} &= \frac{\tilde{F}^{-3}(x)}{\tilde{F}^{-2}(x)} \overset{\text{(v)}}{\le} \frac{\log \tilde{F}^{-2}(x)}{\tilde{F}^{-2}(x)} \xrightarrow{x\to\infty} 0,
\end{aligned}$$

where (iv) and (v) follow from fact 4 in Step 2 and property 2 in Step 1, respectively. From property 1 in Step 1, (H2), and (12), it follows immediately that

$$\frac{h'\bigl(h^{-1}(F(N)+1)\bigr)}{N} \leq \frac{h'\bigl(h^{-1}(\tilde{F}(N))\bigr)}{N} \xrightarrow{N\to\infty} 0,$$

which is just (iii).

Summarizing the above two cases (10) and (11), we have

$$\limsup_{N\to\infty} \frac{\max_{n\geq 0}|A\cap[n+1,n+N]|}{N} = \limsup_{N\to\infty} \frac{|A\cap[F(N)+1,F(N)+N]|}{N} \leq 4\varepsilon.$$

Since $0<\varepsilon<1/4$ is arbitrary, the set A does not have positive upper Banach density, which is a contradiction. □

Proof of Theorem 4. Thanks to Proposition 9, it is enough to show that every set $A\subset\mathbb{N}$ with positive upper Banach density satisfies (C). Let $k\geq 3$ be an integer. Lemma 10 implies that there exist a function $h=h_\alpha$ (to be independent of α) satisfying (H1)–(H4) and a positive number $\delta\leq 1$ such that

$$\limsup_{x\to\infty} \frac{|A\cap(h(x),h(x+1)]|}{|\mathbb{N}\cap(h(x),h(x+1)]|} = \delta.$$

This equation implies that there exists a strictly increasing sequence $\{x_n\}_{n=1}^{\infty}\subset\mathbb{R}^+$ which diverges such that every integer $n>0$ satisfies

$$|A\cap(h(x_n),h(x_n+1)]| \geq \frac{\delta}{2}|\mathbb{N}\cap(h(x_n),h(x_n+1)]|.$$

We can take an integer $n_0>0$ satisfying $h'(x_{n_0})\geq N(\delta/2,k)$ due to (H4). Since the mean value theorem implies $h(x_n+1)-h(x_n)\geq h'(x_n)$ for all $n\in\mathbb{N}$, every integer $n\geq n_0$ satisfies

$$h(x_n+1)-h(x_n)\geq h'(x_n)\geq h'(x_{n_0})\geq N(\delta/2,k).$$

Thus Proposition 1 implies that the set $A\cap(h(x_n),h(x_n+1)]$ with $n\geq n_0$ contains an AP of length k. Therefore, every $A\subset\mathbb{N}$ with positive upper Banach density satisfies (C). □

6 Future work

Question 11. Suppose that a twice differentiable function $f:\mathbb{R}^+\to\mathbb{R}$ satisfies (1). If a strictly increasing sequence $\{a(n)\}_{n=1}^{\infty}\subset\mathbb{N}$ can be written as

$$a(n)=f(n)+O(f'(n)),$$

then does the sequence $\{a(n)\}_{n=1}^{\infty}$ contain arbitrarily long arithmetic progressions?

We do not know the answer to this question, but it is affirmative when the coefficients of the error term $O(f'(n))$ are contained in a finite set $\{c_1, c_2, \dots, c_m\}$. See Appendix C. Hence Question 11 is probably affirmative. As a question related to Question 11, we are interested in how large the error term can be taken. In particular, we do not know whether the error term $O(f'(n))$ is best or not.

Next, we remark that the sequence of all primes is a slightly curved sequence: the asymptotic expansion [17]

$$p_n = f(n) + o\Big(\frac{n}{\log n}\Big)$$

holds. Here p_n is the n-th prime and the function $f(x)$ is

$$x\{\log x + \log\log x - 1 + (\log\log x - 2)/\log x\},$$

which satisfies $f''(x) = O(1/x)$. Thus $\{p_n\}_{n=1}^{\infty}$ is a slightly curved sequence with error $o(n/\log n)$. If we can improve the error term $O((\log\log n)^{1/c_k})$ in Theorem 2 to $o(n/\log n)$, the set of all primes contains arbitrarily long arithmetic progressions, which was shown by Green and Tao [8]. If assuming the Riemann hypothesis, the evaluation

$$|p_n - \mathrm{li}^{-1}(n)| \leq \frac{1}{\pi}\sqrt{n}(\log n)^{5/2}$$

holds [18, Theorem 6.1], where $\mathrm{li}^{-1}(x)$ is the inverse function of the logarithmic integral function $\mathrm{li}(x)$. Thus $\{p_n\}_{n=1}^{\infty}$ is a slightly curved sequence with error $O(\sqrt{n}(\log n)^{5/2})$ if assuming the Riemann hypothesis.

Question 12. Does the set of all primes satisfy (C)? In particular, does the graph of $\{a(p)\}_{p\ \text{prime}}$ contains arbitrarily long arithmetic progressions if $\{a(n)\}_{n=1}^{\infty}$ is a slightly curved sequence with error $O(1)$?

Since the set of all primes does not have positive upper Banach density [12, Corollary 3.4], we can not apply Theorem 4 to $\{a(p)\}_{p\ \text{prime}}$. Nevertheless, we can answer this question if replacing p with $\mathrm{li}^{-1}(n) + O(1)$. See Theorem 17.

Question 13. Is it true that

$$\sup\{\, a \geq 1 \mid \text{the sequence } \{\lfloor n^a \rfloor\}_{n=1}^{\infty} \text{ contains arbitrarily long APs} \,\} = 2?$$

Instead of answering this question, we show that the *graph* of $\{\lfloor n^a \rfloor\}_{n=1}^{\infty}$ with $a \geq 1$ contains an AP of length 3 if and only if $1 \leq a < 2$. The if part follows from Corollary 6. We show the only if part, i.e., the graph of $\{\lfloor n^a \rfloor\}_{n=1}^{\infty}$ with $a \geq 2$ does not contain any APs of length 3 by contradiction. Suppose that a sequence $\{\lfloor (e+dj)^a \rfloor\}_{j=0}^{2}$ is an AP for some two integers $d, e > 0$ and some $a \geq 2$. Then the inequality $|(e+2d)^a + e^a - 2(e+d)^a| < 2$ holds. The mean value theorem implies that

$$(e+2d)^a + e^a - 2(e+d)^a = d^2 a(a-1)(e + d\theta + d\eta)^{a-2},$$

where θ and η are real numbers satisfying $\theta, \eta \in (0,1)$. Thus it follows that

$$2 > |(e+2d)^a + e^a - 2(e+d)^a| = d^2 a(a-1)(e + d\theta + d\eta)^{a-2} \geq 2,$$

which is a contradiction.

The above argument implies

$$\sup \{\, a \geq 1 \mid \text{the graph of } \{\lfloor n^a \rfloor\}_{n=1}^{\infty} \text{ contains an AP of length } 3 \,\} = 2,$$

but we do not achieve the answer to Question 13.

7 Acknowledgments

KS is financially supported by Yoshida Scholarship Foundation. YY is grateful to Seunghoan Song for providing many helpful comments for this paper.

A Slightly curved sequences without the assumption $f''(x) = O(1/x^{\alpha})$

As stated in Section 1, Theorem 2 requires the assumption $f''(x) = O(1/x^{\alpha})$ for some $\alpha > 0$. This appendix addresses a slightly curved sequence with error $O(1)$ without this assumption. First, the following theorem holds:

Theorem 14. *Let f be a twice differentiable function satisfying $\lim_{x\to\infty} f''(x) = 0$ and $\{a(n)\}_{n=1}^{\infty} \subset \mathbb{N}$ be a strictly increasing sequence satisfying $a(n) = f(n) + O(1)$. Then the graph of $\{a(n)\}_{n=1}^{\infty}$ contains arbitrarily long APs.*

Of course, Theorem 2 does not completely contain Theorem 14. For example, the function $f(x) = \int_2^x \mathrm{Li}(t)\, dt$ satisfies the assumption in Theorem 14 but does not satisfy the assumption in Theorem 2 because $f''(x) = 1/\log x$. Here $\mathrm{Li}(t)$ is the offset logarithmic integral function, i.e., $\mathrm{Li}(t) := \int_2^t (1/\log s)\, ds$. Theorem 14 is derived from Theorem 15 below immediately, which is more exact. For example, when length k is given, Theorem 15 implies that the graph of $\{\lfloor c_2 n^2 + c_1 n + c_0 \rfloor\}_{n=1}^{\infty}$ with a sufficiently small $c_2 > 0$ contains an AP of length k, but Theorem 14 does not imply this statement.

Theorem 15. *Let $k \geq 3$ and $r \geq 1$ be integers, f be a twice differentiable function, $R(n)$ be a bounded function satisfying $M_1 \leq R(n) \leq M_2$ for some two real numbers M_1 and M_2, and $\{a(n)\}_{n=1}^{\infty} \subset \mathbb{N}$ be a strictly increasing sequence satisfying $a(n) = f(n) + R(n)$. If the inequality*

$$\limsup_{x\to\infty} |f''(x)| < \left(\frac{k-1}{W(r,k)-1}\right)^2 \left(\frac{1}{k-2} - \frac{2(M_2 - M_1)}{r}\right) \tag{13}$$

holds, then the graph of $\{a(n)\}_{n=1}^{\infty}$ contains an AP of length k.

In general, the van der Waerden number $W(r,k)$ is large and thus the effect of the factor $W(r,k)$ is larger than that of the other factors of the right-hand side in (13). Hence $W(r,k)$ should be small in order to make the right-hand side in (13) large. To make $W(r,k)$ smallest, we should make r smallest. If the right-hand side in (13) is positive, the smallest integer $r > 0$ is

$$r = \lfloor 2(M_2 - M_1)(k-2) \rfloor + 1 > 2(M_2 - M_1)(k-2). \tag{14}$$

Thus the integer r with (14) makes $W(r,k)$ smallest. Let us use (14). When $M_2 - M_1 = 1$ and $k = 3$, the right-hand side in (13) with (14) equals $1/(3 \cdot 13^2)$ because $r = 3$ and $W(3,3) = 27$ [19].

Proof. Thanks to (13), there exists an integer $n_0 > 0$ such that every $x \geq n_0$ satisfies

$$|f''(x)| < \left(\frac{k-1}{W-1}\right)^2 \left(\frac{1}{k-2} - \frac{2M}{r}\right),$$

where $W = W(r,k)$ and $M = M_2 - M_1$. Then we define the following sets:

$$I_j = \left[M_1 + \frac{M}{r}j, M_1 + \frac{M}{r}(j+1)\right] \quad (0 \leq j \leq r-1),$$
$$A = \mathbb{N} \cap [n_0, n_0 + W - 1],$$
$$A_j = \{\, n \in A \mid R(n) \in I_j \,\} \quad (0 \leq j \leq r-1).$$

It can be easily checked that the interval $[M_1, M_2]$ is the union of all the small intervals I_j and the set A is the union of all the small sets A_j. Hence, there exists an integer $0 \leq q \leq r-1$ such that A_q contains an arithmetic progression $P = \{b(j)\}_{j=0}^{k-1}$. Here $b(j)$ is expressed as $b(j) = dj + e$ with two integers $d, e > 0$. The inequalities $(k-1)d \leq W - 1$ and $e \geq n_0$ imply

$$\begin{aligned} N_P(f) &= \sum_{j=0}^{k-3} \left|\Delta^2[f \circ b](j)\right| = \sum_{j=0}^{k-3} d^2 \left|f''(d(j + \theta_j + \eta_j) + e)\right| \\ &< \sum_{j=0}^{k-3} \left(\frac{W-1}{k-1}\right)^2 \left(\frac{k-1}{W-1}\right)^2 \left(\frac{1}{k-2} - \frac{2M}{r}\right) = 1 - 2(k-2)\frac{M}{r}, \end{aligned} \tag{15}$$

where $\theta_j = \theta_j(d,e)$ and $\eta_j = \eta_j(d,e)$ are real numbers satisfying $\theta_j, \eta_j \in (0,1)$. Moreover, the relation $P \subset A_q$ implies

$$N_P(R) = \sum_{j=0}^{k-3} \left|\Delta^2[R \circ b](j)\right| \leq \sum_{j=0}^{k-3} (|\Delta[R \circ b](j+1)| + |\Delta[R \circ b](j)|) \leq 2(k-2)\frac{M}{r}. \tag{16}$$

Hence the inequality $N_P(a) < 1$ follows from (15) and (16). Since $N_P(a)$ is a non-negative integer, the equation $N_P(a) = 0$ follows: $a(P)$ is an AP of length k. Therefore, the graph of $\{a(n)\}_{n=1}^{\infty}$ contains an AP of length k. $\square$

Remark 16. As proved above, van der Waerden's theorem implies Theorem 15 and then Theorem 15 implies Theorem 14, i.e., van der Waerden's theorem implies Theorem 14. Surprisingly, the converse also holds, and thus Theorem 14 is equivalent to van der Waerden's theorem. Let us verify the converse. Here it is well-known that van der Waerden's theorem is equivalent to the statement that if the set of all positive integers is partitioned into finitely many sets then at least one set must contain arbitrarily long APs. We now show that Theorem 14 implies this statement. Suppose that $\mathbb{N} = C_1 \cup C_2 \cup \cdots \cup C_r$, where C_i and C_j are disjoint for every $1 \le i < j \le r$. We define $R(n) = j$ if $n \in C_j$ and also define $a(n) = rn + R(n)$ for all $n \in \mathbb{N}$. Then $\{a(n)\}_{n=1}^{\infty}$ is a strictly increasing sequence. Due to Theorem 14, the graph of $\{a(n)\}_{n=1}^{\infty}$ contains arbitrarily long APs. Fix an arbitrary integer $k > r$. Then there exists an arithmetic progression $P = \{b(j)\}_{j=0}^{k-1}$ such that

$$0 = N_P(a) = \sum_{j=0}^{k-3} |\Delta^2[a \circ b](j)| = \sum_{j=0}^{k-3} |\Delta^2[R \circ b](j)| = N_P(R).$$

Since the inequalities $1 \le R(n) \le r$ and $k > r$ hold, the set $R(P)$ is not an AP of length k: $R(P) = \{j(k)\}$ for some $1 \le j(k) \le r$, which implies that $C_{j(k)}$ contains an AP of length k. From the finiteness of the range $1 \le j(k) \le r$, there exist $1 \le j_0 \le r$ and a sequence $k_1 < k_2 < \cdots$ such that

$$C_{j_0} = C_{j(k_1)} = C_{j(k_2)} = \cdots$$

Note that j_0 does not depend on k. Thus C_{j_0} contains arbitrarily long APs. From the above discussion, Theorem 14 is equivalent to van der Waerden's theorem.

We use the proof of Theorem 15 to prove the next theorem. As stated in Section 6, we cannot answer Question 12, but the following theorem holds:

Theorem 17. *Let $\{a(n)\}_{n=1}^{\infty}$ be a slightly curved sequences with error $O(1)$ and g be a function satisfying the following properties:*

(G1) *g is a twice differentiable function defined on the interval (x_0, ∞) for some $x_0 > 0$;*

(G2) *there exists a positive number $x_1 \ge x_0$ such that g and g' increase on the interval (x_1, ∞);*

(G3) $\lim_{x\to\infty} g'(2x)/g(x)^{\alpha} = 0$ *for all $\alpha > 0$;*

(G4) $\lim_{x\to\infty} g''(x) = 0$.

If $A = \{\tilde{a}(n)\}_{n=n_0}^{\infty}$ with $n_0 > x_1$ is a strictly increasing sequence satisfying $\tilde{a}(n) = g(n) + O(1)$, then the graph of $\{a(n)\}_{n\in A}$ contains arbitrarily long APs.

Proof. Thanks to Proposition 9, it is enough to show that the set $A = \{\tilde{a}(n)\}_{n=n_0}^{\infty}$ satisfies (C). Let $k \ge 3$ be an integer and α be a positive number. We choose the function h_α as $h_\alpha(x) = x^{\beta}$ with $\beta = 1 + \alpha/2$, which satisfies (H1)–(H4). Hence all we need is to show that there exists a strictly increasing sequence $\{x_n\}_{n=1}^{\infty} \subset \mathbb{R}^+$ which diverges such that for every $n > 0$ the set $A \cap (h_\alpha(x_n), h_\alpha(x_n+1)]$ contains an AP of length k.

Take two real numbers M_1 and M_2 satisfying $M_1 \leq \tilde{a}(n) - g(n) \leq M_2$ and the integer r with (14), and let $W = W(r,k)$ and $M = M_2 - M_1$. Since g satisfies (G4), Theorem 15 can be applied to $A = \{\tilde{a}(n)\}_{n=n_0}^{\infty}$. More precisely, the proof of Theorem 15 implies that there exists an integer $n_1 \geq n_0$ such that for every $n \geq n_1$ the set $A \cap [\tilde{a}(n), \tilde{a}(n+W-1)]$ contains an AP of length k. Thus (C) can be reduced to the inclusion relation

$$A \cap (h_\alpha(x_n), h_\alpha(x_n+1)] \supset A \cap [\tilde{a}(n), \tilde{a}(n+W-1)]. \tag{17}$$

Let us show (17). Put $x_n = (g(n) + M_1 - 1)^{1/\beta} < \tilde{a}(n)^{1/\beta}$. Then $h_\alpha(x_n) = g(n) + M_1 - 1 < \tilde{a}(n)$. The mean value theorem implies

$$\begin{aligned}
\tilde{a}(n+W-1) &\leq g(n+W-1) + M_2 \leq g(n) + (W-1)g'(n+W-1) + M_2 \\
&= h_\alpha(x_n) - (M_1 - 1) + (W-1)g'(n+W-1) + M_2 \\
&= h_\alpha(x_n) + (W-1)g'(n+W-1) + M + 1, \qquad (18)\\
h_\alpha(x_n+1) &\geq h_\alpha(x_n) + h'_\alpha(x) = h_\alpha(x_n) + \beta x_n^{\beta-1} \qquad (19)
\end{aligned}$$

due to (G2). Using (G2), (G3), and $\lim_{x\to\infty} g(x) = \infty$ (this limit follows from the fact that $\{\tilde{a}(n)\}_{n_0}^{\infty}$ is a strictly increasing sequence), we have

$$\lim_{x\to\infty} \frac{(W-1)g'(x+W-1) + M + 1}{\beta(g(x) + M_1 - 1)^{1-1/\beta}} \leq \lim_{x\to\infty} \frac{(W-1)g'(2x) + M + 1}{(g(x)/2)^{1-1/\beta}} = 0,$$

whence $(W-1)g'(n+W-1) + M + 1 = o(\beta x_n^{\beta-1})$. Thus there exists an integer $n_2 \geq n_1$ such that every $n \geq n_2$ satisfies

$$(W-1)g'(n+W-1) + M + 1 \leq \beta x_n^{\beta-1}. \tag{20}$$

The equations (18), (19), and (20) yield $h_\alpha(x_n+1) \geq \tilde{a}(n+W-1)$ for every $n \geq n_2$. Therefore, the inclusion relation (17) holds for every $n \geq n_2$. □

Remark 18. The function $g(x) = \mathrm{li}^{-1}(x)$ satisfies (G1)–(G4). (The definition of $\mathrm{li}^{-1}(x)$ is in Section 6.) Moreover, when $\tilde{a}(n) = \mathrm{li}^{-1}(n) + O(1)$, the sequence $A = \{\tilde{a}(n)\}_{n=n_0}^{\infty}$ for some n_0 is a strictly increasing sequence. Therefore, if $\{a(n)\}_{n=1}^{\infty}$ is a slightly curved sequence with error $O(1)$, the graph of $\{a(n)\}_{n\in A}$ contains arbitrarily long APs.

B Result similar to Theorem 15

This appendix states a result similar to Theorem 15, which is not completely contained by Theorem 15. Indeed, Theorem 21 below implies that the graph of $\{\lfloor c_2 n^2 + c_1 n + c_0 \rfloor\}_{n=1}^{\infty}$ with $0 < c_2 < 1/18$ contains an AP of length 4, which is better evaluation than that of Theorem 15.

	$\Sigma = \{0,1\}$	$\Sigma = \{0,1,2\}$	$\Sigma = \{0,1,2,3\}$	$\Sigma = \{0,1,3,4\}$
$k = 2$	$n = 4$	$n = 8$	$n = 51$	
$k = 3$	$n = 10$			(S_3) does not hold [20].

Table 1: Integers n in the statement (S_k) for several alphabets Σ. The case when $k = 2$ and $\Sigma = \{0,1,2,3\}$ was computed by a personal computer. If the statement (S_2) for $\Sigma = \{0,1,2,3,4\}$ holds, then $n > 152$ by computing. If the statement (S_3) for $\Sigma = \{0,1,2\}$ holds, then $n > 1288$ [20].

Theorem 19. *If a function $f : \mathbb{N} \to \mathbb{R}^+$ satisfies $\Delta f > 0$, $\Delta^2 f \geq 0$, and*

$$\sup_{x>0} N(x)^2/x = \infty, \tag{21}$$

then the graph of $\{\lfloor f(n) \rfloor\}_{n=1}^{\infty}$ contains an AP of length 4. Here $N(x)$ denotes the number of positive integers satisfying $f(n) \leq x$, i.e., $N(x) := |\{\, n \in \mathbb{N} \mid f(n) \leq x \,\}|$.

Remark 20. Theorem 19 does not assume that the sequence $\{\lfloor f(n) \rfloor\}_{n=1}^{\infty}$ strictly increases, but when the function f satisfies $\Delta f(n_0) \geq 1$ for some $n_0 > 0$, the sequence $\{\lfloor f(n) \rfloor\}_{n=n_0}^{\infty}$ strictly increases due to the assumption $\Delta^2 f \geq 0$.

Let Σ be a nonempty finite set of non-negative integers. In order to show Theorem 19, we address *words* over the *alphabet* Σ. Here we allow words to continue infinitely on the right side such as $000\cdots$, and to be the empty word. The length of a word w (i.e., the number of all *letters* of w) is denoted by $|w|$ and the sum of all letters of a word w is denoted by $\sum w$. Then we focus on the following condition (C_k) for a word w: *there exist $k+1$ finite length words $w_0, \ldots, w_k$ and a word w_{k+1} such that*

$$w = w_0 \cdots w_k w_{k+1},\ |w_1| = \cdots = |w_k| > 0, \text{ and } \sum w_1 = \cdots = \sum w_k.$$

We also focus on the following statement: *there exists an integer $n > 0$ such that every word w with $|w| \geq n$ satisfies the condition (C_k).* We call it the statement (S_k). The statement (S_k) depends on the alphabet Σ. Table 1 summarizes integers n in the statements (S_2) and (S_3) for several alphabets Σ. The proof of Theorem 19 uses the statement (S_3). The statement (S_3) for $\Sigma = \{0,1\}$ can be checked by taking $n = 10$. Cassaigne et al. [20] proved that the statement (S_3) did not hold for $\Sigma = \{0,1,3,4\}$. For details, see [20].

Proof of Theorem 19. We show that for every integer $k > 0$ there exists an integer $n_0 > 0$ such that the function $\lfloor \Delta f(n) \rfloor$ takes a constant value for every $n_0 \leq n \leq n_0 + k$, by contradiction. Suppose that our assertion does not hold, namely, for some integer $k > 0$, there exists no positive integer n_0 such that the function $\lfloor \Delta f(n) \rfloor$ takes a constant value for every $n_0 \leq n \leq n_0 + k$. Putting the function $\lambda(x) := \lceil N(x)/k \rceil - 1$, we find $N(x) = kN(x)/k > k\lambda(x)$. Hence the inequality $N(x) \geq k\lambda(x) + 1$ holds. Then the assumption

$\Delta^2 f(n) \geq 0$ yields $\lfloor \Delta f(n+1) \rfloor \geq \lfloor \Delta f(n) \rfloor$. Using this inequality and the assumption of the proof by contradiction to obtain the following (i), we have

$$\begin{aligned} x > f(k\lambda(x)+1) - f(1) &= \sum_{j=1}^{k\lambda(x)} \Delta f(j) \geq \sum_{j=1}^{k\lambda(x)} \lfloor \Delta f(j) \rfloor \overset{\text{(i)}}{\geq} k \sum_{j=0}^{\lambda(x)-1} j \\ &= \frac{k}{2}\lambda(x)(\lambda(x)-1) \geq \frac{k}{2}\Big(\frac{N(x)}{k} - 1\Big)\Big(\frac{N(x)}{k} - 2\Big), \end{aligned}$$

that is, $(N(x)-k)(N(x)-2k) < 2kx$. This inequality implies that $N(x)^2/x$ is upper bounded by a constant, which contradicts the assumption (21).

Putting $k = 9$, we can take two integers $n_0 > 0$ and $c \geq 0$ satisfying $\lfloor \Delta f(n) \rfloor = c$ for all $n_0 \leq n \leq n_0 + 9$. Thus $\Delta \lfloor f(n) \rfloor - c \in \{0, 1\}$ for all $n_0 \leq n \leq n_0 + 9$. Since the statement (S_3) holds for $\Sigma = \{0, 1\}$, there exist two integers $d, e > 0$ such that

$$\sum_{j=e}^{e+d-1} (\Delta \lfloor f(j) \rfloor - c) = \sum_{j=e+d}^{e+2d-1} (\Delta \lfloor f(j) \rfloor - c) = \sum_{j=e+2d}^{e+3d-1} (\Delta \lfloor f(j) \rfloor - c).$$

Thus the equation

$$\lfloor f(e+d) \rfloor - \lfloor f(e) \rfloor = \lfloor f(e+2d) \rfloor - \lfloor f(e+d) \rfloor = \lfloor f(e+3d) \rfloor - \lfloor f(e+2d) \rfloor$$

holds: $\{\lfloor f(e+dj) \rfloor\}_{j=0}^{3}$ is an AP of length 4. That is, the graph of $\{\lfloor f(n) \rfloor\}_{n=0}^{\infty}$ contains an AP of length 4. $\square$

Theorem 21. *If a function $f : \mathbb{N} \to \mathbb{R}^+$ satisfies $\Delta f > 0$, $\Delta^2 f \geq 0$, and*

$$\limsup_{x \to \infty} N(x)^2/x > 18, \tag{22}$$

then the graph of $\{\lfloor f(n) \rfloor\}_{n=1}^{\infty}$ contains an AP of length 4.

Proof. If the function $N(x)^2/x$ is not bounded, our assertion follows from Theorem 19. Hence we assume that the function $N(x)^2/x$ is bounded. We show that there exists an integer $n_0 > 0$ such that the function $\lfloor \Delta f(n) \rfloor$ takes a constant value for every $n_0 \leq n \leq n_0 + 9$, by contradiction. In the same way as the proof of Theorem 19, the inequality $(N(x) - 9)(N(x) - 18) < 18x$ holds. Moreover, since the function $N(x)^2/x$ is bounded, the limit

$$\frac{N(x)}{x} = \frac{N(x)}{x^{1/2}} \frac{1}{x^{1/2}} \xrightarrow{x \to \infty} 0$$

holds. Thus the inequality $\limsup_{x \to \infty} N(x)^2/x \leq 18$ follows, which contradicts the assumption (22). Therefore, there exists two integers n_0 and c such that $\lfloor \Delta f(n) \rfloor = c$ for every $n_0 \leq n \leq n_0 + 9$. The remaining can be shown in the same way as the proof of Theorem 19. $\square$

Remark 22. Theorem 21 implies that the graph of $\{\lfloor c_2 n^2 + c_1 n + c_0 \rfloor\}_{n=1}^{\infty}$ with $0 < c_2 < 1/18$ contains an AP of length 4. However, Theorem 15 only implies that the same graph with $0 < c_2 < 1/(6 \cdot 13^2)$ contains an AP of length 3.

The following propositions provide alternative conditions of (21) and (22).

Proposition 23. *If a function $f : \mathbb{N} \to \mathbb{R}^+$ satisfies $\Delta f > 0$ and $\Delta^2 f \geq 0$, then for every $c > 0$ the following three conditions are equivalent:*

$$\text{(i) } \limsup_{x\to\infty} N(x)^2/x > c, \text{ (ii) } \limsup_{n\to\infty} n^2/f(n) > c, \text{ and (iii) } \liminf_{n\to\infty} f(n)/n^2 < 1/c.$$

Proof. First, note that the limit $f(n) \geq f(1) + (n-1)\Delta f(1) \to \infty$ holds as $n \to \infty$. The implication (ii)$\Rightarrow$(i) follows from the definition of $N(x)$. Next, we show the implication (i)$\Rightarrow$(ii) by contradiction. Suppose that the condition (i) holds and the condition (ii) does not hold, i.e., $\limsup_{x\to\infty} N(x)^2/x > c$ and $\limsup_{n\to\infty} n^2/f(n) \leq c$. The assumption $\Delta f > 0$ and the limit $f(n) \to \infty$ imply that for every $x > 0$ we can take only one positive integer $n = n(x)$ satisfying $f(n) \leq x < f(n+1)$. Hence every $x > 0$ satisfies

$$N(x)^2/x < (n+1)^2/f(n) = n^2/f(n) + 2n/f(n) + 1/f(n), \tag{23}$$

where $n = n(x)$. Now, the assumption $\limsup_{n\to\infty} n^2/f(n) \leq c$ yields $\limsup_{n\to\infty} n/f(n) = 0$ and $\limsup_{n\to\infty} 1/f(n) = 0$. Thus, noting $n = n(x) \to \infty$ as $x \to \infty$ and taking the limit in (23) as $x \to \infty$, we have

$$c < \limsup_{x\to\infty} N(x)^2/x \leq \limsup_{n\to\infty} n^2/f(n) \leq c,$$

which is a contradiction. Therefore, the implication (i)$\Rightarrow$(ii) holds. The remaining, i.e., the equivalence (ii)$\Leftrightarrow$(iii) is trivial. $\square$

Proposition 24. *If a function $f : \mathbb{N} \to \mathbb{R}^+$ satisfies $\Delta f > 0$ and $\sum_{n=1}^{\infty} 1/f(n)^s = \infty$ for some $s > 1/2$, then the condition $\sup_{x>0} N(x)^2/x = \infty$ holds.*

Proof. Let ζ be the Riemann zeta function. We show our assertion by contradiction. Suppose that the function $N(x)^2/x$ is bounded. Then the inequality $N(x)^2/x \leq M$ holds for some $M > 0$. Thus the inequality $n^2/f(n) \leq M$ also holds and every $s > 1/2$ satisfies $\sum_{n=1}^{\infty} 1/f(n)^s \leq M\zeta(2s) < \infty$, which is a contradiction. $\square$

C Partial answer to Question 11

This section proves that Question 11 is affirmative when the coefficients of the error term $O(f'(n))$ are contained in a finite set $\{c_1, c_2, \ldots, c_m\}$.

Theorem 25. *Let $f : \mathbb{R}^+ \to \mathbb{R}$ be a twice differentiable function satisfying (1), and $\tilde{R} : \mathbb{N} \to \{c_1, c_2, \dots, c_m\}$ be a bounded function. If a strictly increasing sequence $\{a(n)\}_{n=1}^{\infty} \subset \mathbb{N}$ can be written as*

$$a(n) = f(n) + \tilde{R}(n) f'(n) + O(1),$$

then the sequence $\{a(n)\}_{n=1}^{\infty}$ contains arbitrarily long APs.

Proof. Fix an integer $k \geq 3$. First, the assumption (1) implies

$$a(n) = f(n) + \tilde{R}(n) f'(n) + O(1) = f(n + \tilde{R}(n)) + O(1). \tag{24}$$

We define $\tilde{a}(n) = n + \tilde{R}(n)$, $W = W(m, W(r, k))$, and the following sets:

$$\tilde{A} = \mathbb{N} \cap [\lfloor W^{3/\alpha} \rfloor, \lfloor W^{3/\alpha} \rfloor + W - 1],$$
$$\tilde{A}_j = \{\, n \in \tilde{A} \mid \tilde{R}(n) = c_j \,\} \quad (1 \leq j \leq m).$$

Since the union of all the small sets $\tilde{A}_j$ is $\tilde{A}$, a small set $\tilde{A}_{\tilde{q}}$ contains an arithmetic progression $\tilde{P}$ of length $W(r, k)$. Thus it follows that

$$N_{\tilde{P}}(\tilde{a}) \leq N_{\tilde{P}}(\mathrm{id}_{\mathbb{N}}) + N_{\tilde{P}}(\tilde{R}) = 0.$$

In other words, the sequence $\tilde{a}(\tilde{P})$ (of real numbers) is also an AP of length $W(r, k)$.

Noting that the sequence $\tilde{a}(\tilde{P})$ is a strictly increasing sequence, we define the function $R : \tilde{a}(\tilde{P}) \to \mathbb{R}$ as

$$R(\tilde{a}(n)) = a(n) - f(\tilde{a}(n)) \quad (n \in \tilde{P}). \tag{25}$$

Then the function R is bounded. Thanks to (24), we can take two real numbers M_1 and M_2 satisfying $M_1 \leq R(x) \leq M_2$ such that M_1 and M_2 are independent of r. We define $M = M_2 - M_1$ and the following sets:

$$I_j = \left[M_1 + \frac{M}{r} j, M_1 + \frac{M}{r}(j+1) \right] \quad (0 \leq j \leq r - 1),$$
$$A_j = \{\, x \in \tilde{a}(\tilde{P}) \mid R(n) \in I_j \,\} \quad (0 \leq j \leq r - 1).$$

Since the union of all the small sets A_j is $\tilde{a}(\tilde{P})$, a small set A_q contains an arithmetic progression $P = \{b(j)\}_{j=0}^{k-1}$ of length k. Here $b(j)$ is expressed as $b(j) = dj + e$ with two integers $d, e > 0$. Moreover, the inequalities

$$d \leq \tilde{a}(\lfloor W^{3/\alpha} \rfloor + W - 1) - \tilde{a}(\lfloor W^{3/\alpha} \rfloor) \leq W - 1 + M,$$
$$e \geq \tilde{a}(\lfloor W^{3/\alpha} \rfloor) \geq \lfloor W^{3/\alpha} \rfloor + M_1$$

hold. These inequalities imply that

$$N_P(R) \leq 2(k-2)\frac{M}{r} = o(1),$$
$$N_P(f) = \sum_{j=0}^{k-3} d^2 \left| f''(d(j + \theta_j + \eta_j) + e) \right| = O\Big(\frac{(W - 1 + M)^2}{(\lfloor W^{3/\alpha} \rfloor + M_1)^\alpha} \Big) = O(1/W) = o(1)$$

as $r \to \infty$, where $\theta_j = \theta_j(d, e)$ and $\eta_j = \eta_j(d, e)$ are real numbers satisfying $\theta_j, \eta_j \in (0, 1)$. Thus every sufficiently large r satisfies $N_P(f) + N_P(R) < 1$.

Recall (25). The domains of f and R are the same, but the domain of a is different from them. Hence we must take a set like the inverse image of P under $\tilde{a}$. Now, the restricted function $\tilde{a}|\tilde{P}$ is injective. When we denote by P' the inverse image of $P \subset \tilde{a}(\tilde{P})$ under this restricted function, the sequence P' is also an AP of length k. Then the inequality $N_{P'}(a) \leq N_P(f) + N_P(R) < 1$ holds. Since $N_{P'}(a)$ is a non-negative integer, the equation $N_{P'}(a) = 0$ follows: $a(P')$ is an AP of length k. $\square$

References

[1] E. Szemerédi. On sets of integers containing no k elements in arithmetic progression, *Acta Arith.* **27** (1975), 199–245.

[2] J. Jasinski. Large sets containing copies of small sets, *Real Anal. Exchange* **21** (1995/96), 758–766.

[3] I. I. Piatetski-Shapiro. On the distribution of prime numbers in sequences of the form $[f(n)]$, *Mat. Sbornik N.S.* **33** (75) (1953), 559–566.

[4] J. Rivat and J. Wu. Prime numbers of the form $[n^c]$, *Glasg. Math. J.* **43** (2001), 237–254.

[5] M. Mirek. Roth's theorem in the Piatetski-Shapiro primes, *Rev. Mat. Iberoam.* **31** (2015), 617–656.

[6] H. Li and H. Pan. The Green-Tao theorem for Piatetski-Shapiro primes, preprint, 2019. Available at https://arxiv.org/abs/1901.09372.

[7] P. Erdős and P. Turán. On some sequences of integers, *J. London Math. Soc.* **11** (1936), 261–264.

[8] B. Green and T. Tao. The primes contain arbitrarily long arithmetic progressions, *Ann. of Math. (2)* **167** (2008), 481–547.

[9] L. E. Dickson. *History of the Theory of Numbers. Vol. II: Diophantine Analysis*, Chelsea Publishing Co., 1966.

[10] P. Dénes. Über die Diophantische Gleichung $x^l + y^l = cz^l$, *Acta Math.* **88** (1952), 241–251.

[11] H. Darmon and L. Merel. Winding quotients and some variants of Fermat's last theorem, *J. Reine Angew. Math.* **490** (1997), 81–100.

[12] H. L. Montgomery and R. C. Vaughan. *Multiplicative Number Theory I. Classical Theory*, Cambridge University Press, 2007.

[13] B. L. van der Waerden. Beweis einer Baudetschen Vermutung, *Nieuw Arch. Wisk.* **15** (1927), 212–216.

[14] W. T. Gowers. A new proof of Szemerédi's theorem, *Geom. Funct. Anal.* **11** (2001), 465–588.

[15] T. F. Bloom. A quantitative improvement for Roth's theorem on arithmetic progressions, *J. London Math. Soc.* **93** (2016), 643–663.

[16] B. Green and T. Tao. New bounds for Szemerédi's theorem, III: a polylogarithmic bound for $r_4(N)$, *Mathematika* **63** (2017), 944–1040.

[17] M. Cipolla. La determinazione assintotica dell'n^{imo} numero primo, *Matematiche Napoli* **3** (1902), 132–166.

[18] J. Arias de Reyna and J. Toulisse. The n-th prime asymptotically, *J. Théor. Nombres Bordeaux* **25** (2013), 521–555.

[19] V. Chvátal. Some unknown van der Waerden numbers, in *Combinatorial Structures and Their Applications (Proc. Calgary Internat. Conf., Calgary, Alberta, 1969)*, Science Publishers, 1970, pp. 31–33.

[20] J. Cassaigne, J. D. Currie, L. Schaeffer, and J. Shallit. Avoiding three consecutive blocks of the same size and same sum, *J. ACM* **61** (2014), 10:1-10:17.

2010 *Mathematics Subject Classification*: Primary 11B25; Secondary 11B30.
Keywords: arithmetic progression, Szemerédi's theorem, Piatetski-Shapiro sequence, van der Waerden number, Gowers' upper bound.

Received October 22 2018; revised versions received February 21 2019; February 22 2019. Published in *Journal of Integer Sequences*, February 22 2019.

Return to Journal of Integer Sequences home page.