Journal of Integer Sequences, Vol. 11 (2008),
Article 08.4.2

A Few New Facts about the EKG Sequence

Piotr Hofman and Marcin Pilipczuk
Department of Mathematics, Computer Science and Mechanics
University of Warsaw
Warsaw
Poland
piotr.hofman@students.mimuw.edu.pl
marcin.pilipczuk@students.mimuw.edu.pl

Abstract

The EKG sequence is defined as follows: $a_1 = 1$, $a_2 = 2$ and a_n is the smallest natural number satisfying $\gcd(a_{n-1}, a_n) > 1$ not already in the sequence. The sequence was previously investigated by Lagarias, Rains and Sloane. In particular, we know that (a_n) is a permutation of the natural numbers and that the prime numbers appear in this sequence in an increasing order.

Lagarias, Rains and Sloane performed many numerical experiments on the EKG sequence up to the 10^7th term and came up with several interesting conjectures. This paper provides proofs for the core part of those conjectures. Namely, let (a'_n) be the sequence (a_n) with all terms of the form p and $3p$, for p prime, changed to $2p$. First, we prove that for any odd prime $a_n = p$ we have $a_{n-1} = 2p$. Then we prove that $\lim_{n\to\infty} \frac{a'_n}{n} = 1$, i.e., we have $a_n \sim n$ except for the values of p and $3p$ for p prime: if $a_n = p$ then $a_n \sim \frac{n}{2}$, and if $a_n = 3p$ then $a_n \sim \frac{3n}{2}$.

1 Introduction

The EKG sequence is defined as follows: $a_1 = 1$, $a_2 = 2$ and a_n is the smallest natural number not already in the sequence satisfying $\gcd(a_{n-1}, a_n) > 1$. It was originally defined by Ayres and later investigated by Lagarias, Rains and Sloane [1]. It appears as entry A064413 in the On-line Encyclopedia of Integer Sequences [4]. The first forty terms of the sequence are

1	2	4	6	3	9	12	8	10	5
15	18	14	7	21	24	16	20	22	11
33	27	30	25	35	28	26	13	39	36
32	34	17	51	42	38	19	57	45	40

Lagarias, Rains and Sloane [1] describe some basic properties of this sequence. In particular, they prove that it is a permutation of the natural numbers with linear asymptotic behavior; more precisely $\frac{1}{260}n \leq a_n \leq 14n$. We recall some of their lemmas and theorems in Section 3. Their numerical experiments resulted in several interesting conjectures. First, if $a_n = p$ and p is a prime, then $a_{n-1} = 2p$, which we will prove in Section 4.2. Second, that $a_n \sim n$ except for when $a_n = p$ is prime; then $a_n = p \sim \frac{n}{2}$ and $a_{n+1} = 3p \sim \frac{3n}{2}$.

We will prove the conjectures mentioned above. To formulate what is proved let us introduce some notation. Let (z_n) be the inverse of the permutation (a_n), let (p_i) be the sequence of prime numbers in increasing order, and let (a'_n) be defined as follows: $a'_n = 2p$ if $a_n = 3p$ or $a_n = p$ for p prime, $a'_n = a_n$ otherwise. This definition was introduced by Lagarias, Rains and Sloane [1]. We prove that that $\lim_{n\to\infty} \frac{a'_n}{n} = 1$. More precisely, we will prove that there exists a universal constant C such that for all natural numbers n we have

$$n - \frac{Cn}{\log\log\log n} \leq a_n \leq n + \frac{Cn}{\log\log\log n}.$$

Note that the speed of convergence is extremely slow. In particular, this result is weaker than a more precise conjecture made by Lagarias, Rains and Sloane [1] that $a'_n \sim n(1+\frac{1}{3\log n})$. In Section 5 we discuss some drawbacks and possible improvements of the proofs in this work.

In this work, we will use several strong results from number theory. These facts are gathered in Section 2.

In this work the letters c, C, C', c', C_i, c_i denote constants. If it is not specified otherwise, they are absolute numerical constants.

2 Number theory tools

Let us define p_n to be the n-th prime number and $\pi(n)$ the number of primes smaller than n. According to Rosser [3], the following version of the prime number theorem holds:

Theorem 2.1. *If $n \geq 55$ then*

$$\pi(n) < \frac{n}{\log n - 4}.$$

We conclude that

Corollary 2.2. *If $n > 25000 > e^{10.1}$ then*

$$\pi(n) < \frac{2n}{\log n}$$

and

$$\pi(n) < \frac{n}{6} - 2$$

Let us now estimate the number of different prime divisors of n.

Lemma 2.3. *Any natural number n has at most $C\frac{\log n}{\log\log n}$ different prime divisors.*

Proof. If n has at least k prime divisors, then $n \geq 2 \cdot 3 \cdot \dots \cdot p_k > k!$, so by Stirling's formula $\log n > C_1 k \log k$. Let $f(x) = C_1 x \log x$. Then

$$f\Big(C_2 \frac{\log n}{\log\log n}\Big) = C_1 C_2 \frac{\log n}{\log\log n}(\log C_2 + \log\log n - \log\log\log n) > C_3 C_2 \log n.$$

Therefore $k < C_4 \frac{\log n}{\log\log n}$. □

The following is a well-known fact from calculus.

Theorem 2.4. *Let $p_1, p_2, \dots$ be the sequence of the prime numbers primes in increasing order ($p_1 = 2, p_2 = 3, \dots$). Then*

$$\prod_{n=1}^{\infty} \Big(1 - \frac{1}{p_n}\Big) = 0.$$

However, in our work we will need a stronger result; namely, Merten's theorem [2].

Theorem 2.5. *Let $p_1, p_2, \dots$ be the sequence of prime numbers in increasing order ($p_1 = 2, p_2 = 3, \dots$). Then*

$$\lim_{n\to\infty} (\log p_n) \prod_{n=1}^{\infty} \Big(1 - \frac{1}{p_n}\Big) = e^{-\gamma}$$

where γ is the Euler-Mascheroni constant.

Since $p_n \sim n \log n$ the following is straightforward:

Corollary 2.6. *There exists a universal constant C such that for all natural numbers n the following holds:*

$$\prod_{n=1}^{\infty} \Big(1 - \frac{1}{p_n}\Big) < \frac{C}{\log n}.$$

Note: all results cited above are independent of the Riemann hypothesis.

3 Already known simple facts

As Lagarias, Rains and Sloane [1] defined, a prime dividing both a_{n-1} and a_n is called a *controlling prime* for a_n. Note that a controlling prime is not unique; if not specified otherwise, in proofs we choose any of the controlling primes. We will define $g(n) = \gcd(a_n, a_{n-1})$. As observed by Lagarias, Rains and Sloane [1], we can define the sequence in a different way. For every prime p let $B_n(p)$ be the smallest multiple of p that has not appeared in the first n terms of the sequence. Then a_{n+1} is smallest among all $B_n(p)$ for primes p dividing a_n. The following lemmas and theorems were proved by Lagarias, Rains and Sloane [1] and for some of them we quote the proofs, since they are quite basic and will be important in the next section.

Lemma 3.1. *Let $p > 2$ be a prime. If a_n is the first term divisible by p, then $a_{n+1} = p$ and $a_n = pq$ where q is the smallest prime divisor of a_{n-1}.*

Proof. For any $k \mid a_{n-1}$ the number kp is a good candidate for a_n, so $a_n = qp$, where q is the smallest possible controlling prime, i.e., the smallest prime divisor. Since q is a controlling prime, q, $2q$, $\ldots$, $q(p-1)$ were already used. So if q was a controlling prime for a_{n+1}, then $a_{n+1} \geq q(p+1) > p$, where p is a good candidate. So p will be a controlling prime and $a_{n+1} = p$. $\square$

Lemma 3.2. *The prime numbers appear in a_n in increasing order.*

Proof. If $a_{n+1} = p$, then $a_n = qp$ is the first term divisible by p. For any $p' < p$ we have $qp' < qp = a_n$, so the term qp' appeared in the sequence earlier. $\square$

Lemma 3.3. *If $\{m, 2m, \ldots, km\} \subset \{a_i : 1 \leq i \leq M\}$, then $\{1, 2, \ldots, k\} \subset \{a_i : 1 \leq i \leq M+1\}$.*

Proof. By induction on k. For $k = 1$ it is obvious, since $a_1 = 1$. Let $\{m, 2m, \ldots, km\} \subset \{a_i : 1 \leq i \leq M\}$. Let $a_n = km$, $n \leq M$. Let q be a controlling prime for a_n. If $q \mid m$, then all the numbers $m, 2m, \ldots, (k-1)m$ were good candidates for a_n, so they were before a_n, so, by the induction hypothesis, $\{1, 2, \ldots, k-1\} \subset \{a_i : i \leq n\}$. The number k is a great candidate for a_{n+1}, because all smaller numbers were used, so $\{1, 2, \ldots, k\} \subset \{a_i : i \leq n+1\}$. If $q \mid k$, then k was a candidate for a_n, so it was before a_n. By the induction hypothesis $\{1, 2, \ldots, k-1\} \subset \{a_i : i \leq M+1\}$. This completes the proof. $\square$

Lemma 3.4. *For any prime p, the numbers $1, 2, \ldots p-1$ appear in the EKG sequence before p.*

Proof. As in the proof of Lemma 3.1, before $a_{n+1} = p$ for p prime, there was $a_n = qp$ and all the numbers $q, 2q, \ldots, q(p-1)$ were used before. Therefore, by Lemma 3.3, all the numbers $1, 2, \ldots, p-1$ were before p. $\square$

Theorem 3.5. *For all n the EKG sequence satisfies the following inequality*

$$\frac{1}{260}n < a_n < 14n.$$

Corollary 3.6. *If for some a, b, we have $z_a < z_b$ then $a < 14z_a < 14z_b < 14 \cdot 260b$.*

4 New results

4.1 The fundamental lemma

Most of the proofs in this paper are based on the following observation.

Lemma 4.1. *Fix any real number x. Then for every prime q there can be at most one index n for which q is a controlling prime for a_n and $a_{n-1} < x \leq a_n$.*

Proof. According to the definition, a_n is the smallest multiple of q not yet used in the sequence. Since $a_n \geq x$, all multiples of q smaller than x were already used, and later in the sequence there is no term divisible than q smaller than x, i.e., if $k > n$ and $q \mid a_k$, then $a_k \geq x$. $\square$

This lemma has important consequences; it strongly limits the number of times the sequence can cross the border of x going upwards. Since all the primes in the sequence appear in increasing order, if p_{k+1} has not yet appeared, one can cross the border of x upwards only $k = \pi(p_{k+1}) = \Theta(\frac{p_{k+1}}{\log p_{k+1}})$ times.

4.2 $2p$ is before p

In this section we will prove that each odd new prime in the EKG sequence is introduced by a subsequence $2p$, p, $3p$. We will simply use Lemma 4.1 and the estimate for $\pi(x)$.

We will prove the following theorem:

Theorem 4.2. *For any primes $p, q > 2$ the term qp appears in the sequence after $2p$.*

Proof. We will prove Theorem 4.2 by contradiction. Let us assume that p is the first prime such that for some N we have $a_{N+1} = p$ and $a_N = qp$, where $q > 2$. From Lemmas 3.1 and 3.2 we know that these are the first terms divisible by p and previous terms of the sequence are divisible only by smaller primes than p. From Lemma 4.1 we know that before the term a_N the sequence can cross the border of $2p$ upwards only $\pi(p)$ times — for every prime smaller than p at most once. Now we will show that for large p, the sequence must have crossed this border downwards $\Theta(p)$ times, which is asymptotically bigger than $\pi(p)$. We will try to be tight on the constants, to match the numerical experiments made by Lagarias, Rains and Sloane [1], and therefore the theorem will be proved for the whole sequence.

Since q is the controlling prime for $a_N = qp$, all multiples of q — $\{q, 2q, \ldots, (p-1)q\}$ have appeared before the term a_N. Particularly important for us are numbers greater than $2p$ and divisible by $2q$, because if a_k is such a number for $k < N$ then there is the possibility that $a_{k+1} = 2p$. This cannot hold for $k \le N$, so for every such a_k the next number is smaller than $2p$. Among the terms $\{q, 2q, \ldots, (p-1)q\}$ there are at least $\lfloor \frac{qp-2p}{2q} \rfloor \ge \lfloor \frac{p}{6} \rfloor = \Theta(p)$ numbers both divisible by $2q$ and greater than $2p$. They all appeared in the sequence before a_N. So we have at least $\lfloor \frac{p}{6} \rfloor$ moments before a_N, where the sequence goes downwards through the border $2p$ and, due to Lemma 4.1, at most $\pi(p)$ moments before a_N, where the sequence goes upwards.

From Theorem 2.2 we know that if $p > 25000$ we have $\pi(p) < \lfloor \frac{p}{6} \rfloor - 1$. The number of upward and downward crossing should differ by at most 1, so we have a contradiction for $p > 25000$.

Lagarias, Rains and Sloane [1] computed the first 10 000 000 terms of the sequence. The last prime number not greater than 25 000 is 24 989 and it appears around the 50 000-th term. Up to this bound all primes p were preceded by the term $2p$. Therefore the theorem is proved for the whole sequence. $\square$

From this the following theorem is an obvious corollary:

Theorem 4.3. *If $a_n = p$ and $p > 2$ is prime, then $a_{n-1} = 2p$.*

Proof. From Lemma 3.1 we know that $a_{n-1} = qp$ for some prime q and this is the first term divisible by p. Therefore $q = 2$, since $2p$ could not have appeared before. $\square$

The above idea can be generalized. As in the proof of Theorem 4.2 we can use any natural number n instead of p and any fixed number a instead of 2. This leads to the following theorem, which we do not use later, but which is an interesting result on its own. This theorem could also be deduced from the final theorem — $\lim_{n\to\infty} \frac{a'_n}{n} = 1$, but the proof based on the proof of Theorem 4.2 is much more elementary. We do not include the proof here, since it is very similar to the proof of Theorem 4.2.

Theorem 4.4. *Let $a > 1$ be a natural number. Then there exists a natural number N such that for all natural numbers $b > a$ and $n > N$ the term an appears in the sequence before bn, i.e., $z_{an} < z_{bn}$. In particular, this holds if $\log N > Ca^3$ for some universal constant C.*

Lemma 4.1 can be somewhat enhanced, since we know that $2p$ appears before every odd prime p, and there $3p$ appears afterward. The following lemma shows that the moments when the sequence crosses the border x upwards and it is not the EKG *tick* (i.e., in the subsequence of the form $q, 3q$ for q prime) are very rare.

Lemma 4.5. *Let $x > 1$ be a real number and let B be the set*

$$B = \{n : a_n < x \le a_{n+1};\ a_n \textit{ is not a prime}\}.$$

Then $|B| \le \pi(\sqrt{x})$.

Proof. Let $n \in B$. Since a_n is composite, it has a prime divisor not greater than $\sqrt{a_n}$, which is smaller than $\sqrt{x}$. Let q be any such prime divisor.

Since $a_{n+1} \ge x$ and all multiples of q are candidates for a_{n+1}, all multiples of q smaller than x must have been used before. Therefore, after a_n, there are no free multiples of q smaller than x. So every $n \in B$ *uses* at least one prime q smaller than $\sqrt{x}$ — there are no more unused multiples of q smaller than x, so $|B| \le \pi(\sqrt{x})$. □

4.3 Numbers above the border

By Lemma 4.1, we know that the sequence can cross the border of an integer x upwards at most $\pi(x)$ times. The interesting question now is: before an integer x appears in the sequence, how many terms greater than x can appear?

Now we are giving partial answer to that question. We need to assume that x has got some small divisor, so the numbers greater than x having a common divisor with x (i.e., such terms, after which x is a candidate for the next term) are quite dense.

The following lemma will be used later only for a being a small prime different than 3 and while reading one can think of a as such number — much smaller than n and prime.

Lemma 4.6. *Let $a > 1$ be an integer. Then there are at most $Ca\frac{an}{\log\log an} \le Ca^2\frac{n}{\log\log n}$ terms of the EKG sequence greater than an and appearing before an, where C is a universal constant.*

Proof. Fix $a > 1$ and large n and let $N = z_{an}$, i.e., $a_N = an$. Let $A = \{1 \le i < N : a_i > an\}$; we need to estimate $|A|$. From Lemma 4.1 we know that the sequence can cross the border of an upwards at most $\pi(an)$ times. We will consider several cases.

Let $A_1 = \{i \in A : a \mid a_i\}$. If $i \in A_1$, then an is a candidate for a_{i+1}, so after such i the sequence goes downwards through the border of an if $i+1 < N$. Therefore $|A_1| \le \pi(an)+1$.

Now let $i \in A \setminus A_1$ and let q be a controlling prime for a_i. The idea now is as follows: either a_i is quite close to the border an, or there is an already used multiple of qa below a_i. The formal argument follows.

Note that Corollary 3.6 implies $q \le a_i \le C_0 \cdot an$ for some universal constant C_0.

Let $f(s) = a_i - sq$. Note that $\gcd(q, a) = 1$, because otherwise an would be a better candidate for a_i. For all s (if $f(s) > 0$) $f(s)$ was already used in the sequence before a_i. Among $f(1), f(2), \dots, f(a-1)$ there is exactly one $f(s_0(a_i))$, such that $a \mid f(s_0(a_i))$.

Let $A_2 = \{i \in A \setminus A_1 : f(s_0(a_i)) \le an\}$. For every prime q there are at most $a-1$ such numbers a_i, for which q is a controlling prime and $i \in A_2$ — these are the first at most $a-1$ multiples of q greater than an. Therefore $|A_2| \le (a-1)\pi(C_0 an) \le C_0'(a-1)\pi(an)$.

Let $A_3 = A \setminus A_1 \setminus A_2 = \{i \in A \setminus A_1 : f(s_0(a_i)) > an\}$. The term $f(s_0(a_i))$ was used before a_i, thus for some $j < i$ we have $a_j = f(s_0(a_i))$ and obviously $j \in A_1$. For how many indices i' can we have $f(s_0(a_{i'})) = a_j$? For every prime divisor r of a_j, at most $a-1$ times: $a_{i'}$ can be one of the numbers $a_j + r$, $a_j + 2r$, $\dots$, $a_j + (a-1)r$. According to Lemma 2.3, the number a_j can have at most $C_1 \frac{\log a_j}{\log\log a_j}$ distinct prime divisors.

According to Corollary 3.6 we have $a_j < 14 \cdot 260an$, so a_j can have at most $C_2 \frac{\log(an)}{\log\log(an)}$ distinct prime divisors, where C_2 is a universal constant.

Therefore

$$|A_3| \le |A_1| \cdot (a-1) \cdot C_2 \frac{\log(an)}{\log\log(an)}.$$

Since $\pi(an) \le C \frac{an}{\log(an)}$ for some universal constant C, we have

$$|A| = |A_1| + |A_2| + |A_3| \le \pi(an) + C_0'(a-1)\pi(an) + C_2(a-1)\pi(an)\frac{\log(an)}{\log\log(an)} \le$$

$$\le Ca\frac{an}{\log\log an} \le Ca^2\frac{n}{\log\log n}.$$

□

The following theorem is an obvious corollary:

Theorem 4.7. *There exists a universal constant C such that for every integer $a > 1$ the following holds:* $z_{an} \le an + Ca\frac{an}{\log\log an} \le an + Ca^2\frac{n}{\log\log n}$.

Proof. Fix $a > 1$. From Lemma 4.6 we know that there are at most $Ca\frac{an}{\log\log an}$ indices $i < z_{an}$ such that $a_i > an$. Obviously there are at most an indices $i < z_{an}$ for which $a_i \le an$. Therefore there are at most $an + Ca\frac{an}{\log\log an} = an + o(n)$ indices $i < z_{an}$. □

However, Lemma 4.6 is not sufficient for all our purposes. It limits the number of terms greater than an, but does not limit how big the terms are. The following lemma is an improvement to Lemma 4.6. It uses almost the same technique, but the proof is more complicated and we will use the enhanced version of Lemma 4.1, i.e., Lemma 4.5 and Lemma 4.6 itself.

Lemma 4.8. *Let $a > 1$ be an integer not divisible by 3 and let $\frac{1}{3} > \varepsilon > 0$. Then there exists an integer $N(a)$ such that for all integers $n > N(a)$ if the number $x > (a+2\varepsilon)n$ appeared in the sequence before term an, then x is of the form $3p$ for some prime p. In particular, the sufficient condition is that $\log\log N(a) > \frac{Ca^2}{\varepsilon}$ for some universal constant C.*

Proof. First, due to Lemma 4.6, we know that there are at most $C_1 a \frac{an}{\log\log(an)}$ terms greater than an before the term an appears in the sequence. Choose $N(a)$ so big, that for all $n > N(a)$ we have $C_1 a \frac{an}{\log\log an} < 1 + \frac{\varepsilon n}{3}$, i.e., there are more numbers divisible by 3 in the segment $(an, (a+\varepsilon)n)$ than terms of the sequence greater than an before an appears. Note that this one holds if $\log\log N(a) > \frac{C_2 a^2}{\varepsilon}$.

For such a big n, whenever any term of form $3p$ appears in the sequence before the term an appears, there is an unused divisible by three candidates for the next term smaller than $(a+\varepsilon)n$. So the next term will be smaller than the border of $(a+\varepsilon)n$.

From Lemma 4.5, there are at most $\pi(\sqrt{(a+\varepsilon)n})$ indices i smaller than z_{an}, for which $a_i < (a+\varepsilon)n$, and $a_{i+1} \geq (a+\varepsilon)n$ and a_i is composite. In the other case, if $a_i < (a+\varepsilon)n$ and $a_{i+1} \geq (a+\varepsilon)n$ and a_i is prime, the next term a_{i+2} will be smaller than $(a+\varepsilon)n$.

Suppose there is an index $i < z_{an}$ for which a_i is composite and $a_{i+1} > (a+2\varepsilon)n$. Let i be the first such index. Since $a_i \leq (a+2\varepsilon)n$, a_i has a prime divisor q smaller than $\sqrt{(a+2\varepsilon)n}$. If a_{i+1} is greater than $(a+2\varepsilon)n$, then all multiples of q in the segment $((a+\varepsilon)n, (a+2\varepsilon)n)$ must have been used before. In this segment there are at least $\lfloor \frac{\varepsilon}{aq} n \rfloor$ numbers divisible by both q and a. Let us call such terms *interesting* ones. All *interesting* terms were used before an appears in the sequence. Therefore after such an *interesting* term, number an is good candidate for the next term, so after such terms the sequence goes downwards through the border of $(a+\varepsilon)n$ — the next term will be at most an.

Moreover, since any *interesting* term is divisible by a, it is not of the form of $3p$ for some prime p, since $p \geq \frac{an}{3} > a$ and a is not divisible by 3.

Except for the ticks of the form p, $3p$, the sequence went upwards through the border of $(a+\varepsilon)n$ at most $\pi(\sqrt{(a+\varepsilon)n})$ times, so there can be at most $\pi(\sqrt{(a+\varepsilon)n})$ *interesting* terms, i.e., from the segment $((a+\varepsilon)n, (a+2\varepsilon)n)$ divisible by aq. But for large n we have

$$\frac{\varepsilon}{aq} n \geq \frac{\varepsilon}{a\sqrt{2+2\varepsilon}} \sqrt{n} > \pi(\sqrt{(a+\varepsilon)n}),$$

since $\pi(x) = \Theta(\frac{x}{\log x})$. Note that this inequality holds if $\log n > \frac{C_3 a^{\frac{3}{2}}}{\varepsilon}$ for some universal constant C_3. This condition is much weaker than the bound $\log\log n > \frac{C_2 a^2}{\varepsilon}$ we assumed earlier.

Therefore if $\log\log n > \frac{C_2 a^2}{\varepsilon}$ the only terms greater than $(2+2\varepsilon)n$ are of the form $3p$ for p prime. $\square$

While analyzing the EKG sequence one question appears very often. Let us fix an integer $a > 1$, and let a be not divisible by 3 (in fact, the most interesting case is when a is a prime different than 3). Imagine the term an has appeared in the sequence; the question is, what is the smallest k such that ak has yet not appeared in the sequence (i.e., before an)? Lemma 4.9, which is in fact a simple corollary of Lemma 4.8, provides an answer to that question.

Lemma 4.9. *Let $a > 1$ be an integer not divisible by 3. Let $m_a(n)$ be smallest k, such that term ak has not appeared in the first n terms of the sequence. Let $M_a(n)$ be the biggest k such that term ak has appeared in the first n terms of the sequence. Then*

$$\lim_{n\to\infty} \frac{M_a(n) - m_a(n)}{n} = 0.$$

More precisely, there exists a universal constant C such that the following holds:

$$M_a(n) - m_a(n) \leq \frac{Can}{\log\log n}.$$

Proof. For large n we can apply Lemma 4.8 for $\varepsilon = \frac{C_1 a^2}{\log\log n}$ obtaining that $aM_a(n) \leq (a + 2\varepsilon)m_a(n)$. Keeping in mind that, according to Lagarias, Rains and Sloane [1], the EKG sequence has linear bounds, (i.e., $ci < a_i < Ci$ for some constants c, C, so both $M_a(n) \to \infty$ and $m_a(n) \to \infty$ with $n \to \infty$ and $m_a(n) \leq Cn + a$), and that $\varepsilon \to 0$ as $n \to \infty$, we have for sufficiently large n:

$$M_a(n) - m_a(n) < \frac{2\varepsilon}{a} m_a(n) < \frac{C_1 a^2}{a \log\log n}(Cn + a) < \frac{C_2 an}{\log\log n}.$$

□

4.4 Linear limit for even terms

In this section we will prove that $\lim_{n\to\infty} \frac{2n}{z_{2n}} = 1$, i.e., that an even term $2n$ is at position approximately $2n$. This will give us *a skeleton* for proving that all the terms x except for p and $3p$ are at position approximately x.

From Theorem 4.7 for $a = 2$ we know that term $z_{2n} \leq 2n + C\frac{n}{\log\log n}$, which gives us $\liminf_{n\to\infty} \frac{2n}{z_{2n}} \geq 1$. What we need to prove is that $z_{2n} \geq 2n - o(n)$. We will try to estimate the asymptotic behavior of the $o(n)$ part in the proof.

Theorem 4.10.

$$\limsup_{n\to\infty} \frac{2n}{z_{2n}} \leq 1.$$

Moreover, there exists a universal constant C such that

$$z_{2n} \geq 2n - \frac{Cn}{\log\log\log n}.$$

Proof. Fix a small $\varepsilon < \frac{1}{7}$. We will prove, that for large n, namely $\log\log\log n > \frac{C}{\varepsilon}$, $z_{2n} \geq (2 - 6\varepsilon)n$. This will lead to the result.

First, from Lemma 4.9, we can assume that n is sufficiently large, so that all even numbers smaller than $(2 - \varepsilon)n$ were already used in the sequence before the term $2n$. For this we need only $\log\log n > \frac{C}{\varepsilon}$, which is one log weaker than the assumption made in this theorem.

Let $p_1, p_2, \dots$ be a sequence of the prime numbers in increasing order ($p_1 = 2$ and $p_2 = 3$). By Merten's theorem (Theorem 2.5 and Corollary 2.6)

$$\prod_{n=1}^{K}\left(1-\frac{1}{p_n}\right) < \frac{C}{\log K}.$$

Let K be sufficiently large (namely $\log K > \frac{C}{\varepsilon}$) such that

$$\prod_{n=3}^{K}\left(1-\frac{1}{p_n}\right) < \varepsilon.$$

Notice, that we omitted $p_1 = 2$ and $p_2 = 3$ in this product.

Let $P = \prod_{n=3}^{K} p_i$. Since $p_i \le C i \log i$,

$$P \le (CK\log K)^K \le (CK)^{2K} \le e^{\frac{2C}{\varepsilon}e^{\frac{C}{\varepsilon}}} \le e^{e^{\frac{C'}{\varepsilon}}}.$$

Therefore $\log\log P \le \frac{C}{\varepsilon}$. Since we assumed that $\log\log\log n > \frac{C}{\varepsilon}$, we can assume that $n > \frac{P}{\varepsilon} + P$ by choosing appropriate universal constants.

Let us concentrate on numbers smaller than $(2-3\varepsilon)n$, not divisible by any of the numbers $p_3, p_4, \dots, p_K$. Let $s = \lfloor\frac{(2-3\varepsilon)n}{P}\rfloor > \frac{1}{\varepsilon}$. Then in every segment $(iP, (i+1)P]$ for $i = 0, 1, \dots, s-1$ there are exactly $P \cdot \prod_{i=3}^{K}(1-\frac{1}{p_i}) < P\varepsilon$ numbers not divisible by any $p_3, p_4, \dots, p_K$. So in $[1, (2-3\varepsilon)n]$ there are at most $sP\varepsilon + ((2-3\varepsilon)n - sP) < 2n\varepsilon + P < 3\varepsilon n$ numbers not divisible by any of $p_3, p_4, \dots, p_K$.

Since $n > \frac{P}{\varepsilon} + P$, then $n > \frac{2p_K}{\varepsilon}$. Then, since all even numbers smaller than $(2-\varepsilon)n$ appear in the sequence before index z_{2n}, for all $3 \le i \le K$ some number divisible by $2p_i$ greater than $(2-2\varepsilon)n$ appears in the sequence before index z_{2n}. Now we are going to use Lemma 4.9 to estimate, that for all $3 \le i \le K$ all numbers divisible by p_i smaller than $(2-3\varepsilon)n$ were used in the sequence before the term $2n$. Let $3 \le i \le K$. To use Lemma 4.9, we need $\frac{Cp_i n}{\log\log n} < \varepsilon n$. However, $p_i \le p_K \le CK\log K < e^{\frac{C'}{\varepsilon}}$, so we need to prove that $\frac{Ce^{\frac{C'}{\varepsilon}}}{\varepsilon} < \log\log n$. By applying logarithm to both sides we get $\frac{C''}{\varepsilon} < \log\log\log n$, which was our assumption.

But, as we proved before, there are at most $3\varepsilon n$ numbers not divisible by any of the $p_3, p_4, \dots, p_K$ smaller than $(2-3\varepsilon)n$. In total, there are at most $(3\varepsilon + 3\varepsilon)n = 6\varepsilon n$ numbers smaller than $2n$, than might not have been used in the sequence before index z_{2n}. Therefore, for large n, $z_{2n} > (2-6\varepsilon)n$, which completes the proof. □

Corollary 4.11.

$$\lim_{n\to\infty}\frac{2n}{z_{2n}} = 1.$$

More precisely, there exists a universal constant C such that for all n:

$$2n - \frac{Cn}{\log\log\log n} \le z_{2n} \le 2n + \frac{Cn}{\log\log n}.$$

Since for every prime $p > 3$, we have $z_p = z_{2p} + 1 = z_{3p} - 1$, and from Corollary 4.11, $\lim_{i\to\infty} \frac{2p_i}{z_{2p_i}} = 1$, the following corollary is obvious:

Corollary 4.12.

$$\lim_{i\to\infty} \frac{p_i}{z_{p_i}} = \frac{1}{2}$$

$$\lim_{i\to\infty} \frac{3p_i}{z_{3p_i}} = \frac{3}{2}$$

From Corollary 4.11 we can conclude a few other facts that can be useful in our final proof.

Corollary 4.13. *Let $x > 0$. Then there exists a universal constant C such that all even terms smaller than x appear in the EKG sequence before index $x + \frac{Cx}{\log\log x}$.*

Corollary 4.14. *Let $x > 0$. Then there exists a universal constant C such that all even terms greater than $x + \frac{Cx}{\log\log\log x}$ appear in the sequence after index x.*

Proof. This is straightforward corollary from Corollary 4.11 — if $2n > x + \frac{Cx}{\log\log\log x}$ then $z_{2n} > 2n - \frac{C'n}{\log\log\log n} > x$ by adjusting the constant C. □

4.5 Linear limit for all terms

In this final section we will prove that if (e_i) is a sequence of all **odd** natural numbers except these of the form $3p$ and p for p prime, in increasing order, then $\lim_{i\to\infty} \frac{e_i}{z_{e_i}} = 1$. This, together with linear bounds proved by Lagarias, Rains and Sloane [1] and Corollary 4.11, gives us the desired result: $\lim_{n\to\infty} \frac{a'_n}{n} = 1$.

First we will prove that any e_i could not appear in the sequence too early, i.e., it can appear at position $e_i - o(e_i)$. This will be a quite straightforward corollary from Corollary 4.12: if some e_i appears, there must have been a not much smaller even term before.

Later, we will prove that e_i could not appear in the sequence too late, i.e., it can appear at position $e_i + o(e_i)$. In this proof once again we will be estimating how many times do the sequence cross some border upwards or downwards. We will prove, that not much bigger even term was not used before the term e_i.

Theorem 4.15.

$$\limsup_{i\to\infty} \frac{e_i}{z_{e_i}} \leq 1.$$

More precisely, there exists a universal constant C such that for all e_i

$$z_{e_i} \geq e_i - \frac{Ce_i}{\log\log\log e_i}.$$

Proof. Fix i and let $n = z_{e_i}$, i.e., $a_n = e_i$. Since e_i is not of the form of p or $3p$, then a_{n-1} is composite. Lagarias, Rains and Sloane [1] proved that $a_{n-1} < 14(n-1) < 14n < 14\cdot 260a_n = 14 \cdot 260e_i$. Therefore a_{n-1} has a prime divisor q satisfying $q \leq \sqrt{a_{n-1}} < \sqrt{14 \cdot 260e_i}$. Since

$a_n = e_i$, all multiples of q smaller than e_i must have been used in the sequence before the term a_n. Let $x(i)$ be biggest integer divisible by $2q$ smaller than e_i. Since $q \mid x(i)$, $z_{x(i)} < n$. Obviously $x(i) \geq e_i - 2q > e_i - 2\sqrt{14 \cdot 260}\sqrt{e_i}$. But $x(i)$ is even, so from Corollary 4.11 we have $z_{x(i)} \geq x(i) - \frac{Cx(i)}{\log\log\log x(i)}$ for some universal constant C. Therefore

$$z_{e_i} = n > z_{x(i)} \geq x(i) - \frac{Cx(i)}{\log\log\log x(i)} > e_i - C\sqrt{e_i} - \frac{Ce_i}{\log\log\log e_i} \geq e_i - \frac{C'e_i}{\log\log\log e_i}.$$

□

Theorem 4.16.

$$\liminf_{i\to\infty} \frac{e_i}{z_{e_i}} \geq 1.$$

More precisely, there exists a universal constant C such that for all e_i

$$z_{e_i} \leq e_i + \frac{Ce_i}{\log\log\log e_i}.$$

Proof. We will prove the theorem by contradiction. Assume there exists an increasing unbounded sequence $0 < c_0 < c_i \to \infty$ such that for all I there exists $i > I$ such that $z_{e_i} > (1 + 7\varepsilon_i)e_i$ where $\varepsilon_i = \frac{c_i}{\log\log\log e_i}$. We can assume that $\varepsilon_i < \frac{1}{7}$, i.e., $c_i \leq \frac{1}{7}\log\log\log e_i$. Let e_i be a very large term satisfying $z_{e_i} > (1 + 7\varepsilon_i)e_i$.

We will start by using Corollaries 4.13 and 4.14 several times. We can use them since $\limsup_{i\to\infty} \frac{\frac{c_i}{\log\log\log e_i}}{\frac{C}{\log\log\log e_i}} = \infty$ and $\frac{C}{\log\log\log e_i}$ is the worst asymptotic bound in Corollaries 4.13 and 4.14.

First, using Corollary 4.13 we can assume that e_i is large enough so that all even terms smaller than e_i appeared before index $(1 + \varepsilon_i)e_i$. In other words, for all $2n < e_i$ we have $z_{2n} < (1 + \varepsilon_i)e_i$.

Second, using Corollary 4.13 a second time, we can assume that e_i is large enough so that all even terms smaller than $(1 + 6\varepsilon_i)e_i$ have appeared before index $(1 + 7\varepsilon_i)e_i$. In other words, for all $2n < (1 + 6\varepsilon_i)e_i$ we have $z_{2n} < (1 + 7\varepsilon_i)e_i < z_{e_i}$.

Third, using Corollary 4.14 we can assume that e_i is large enough so that all even terms greater than $(1 + 2\varepsilon_i)e_i$ have appeared after index $(1 + \varepsilon_i)e_i$. In other words, for all $2n > (1 + 2\varepsilon_i)e_i$ we have $z_{2n} > (1 + \varepsilon_i)e_i$.

To sum up, we have an interval of indices (i.e., a part of sequence) between $(1+\varepsilon_i)e_i$ and $(1+7\varepsilon_i)e_i$, where all even terms between $(1+2\varepsilon_i)e_i$ and $(1+6\varepsilon_i)e_i$ appear and no even term smaller than e_i appear. Let us call these even terms between $(1 + 2\varepsilon_i)e_i$ and $(1 + 6\varepsilon_i)e_i$ the *important* terms.

Since e_i is composite, it has prime divisor $q(i)$ not greater than $\sqrt{e_i}$. Therefore, among all *important* terms, we have at least

$$\left\lfloor \frac{4\varepsilon_i e_i}{2q(i)} \right\rfloor \geq \lfloor 2\varepsilon_i\sqrt{e_i} \rfloor \geq \varepsilon_i\sqrt{e_i} \geq \frac{c_0\sqrt{e_i}}{\log\log\log e_i}$$

terms divisible by $q(i)$. Let us call these terms *very important* ones.

Every *very important* term is not of the form $2p$ for $p > 3$ prime, since it is divisible by $2q(i)$ and greater than $2q(i)$. And obviously after every *very important* term the number e_i is a candidate for the next term. Therefore we have at least $\frac{c_0\sqrt{e_i}}{\log\log\log e_i}$ moments between indices $(1+\varepsilon_i)n$ and $(1+7\varepsilon_i)n$, when the EKG sequence crosses the border of e_i downwards. Notice, that none of these moments are subsequences of the form $2p, p$, since a *very important* term is not of the form $2p$.

From Lemma 4.5 we know, that before index z_{e_i} the EKG sequence can cross the border of e_i upwards at most $\pi(\sqrt{e_i})$ times, except for the subsequences of the form $p, 3p$ for p prime.

Let us look at the subsequence $p, 3p$ for p prime that appeared in the EKG sequence after index $(1+\varepsilon_i)e_i$ and before index $(1+7\varepsilon_i)e_i$. The term before was obviously $2p$ and, since every even term smaller than e_i has appeared before index $(1+\varepsilon_i)e_i$, we have that $2p > e_i$, except for maybe one moment where $z_{2p} = \lfloor (1+\varepsilon_i)e_i \rfloor$. Otherwise we have $p < e_i < 2p < 3p$, so the EKG sequence crossed here the border of e_i first downwards and then upwards.

Therefore between indices $(1+\varepsilon_i)e_i$ and $(1+7\varepsilon_i)e_i$ the number of times the sequence crosses the border of e_i upwards and downwards, without taking into account subsequences of the form $2p, p$ and $p, 3p$ for p prime, should differ by at most two. The other ones just form pairs — subsequences of the form $2p, p, 3p$ with $p < e_i < 2p < 3p$. But we proved that we have at least $\frac{c_0\sqrt{e_i}}{\log\log\log e_i}$ downward crossings and at most $\pi(\sqrt{e_i}) = O(\frac{\sqrt{e_i}}{\log e_i})$ upward crossings, which is a contradiction. This completes the proof. $\square$

To sum up, in the end, we have proved the following theorem, by Theorems 4.15, 4.16 and Corollaries 4.11 and 4.12:

Theorem 4.17.

$$\lim_{n\to\infty} \frac{a'_n}{n} = 1.$$

More precisely, there exists a universal constant C such that

$$n - \frac{Cn}{\log\log\log n} \le a'_n \le n + \frac{Cn}{\log\log\log n}.$$

5 Conclusion

In this paper we presented a quite elementary proof that in the EKG sequence odd primes appear in the subsequences $2p$, p, $3p$ (Theorem 4.3) and, building some technical tools on top of observation made in Lemma 4.1 and Lemma 4.5, we managed to prove asymptotic behavior of the EKG sequence, namely, that $\lim_{n\to\infty} \frac{a'_n}{n} = 1$. We tried to estimate the speed of the convergence through the proofs to see if we could use these techniques to cope with the stronger conjecture made by Lagarias, Rains and Sloane [1]:

Conjecture 5.1 (Lagarias, Rains, Sloane)**.**

$$a'_n \sim n\Big(1 + \frac{1}{3\log n}\Big).$$

The proved result, Theorem 4.17, is far from this goal. We were able to prove a very weak convergence speed — $\frac{C}{\log\log\log n}$ tail. This is the reason we were not able to apply any argument similar to those supporting Conjecture 5.1. Let us repeat the argument. In the EKG sequence integers appear more or less in the increasing order, except for terms of the form p and $3p$. If $a_n = m$, then earlier in the sequence we had all integers smaller than m except for numbers of the form p and $3p$ smaller than m, but we had all p and $3p$ for $2p \leq m$. This leads to the Conjecture 5.1 by

$$n \sim m - \pi(m) - \pi\left(\frac{m}{3}\right) + 2\pi\left(\frac{m}{2}\right) \sim m - \frac{m}{3\log m}.$$

Unfortunately, in this argument we need to investigate sets of missing terms (numbers of the form p and $3p$) that are of the size $\frac{Cn}{\log n}$. Such sizes are lost comparing to $\frac{Cn}{\log\log\log n}$ part. Moreover, in this paper we get weaker convergence speed ($\frac{Cn}{\log\log n}$) very early, namely in the Lemma 4.6, so we were not able to use this technique.

To get closer to proving Conjecture 5.1 using techniques presented in this paper, one needs to improve the convergence speed. Note that there are two places where we gain one more log in the speed. One is Lemma 4.6, where we need to multiply $\pi(x)$ by the number of different prime divisors of x, which might be $\frac{\log x}{\log\log x}$. The proof of Lemma 4.6 is quite tricky and we were not able to improve it.

The second place where we lose out is the proof of Theorem 4.10, probably the most technical proof in this paper. The estimations made there are quite rough, therefore we think they might be somehow improved. However, we do not know how to do it now.

To sum up, although this paper provides some new tools and techniques to cope with the EKG sequence, and proves the core part of the conjectures made by Lagarias, Rains and Sloane [1], there is still quite a significant hole between what is done and what is conjectured. The aforementioned Conjecture 5.1 seems to us to be the most interesting question remaining.

Acknowledgments

We would like to say a huge “thank you” to few our professors and colleagues, without whom this work wouldn’t exist. Chronologically first, to Professor Jaroslaw Grytczuk and to Andrzej Grzesik for showing us the EKG sequence. Later, to our RPG Fellowship, Marek Grabowski, Kaja Malek, Michal Pilipczuk and Jakub Wojtaszczyk, for patience in listening to our proofs and concepts instead of playing Earthdawn. Finally, to Joanna Jaszunska and Jakub Wojtaszczyk for pointing out many errors and an almost infinite number of missing articles in the paper. And finally, to Professor Jerzy Urbanowicz, for being our mentor while working on this problem.

References

[1] J. C. Lagarias, E. M. Rains, and N. J. A. Sloane. The EKG sequence. *Experimental Math.*, **11** (2002), 437–446.

[2] F. Mertens. Ein Beitrag zur analytischen Zahlentheorie. *J. reine angew. Math.*, **78** (1874), 46–62.

[3] B. Rosser. Explicit bounds for some functions of prime numbers. *American J. Math.*, **63** (1941), 211–232.

[4] N. J. A. Sloane. The On-Line Encyclopedia of Integer Sequences. `http://www.research.att.com/~njas/sequences/`.

2000 *Mathematics Subject Classification*: Primary 11B83.
Keywords: EKG sequence, integer sequence, prime numbers.

(Concerned with sequence A064413.)

Received March 4 2008; revised version received September 20 2008. Published in *Journal of Integer Sequences*, October 2 2008.

Return to Journal of Integer Sequences home page.