LINES IN THE PRIME NUMBER GRAPH

Carl Pomerance
Dartmouth College, Hanover, New Hampshire, USA
`carlp@math.darmouth.edu`

Patrick Solé
I2M, (CNRS, Aix-Marseille University), Marseille, France
`patrick.sole@telecom-paris.fr`

Received: , Revised: , Accepted: , Published:

Abstract

The prime number graph is the set of points (n, p_n) where p_n denotes the n^{th} prime. Let $L(n)$ be the minimum number of straight line segments needed to cover the first n points in this set. Let $B(n)$ be the largest number of points (k, p_k) with $k \leq n$ covered by a single line. Recently Sloane conjectured that $L(n) = O(n/\log n)$. We show that $L(n) = O(n \log\log n/\log n)$ and $B(n) \geq c \log n$ for a constant $c > 0$ and all large n. Under RH we show that for large n we have $B(n) = O(n^{3/4}(\log n)^{1/2})$ and $L(n) \geq c' n^{1/4}(\log n)^{-1/2}$ for some constant $c' > 0$.

AMS Subject Classification: 11N05.
Keywords: Prime Number Theorem.

1. Introduction

Let $p_1, p_2, \ldots$ denote the sequence of primes. A **prime point** is a point of the plane of the form (k, p_k) for some k. This graphical representation of the primes was considered in [8]. It is interesting to look at sets of these prime points that are colinear, such as

$$(6, 13), (7, 17), (10, 29), (12, 37), (13, 41), (16, 53),$$

which are all on the line $y = 4x - 11$. Let $L(n)$ be the minimum number of lines needed to cover the first n prime points. For example, $L(2) = 1$ and $L(3) = 2$. A prime p_n is **awkward** if $L(n) > L(n-1)$. These concepts were introduced in the recent Numberphile videos [1], [2]. See [5] for numerics of $L(n)$ (called there $a(n)$) for small n. See [6] for the list of the first awkward primes.

In this note we study how the function $L(n)$ behaves for large n. Since the primes have asymptotic density 0, it is clear that no line can contain infinitely many prime points (k, p_k). We introduce the function $B(n)$ which is the largest number of prime points among the first n of them covered by a single line, and derive upper and lower bounds for it. The arguments are based on the Prime Number Theorem with remainder.

The material is arranged as follows. The next section recalls some known results on the topic of the prime number graph and the Prime Number theorem with remainder. Section 3 studies the function $L(n)$ and Section 4 the function $B(n)$. Section 5 concludes this note and underlines some challenging open problems.

2. Some old results

We begin by an observation due to Erdős and quoted without proof in [8].

Theorem 1. *For any integer k almost all prime points (n, p_n) lie on a line with k other prime points. That is, the set of primes for which this is so has relative density 1 in the set of primes.*

From there we deduce an asymptotic upper bound on $L(n)$.

Corollary 1. *For $n \to \infty$ we have $L(n) = o(n)$.*

We shall present a detailed proof of a more quantitative version of this theorem and corollary in the next section.

We also derive a lower bound for $B(n)$ with the proof strongly based on the proof of [8, Theorem 4.1].

Based on the experimental data of [5], Sloane conjectures in [1] that $L(n) = O(\frac{n}{\log n})$. This motivates Theorem 3 of the next section, where we come close to proving Sloane's conjecture.

As mentioned in the Introduction, our proofs strongly use the Prime Number Theorem with remainder. In particular, let

$$\mathrm{li}(x) = \int_0^x \frac{dt}{\log t}$$

denote the logarithmic integral function (where the principal value is taken for the singularity at $t = 0$). Then $\mathrm{li}(x) \sim x/\log x \sim \pi(x)$ as $x \to \infty$, but the $\mathrm{li}(x)$ approximation to $\pi(x)$ is much more accurate. In particular, we have

$$|\pi(x) - \mathrm{li}(x)| \leq x/\exp(c(\log x)^{3/5}(\log\log x)^{-1/5})$$

for a positive constant c and all large x. If the Riemann Hypothesis is assumed, then this result improves to

$$|\pi(x) - \mathrm{li}(x)| \leq x^{1/2}\log x$$

for all $x \geq 2$. See [3], [4]. (See [7] for an asymptotically weaker, but numerically explicit version of the former result.)

3. The Erdős observation and awkward primes

In this section we give a quantitative proof of Theorem 1. In particular we will prove the following.

Theorem 2. *Let $N(x)$ denote the number of $n \leq x$ such that (n, p_n) lies on a line with at least $\log n/(\log\log n)^2$ other points (m, p_m). Then $N(x) \sim x$ as $x \to \infty$.*

We will prove Theorem 2 as a corollary of the following result.

Theorem 3. *We have $L(n) = O(n \log\log n/\log n)$.*

Proof of Theorem 3. Let k be a large integer and consider the Farey sequence of level k. Let $a/b, a'/b'$ be two consecutive terms. Then

$$a'/b' - a/b = 1/bb', \quad b, b' \leq k, \quad b + b' > k, \quad \gcd(b, b') = 1.$$

Let $u = e^{k+a/b}$, $u' = e^{k+a'/b'}$ and let I denote the interval $(u, u']$. We consider inverse prime points (p_n, n) with $p_n \in I$. The length $|I|$ of I has

$$|I| = u' - u = e^{k+a'/b'} - e^{k+a/b} = e^{k+a/b}(e^{1/bb'} - 1) = (1 + o(1))u/bb', \quad k \to \infty.$$

Let P denote the parallelogram bounded by the vertical lines $x = u$, $x = u'$ and the lines with slope $1/\log u = 1/(k + a/b)$ through $(u, u - \lfloor w \rfloor)$ and $(u, u + \lceil w \rceil)$, where

$$w = \frac{|I|^2}{u \log^2 u} = (1 + o(1))\frac{u}{(bb')^2 \log^2 u}.$$

These lines gain $|I|/\log u$ on the interval $(u, u']$. This is about the same gain as $\mathrm{li}(x)$ on the interval. Note that by Taylor's theorem,

$$\mathrm{li}(u') - \mathrm{li}(u) = \frac{|I|}{\log u} - \left(\frac{1}{2} + o(1)\right)\frac{|I|^2}{u \log^2 u} = \frac{|I|}{\log u} - \left(\frac{1}{2} + o(1)\right) w.$$

Thus, the region

$$|y - \mathrm{li}(x)| \leq w/4, \quad x \in I$$

lies wholly in P, and so by the Prime Number Theorem with remainder, $y = \pi(x)$ with $x \in I$ also lies in P, assuming k is sufficiently large. Hence all of the inverse prime points (p_n, n) with $p_n \in I$ lie in P. And the total number of such points is

$$\pi(u') - \pi(u) = (1 + o(1))\frac{|I|}{\log u}.$$

Now consider those lines with slope $1/\log u = 1/(k + a/b) = b/(bk + a)$ which pass through a lattice point in P. There are

$$(2 + o(1))w(bk + a) \sim 2wb \log u \sim \frac{2u}{bb'^2 \log u} \ll \frac{e^k}{bb'^2 k}$$

such lines. So if we consider all of the lines appearing in this argument for primes in $(e^k, e^{k+1}]$, the number of them is bounded by a constant times

$$\frac{e^k}{k} \sum \frac{1}{bb'^2}.$$

Here the sum is over the full Farey dissection of level k. We have

$$\sum_{\substack{1 \le b, b' \le k \\ b+b' > k}} \frac{1}{bb'^2} = \sum_{1 \le b' \le k} \frac{1}{b'^2} \sum_{k-b' < b \le k} \frac{1}{b} < \sum_{1 \le b' \le k} \frac{1}{b'(k - b' + 1)}$$

$$< \frac{1}{k} + \sum_{1 \le b' < k} \frac{1}{b'(k - b')} = \frac{1}{k} + \frac{1}{k} \sum_{1 \le b' < k} \left(\frac{1}{b'} + \frac{1}{k - b'} \right) \ll \frac{\log k}{k}.$$

Thus, the number of these lines with slope $1/\log u$ that hit at least one lattice point in P is $O(e^k \log k/k^2)$. Summing this for $k \le K - 1$ we have that the total number of lines that contain some (p_n, n) for $p_n \le e^K$ is $O(e^K \log K/K^2)$. If $n \in (\pi(e^{K-1}), \pi(e^K)]$, then $L(n) = O(e^K \log K/K^2) = O(n \log\log n/\log n)$. This completes the proof. □

As a corollary we have Theorem 2.

Proof of Theorem 2. Consider the $O(n \log\log n/\log n)$ lattice lines that cover all of the points (p_j, j) for $j \le n$. Those lines which cover less than $\log n/(\log\log n)^2$ inverse prime points together cover $O(n/\log\log n)$ points. This leaves still asymptotically all n inverse prime points, where each such point is contained in a line with at least $\log n/(\log\log n)^2$ other inverse prime points. □

Theorem 4. *The number of awkward primes among the first n primes is $L(n)$. Thus, the reciprocal sum of the awkward primes is finite.*

Proof. Let $L(0) = 0$. For each positive integer j, we have $L(j) - L(j - 1)$ equal to 0 or 1, where the value 1 occurs if and only if p_j is awkward. We have

$$L(n) = \sum_{j \le n} (L(j) - L(j - 1)),$$

so it is clear then that $L(n)$ is the number of awkward primes p_j with $j \le n$. Since the counting function of the awkward primes up to x is $O(\pi(x) \log\log x/\log x)$, it follows immediately that their reciprocal sum is finite. □

4. The function $B(n)$

We know by [8, Th. 4.1] and Theorem 2 that $B(n)$ is not bounded above. In fact, we have the following estimate.

Theorem 5. *There is a positive constant* c_1 *such that for all large* n *we have* $B(n) \geq c_1 \log n$.

Proof. First, we note that there is a simple combinatorial relation connecting the functions L and B, namely

$$L(n)B(n) \geq n. \tag{4.1}$$

This is immediate by considering a covering of the first n prime points by $L(n)$ line segments. Each line contains at most $B(n)$ prime points. Thus, from Theorem 3 we have $B(n) \gg \log n/\log\log n$. This almost proves the theorem. We use the proof of Theorem 3 for a slight improvement. In that proof we used the Farey dissection of level k to obtain a dissection of the interval $(e^k, e^{k+1}]$. Now we use only the first (and longest) piece of the dissection: $(e^k, e^{k+1/k}]$. Following the proof we have order of magnitude e^k/k^3 lattice lines of slope $1/k$ and order of magnitude e^k/k^2 inverse prime points. Thus, on average each line has order of magnitude k inverse prime points. This completes the proof. □

We remark that using the interval $(e^k, e^{k+1/k}]$ to show that some lines have many prime points already appeared in the proof of [8, Theorem 4.1].

A trivial upper bound for $B(n)$ is n, of course. We can do considerably better. Let $R(x)$ be a smooth, increasing, concave down function such that

$$|\pi(x) - \mathrm{li}(x)| \leq R(x) \tag{4.2}$$

for all large x. The comments in Section 2 about the Prime Number Theorem with remainder pertain to $R(x)$.

Theorem 6. *For* x *sufficiently large and all* $n \leq x$*, we have*

$$B(n) = O((xR(x))^{1/2}).$$

Proof. From the discussion above, we may assume that $R(x) = o(\mathrm{li}(x))$ as $x \to \infty$ and that both functions $y = \mathrm{li}(x) + R(x)$ and $y = \mathrm{li}(x) - R(x)$ are smooth, strictly increasing, and strictly concave down. Thus, a line may intersect these two curves in at most two points each. In fact, a line can intersect the region $|y - \mathrm{li}(x)| \leq R(x)$ at most twice, i.e., either for a single bounded interval I on the positive x-axis or for two disjoint bounded intervals. A calculation shows that the length of such an interval is $O((xR(x))^{1/2} \log x)$. Since the number of primes in such an interval is $O((xR(x))^{1/2})$, the theorem follows. □

Assuming the Riemann Hypothesis (RH) we can give a more explicit upper bound on $B(n)$.

Corollary 2. *Under RH we have for x large and all $n \leq x$,*

$$B(n) = O(x^{3/4}(\log x)^{1/2}),$$

and for some constant $c_2 > 0$,

$$L(n) \geq c_2 x^{1/4}(\log x)^{-1/2}.$$

Proof. By the discussion in Section 2, we can take $R(x) = \sqrt{x}\log x$ under RH. The first bound follows then by Theorem 6 . Plugging this bound into (4.1) yields the second bound. □

5. Conclusion and open problems

In this note we have studied the covering properties of line segments in the prime number graph. We have derived an asymptotic upper bound for the minimum size of a cover, and estimates for the largest number of prime points on a single segment. Our estimates seem far from optimal, as is also suggested from the numerical work in [5], [6]. Regarding the function $L(n)$ there is a gap between Theorem 3 and Corollary 2. As for the function $B(n)$, it would be nice to reduce the gap between Theorem 5 and Corollary 2.

References

[1] Brady Haran and Neil Sloane, Awkward primes, Numberphile video, 2026. https://www.youtube.com/watch?v=VFoIPlUalRY&t=525s

[2] Brady Haran and Neil Sloane, Primes at the end of the line, Numberphile video, 2026. https://www.youtube.com/watch?v=u-_8wX4cECo

[3] H. von Koch, Sur la distribution des nombres premiers, Acta Math. **24** (1901), 159–182.

[4] H. L. Montgomery and R. C. Vaughan, Multiplicative number theory I. Classical Theory, Cambridge University Press, 2007.

[5] A373813 in Online Encyclopedia of Integer Sequences.

[6] A393445 in Online Encyclopedia of Integer Sequences.

[7] D. J. Platt and T. S. Trudgian, The error term in the prime number theorem', Math. Comp. **90** (2021), 871–881.

[8] C. Pomerance, The prime number graph, Math. Comp. **33** (1979), 399–408.