ON THE SELBERG-ERDŐS PROOF OF THE PRIME NUMBER THEOREM

ASHVIN A. SWAMINATHAN

Abstract. In this article, we discuss the first elementary proof, due to Selberg and Erdős, of the Prime Number Theorem. In particular, we begin with a presentation of the Selberg symmetry formula and proceed to give a detailed account of the proof as published by Erdős in 1949.

1. Introduction

The prime numbers, and specifically their enumerating function π defined by

$$\pi(x) = \#\{\text{primes } p : p \leq x\},$$

have been a source of fascination to mathematicians for thousands of years. As early as 300 BC, Euclid discussed a simple proof of the infinitude of primes (i.e. the unboundedness of π) in his famed *Elements*. As late as the 1700s, Euler discovered another, more informative proof that there are infinitely many primes: in fact, he showed that the primes are so frequent among the integers that the series

$$\sum_{\text{prime } p} \frac{1}{p}$$

diverges. Drawing on Euler's methods and employing inputs from complex analysis, like the study of Dirichlet series and Landau's Theorem, Dirichlet proved that there are infinitely many primes in arithmetic progressions. Such tools, and in particular the complex analytic properties of the Riemann ζ-function, were fundamental in establishing the first asymptotic for $\pi(x)$. As an adolescent in the late 1700s, Gauss had correctly posited that $\pi(x) \sim x/\log x \sim \mathrm{Li}(x) := \int_2^x dt/\log t$, but it took a century for his conjecture to be proven. In the 1840s via elementary methods, Chebyshev explicitly computed constants $c_1 < 1 < c_2$, both close to 1 in value, such that

$$c_1 < \frac{\pi(x)}{x/\log x} < c_2, \tag{1}$$

a result that provided a partial validation of Gauss' conjecture. And by 1896, with such implements as Stirling's formula, Jensen's formula, and von Mangoldt's Theorem in hand, Hadamard and Poussin completed the proof of Gauss' conjecture by showing that $\zeta(1 + iT) \neq 0$ for all $T \in \mathbb{R}$.

The aforementioned asymptotic, $\pi(x) \sim x/\log x$, came to be known in the twentieth century as the "Prime Number Theorem" (PNT), and it was believed by many a great mathematician, including Hardy, that the PNT and its proof were intimately connected to complex analysis; i.e. it was thought that the PNT is a "deep" theorem that cannot be deduced by elementary means. This belief was strikingly disproved by Selberg and Erdős in the 1940s, when Selberg employed elementary methods to obtain the so-called *Selberg symmetry formula*

$$\sum_{\substack{\text{prime } p \\ p \leq x}} \log^2 p + \sum_{\substack{\text{prime } p,q \\ pq \leq x}} \log p \log q = 2x \log x + O(x), \tag{2}$$

with which Erdős and Selberg individually managed to find a completely elementary proof of the PNT. In this article, we begin with a presentation of the Selberg symmetry formula along with is proof, and we then proceed to give a detailed account of the elementary proof of the PNT as published by Erdős in 1949 ([Erd49]).

2. The Selberg Symmetry Formula

The Selberg symmetry formula (2) is a key input in the elementary proof of the PNT that is given in Section 3 (as it happens, (2) can be very easily deduced from the PNT!). The following account of the formula and its proof is inspired by a weblog post of Tao ([Tao07]) and by an annotation of Balady ([Bal06]).

To prove (2), it seems natural to consider defining a function that is supported not only on primes (as is suggested by the first sum in the equation), but also on those numbers that can be expressed as a product of two primes (as is suggested by the second sum). Given such a function, call it f, we might be able to replace the left-hand-side of (2) with a single, more tractable sum over all $n \leq x$ of $f(n)$. In this regard, we make the following definition: for positive integers n, let the *2^{nd} von Mangoldt function* $\Lambda_2(n)$ be defined by

$$\Lambda_2(n) := \sum_{d|n} \mu(d) \log^2\left(\frac{n}{d}\right).$$

We claim that $\Lambda_2(n)$ is an approximate indicator function for the primes and products of two primes; i.e. we claim that $\Lambda_2(n)$ is, in large part, supported on the primes and products of two primes. To begin with, notice that $\Lambda_2(p) = \log^2 p \neq 0, \Lambda_2(p^2) = 3\log^2 p \neq 0$ for any prime p, and that $\Lambda_2(pq) = 2\log p \log q \neq 0$ when p, q are distinct primes. Moreover, as we will now show, $\Lambda_2(n) = 0$ whenever n has at least 3 distinct prime factors:

Proposition 1. *For all positive integers n such that either $n = 1$ or n has at least 3 distinct prime factors, we have that $\Lambda_2(n) = 0$.*

Proof. It is obvious that $\Lambda_2(1) = 0$. Now suppose that n has at least 3 distinct prime factors. Write n in terms of its prime factorization as $n = p_1^{e_1} \dots p_k^{e_k}$. Then, we have the following equalities:

$$\begin{aligned}
\Lambda_2(n) &= \sum_{d|n} \mu(d) \log^2 \left(\frac{n}{d}\right) \\
&= \sum_{\substack{d|n \\ d \text{ sq-free}}} \mu(d) \log^2 \left(\frac{p_1^{e_1} \dots p_k^{e_k}}{d}\right) \\
&= \sum_{\substack{d|n \\ d \text{ sq-free}}} \mu(d) \left(\sum_{i=1}^{k} e_i \log p_i - \log d\right)^2 \\
&= \left(\sum_{i=1}^{k} e_i \log p_i\right)^2 \sum_{\substack{d|n \\ d \text{ sq-free}}} \mu(d) - 2\sum_{i=1}^{k} e_i \log p_i \sum_{\substack{d|n \\ d \text{ sq-free}}} \mu(d) \log d + \\
&\quad \sum_{\substack{d|n \\ d \text{ sq-free}}} \mu(d) \log^2 d.
\end{aligned}$$

Now, to complete the proof, it suffices to show that the following three sums are 0:

$$\sum_{\substack{d|n \\ d \text{ sq-free}}} \mu(d) = \sum_{\substack{d|n \\ d \text{ sq-free}}} \mu(d) \log d = \sum_{\substack{d|n \\ d \text{ sq-free}}} \mu(d) \log^2 d = 0.$$

But the above equalities follow easily from applying the binomial theorem to the fact that $(1-1)^m = 0$ for all $m \in \mathbb{Z}_{>0}$ (it is also necessary to assume that n has sufficiently many prime factors; e.g. if $n > 1$, then n must have at least three distinct prime factors for the third sum to be 0). □

It follows from our above computation of $\Lambda_2(p)$, $\Lambda_2(p^2)$, and $\Lambda_2(pq)$ and from Proposition 1 that $\Lambda_2(n)$ that the function $\Lambda_2(n)$ serves as an approximate indicator function for the primes and products of two primes. To understand the left-hand-side of (2), it now seems useful to study the following sum:

$$\sum_{n \le x} \Lambda_2(n) = \sum_{n \le x} \sum_{d|n} \mu(d) \log^2 \left(\frac{n}{d}\right). \tag{3}$$

We will now show that the double sum on the right-hand-side of (3) gives rise to the left-hand-side of (2). To evaluate this double sum, we introduce a new function called the *1st von Mangoldt function* and prove the following lemma.

Lemma 2. *For positive integers n, let the 1st von Mangoldt function $\Lambda(n)$ be defined by*

$$\Lambda(n) = \sum_{d|n} \mu(d) \log\left(\frac{n}{d}\right).$$

Then, we have that

$$\sum_{d|n} \mu(d) \log^2\left(\frac{n}{d}\right) = \Lambda(n)\log n + \sum_{d|n} \Lambda(d)\Lambda\left(\frac{n}{d}\right).$$

Proof. It follows from the above definition of $\Lambda(n)$ that $\Lambda(n) = \log p$ when $n = p^k$ for p prime and positive integers k, and that $\Lambda(n) = 0$ otherwise. By expressing n in its prime factorization, we can then deduce that

$$\log n = \sum_{d|n} \Lambda(d).$$

By applying the above equality repeatedly, we obtain the following equalities:

$$\begin{aligned}
\log^2 n &= \sum_{d|n} \Lambda(d) \log n \\
&= \sum_{d|n} \Lambda(d) \log\left(\frac{n}{d}\right) + \sum_{d|n} \Lambda(d) \log d \\
&= \sum_{d|n} \Lambda(d) \sum_{a|\frac{n}{d}} \Lambda(a) + \sum_{d|n} \Lambda(d) \log d \\
&= \sum_{ad|n} \Lambda(a)\Lambda(d) + \sum_{d|n} \Lambda(d) \log d \\
&= \sum_{b|n} \left(\Lambda(b)\log b + \sum_{d|b} \Lambda(d)\Lambda\left(\frac{b}{d}\right)\right).
\end{aligned}$$

The desired equality then follows by applying the Möbius inversion formula. ☐

By Lemma 2, the double sum on the right-hand-side of (3) may be written as follows:

$$\begin{aligned}
\sum_{n \le x} \sum_{d|n} \mu(d) \log^2\left(\frac{n}{d}\right) &= \sum_{n \le x} \left(\Lambda(n)\log n + \sum_{d|n} \Lambda(d)\Lambda\left(\frac{n}{d}\right)\right) \\
&= \sum_{n \le x} \Lambda(n) \log n + \sum_{mn \le x} \Lambda(m)\Lambda(n).
\end{aligned}$$

Now, observe that we have

$$\sum_{n \le x} \Lambda(n) \log n = \sum_{\substack{\text{prime } p \\ p \le x}} \log^2 p + O(\sqrt{x}\log^2 x),$$

because the prime powers in the support of $\Lambda(n)$ make a negligible contribution to the sum.

We also claim that

$$\sum_{mn\leq x} \Lambda(m)\Lambda(n) = \sum_{\substack{\text{prime } p,q \\ pq\leq x}} \log p \log q + O(x).$$

By rearranging the claimed equality, it suffices to show that

$$\sum_{\substack{mn\leq x \\ m,n \text{ not both prime}}} \Lambda(m)\Lambda(n) = O(x),$$

and by symmetry, it further suffices to show that

$$\sum_{\substack{mn\leq x \\ m \text{ not prime}}} \Lambda(m)\Lambda(n) = O(x).$$

For any fixed $m \leq x$, we have by Chebyshev's bound (1) that

$$\sum_{n\leq x/m} \Lambda(m)\Lambda(n) \ll \Lambda(m) \cdot \frac{x}{m}.$$

Thus, we have that

$$\sum_{\substack{mn\leq x \\ m \text{ not prime}}} \Lambda(m)\Lambda(n) \ll \sum_{\substack{m\leq x \\ m \text{ not prime}}} \Lambda(m) \cdot \frac{x}{m} \ll \sum_{\substack{m\leq x \\ m \text{ not prime}}} x \cdot \frac{\log m}{m} \ll x,$$

which yields the above claim. From the previous two estimates, it follows that we have

$$\begin{aligned} \sum_{n\leq x}\sum_{d|n} \mu(d) \log^2\left(\frac{n}{d}\right) &= \sum_{n\leq x} \Lambda(n)\log n + \sum_{mn\leq x} \Lambda(m)\Lambda(n) \\ &= \sum_{\substack{\text{prime } p \\ p\leq x}} \log^2 p + \sum_{\substack{\text{prime } p,q \\ pq\leq x}} \log p \log q + O(x). \end{aligned}$$

In this sense, the right-hand-side of (3) does give rise to the left-hand-side of (2). Now, to prove (2), it suffices to show that the right-hand-side of (3) can be estimated as follows:

$$\sum_{n\leq x}\sum_{d|n} \mu(d) \log^2\left(\frac{n}{d}\right) = 2x\log x + O(x). \tag{4}$$

To obtain the above estimate, we require the following lemma, which contains a number of other useful estimates:

Lemma 3. *We have the following estimates:*

$$\sum_{n \le x} \frac{\mu(n)}{n} = O(1), \tag{5}$$

$$\sum_{n \le x} \frac{\mu(n)}{n} \log\left(\frac{x}{n}\right) = O(1), \tag{6}$$

$$\sum_{n \le x} \frac{\mu(n)}{n} \log^2\left(\frac{x}{n}\right) = 2\log x + O(1). \tag{7}$$

Proof. For (5), the Möbius Inversion Formula tells us that

$$1 = \sum_{n \le x} \mu(n) \left(\frac{x}{n} + O(1)\right) = x \sum_{n \le x} \frac{\mu(n)}{n} + O(x),$$

from which we deduce that

$$\sum_{n \le x} \frac{\mu(n)}{n} = O(1),$$

as desired. For (6), approximating with an integral yields that

$$\sum_{n \le x} \frac{x}{n} = x \log x + Cx + O(1),$$

where C is a fixed constant. Then, the Möbius Inversion Formula tells us that

$$\begin{aligned} x &= \sum_{n \le x} \mu(n) \left(\frac{x}{n} \log\left(\frac{x}{n}\right) + \frac{Cx}{n} + O(1)\right) \\ &= x \sum_{n \le x} \frac{\mu(n)}{n} \log\left(\frac{x}{n}\right) + Cx \sum_{n \le x} \frac{\mu(n)}{n} + O(x). \end{aligned}$$

By applying (5) and rearranging, we obtain that

$$\sum_{n \le x} \frac{\mu(n)}{n} \log\left(\frac{x}{n}\right) = O(1),$$

as desired. For (7), approximating with an integral yields that

$$\sum_{n \le x} \frac{x}{n} \log\left(\frac{x}{n}\right) = \frac{x}{2} \log^2 x + Cx \log x - Dx + O(\log x),$$

where C and D are fixed constants. Then, the Möbius Inversion Formula tells us that

$$\begin{aligned} x \log x &= \frac{x}{2} \sum_{n \leq x} \frac{\mu(n)}{n} \log^2 \left(\frac{x}{n}\right) + Cx \sum_{n \leq x} \frac{\mu(n)}{n} \log \left(\frac{x}{n}\right) - \\ & Dx \sum_{n \leq x} \frac{\mu(n)}{n} + \sum_{n \leq x} O\left(\log\left(\frac{x}{n}\right)\right). \end{aligned}$$

Approximating with an integral also yields that

$$\sum_{n \leq x} \log \left(\frac{x}{n}\right) = x + O(\log x),$$

and this estimate along with (5) and (6) tell us that

$$\begin{aligned} x \log x &= \frac{x}{2} \sum_{n \leq x} \frac{\mu(n)}{n} \log^2 \left(\frac{x}{n}\right) + Cx \sum_{n \leq x} \frac{\mu(n)}{n} \log \left(\frac{x}{n}\right) - \\ & Dx \sum_{n \leq x} \frac{\mu(n)}{n} + \sum_{n \leq x} O\left(\log\left(\frac{x}{n}\right)\right) \\ &= \frac{x}{2} \sum_{n \leq x} \frac{\mu(n)}{n} \log^2 \left(\frac{x}{n}\right) + O(x) + O(x) + O(x), \end{aligned}$$

from which we deduce that

$$\sum_{n \leq x} \frac{\mu(n)}{n} \log^2 \left(\frac{x}{n}\right) = 2 \log x + O(1),$$

which is the desired result. □

The last estimate we will require follows easily from approximation by an integral:

$$\sum_{n \leq x} \frac{\log n}{n} = \frac{1}{2} \log^2 x + C + O\left(\frac{\log x}{x}\right) \tag{8}$$

for some fixed constant C. Now, observe that we may expand the double sum on the left-hand-side of (4) in the following way:

$$\begin{aligned}\sum_{n\leq x}\sum_{d|n}\mu(d)\log^2\left(\frac{n}{d}\right) &= \sum_{ad\leq x}\mu(d)\log^2 a \\ &= \sum_{d\leq x}\mu(d)\sum_{a\leq x/d}\log^2 a \\ &= \sum_{d\leq x}\frac{\mu(d)}{d}\log^2\left(\frac{x}{d}\right) - 2x\sum_{d\leq x}\frac{\mu(d)}{d}\log\left(\frac{x}{d}\right) + \\ &\quad 2x\sum_{d\leq x}\frac{\mu(d)}{d} + O\left(\sum_{d\leq x}\log^2\left(\frac{x}{d}\right)\right),\end{aligned}$$

where the last equality follows from applying the estimate (8). We now apply the estimates given in Lemma 3. By (7), the first term in the above sum gives rise to the main term of $2x\log x + O(x)$, by (6), the second term in the above sum is just $O(x)$, and by (5), the third term in the above sum is also just $O(x)$. Finally, by approximating with an integral, it is easy to see that

$$\sum_{d\leq x}\log^2\left(\frac{x}{d}\right) = O(x),$$

and applying this estimate to the fourth term in the above sum yields that the fourth term is also just $O(x)$. We thus conclude that

$$\sum_{n\leq x}\sum_{d|n}\mu(d)\log^2\left(\frac{n}{d}\right) = 2x\log x + O(x),$$

which yields the desired equality (4). But we already showed that (4) implies (2), so we have indeed proven the Selberg symmetry formula (2). It is worth noting that this formula is a useful tool for devising elementary proofs of many theorems in analytic number theory, not just the PNT ([Erd49]).

3. The Proof of the PNT

We now present the elementary proof of the PNT, due to Selberg and Erdős. We will begin with a proposition of Erdős (i.e. equation (2) from [Erd49]), which makes use of the Selberg symmetry formula (2), and we will conclude by deducing the PNT from this proposition and from the Selberg symmetry formula. As is observed by Erdős in [Erd49], it is not necessary to use the full strength of the Selberg symmetry formula: in the following, we only require that

$$\sum_{\substack{\text{prime } p\\ p\leq x}}\log^2 p + \sum_{\substack{\text{prime } p,q\\ pq\leq x}}\log p\log q = 2x\log x + o(x\log x). \tag{9}$$

Define $\vartheta(x)$ by

$$\vartheta(x) = \sum_{\substack{\text{prime p} \\ p \leq x}} \log p,$$

and let a and A be defined by

$$a = \liminf_{x \to \infty} \frac{\vartheta(x)}{x} \text{ and } A = \limsup_{x \to \infty} \frac{\vartheta(x)}{x}.$$

We now prove the following proposition of Erdős:

Proposition 4 (Erdős). *For every $c > 0$, there exists $\delta(c) > 0$ such that*

$$\pi(x(1+c)) - \pi(x) > \delta(c) \frac{x}{\log x}$$

for sufficiently large x.

Proof of Proposition 4 (from [Erd49]*).* Recall that $\pi(x) \sim x/\log x \Leftrightarrow \vartheta(x) \sim x$. It therefore suffices (and is in fact equivalent) to show that for every $x > 0$ there exists $\delta(c) > 0$ such that

$$\vartheta(x(1+c)) - \vartheta(x) > \delta(c)x$$

for sufficiently large x. We proceed by contradiction. If the proposition is false, then there exist constants $c > 0$ such that

$$\vartheta(x(1+c)) - \vartheta(x) = o(x)$$

for arbitrarily large values of x. Consider the set S of all such constants c, and let $C = \sup S$. Since $0 < a \leq A < \infty$, we have that $C < \infty$. We will show that

$$\vartheta(x(1+C)) - \vartheta(x) = o(x)$$

for arbitrarily large values of x. For this, we require the following lemma:

Lemma 5. *Whenever $x < y$, we have that $\vartheta(y) - \vartheta(x) = 2(y-x) + o(y)$.*

Proof. By the Selberg symmetry formula (9), we have that

$$\sum_{\substack{\text{prime } p \\ x < p \leq y}} (\log p)^2 \leq 2(y-x)\log y + o(y \log y).$$

First suppose that $x \geq y/\log^2 y$. In this case, we have that $\log x = (1+o(1))\log y$, so for all primes p such that $x < p \leq y$, we have that $\log p = (1+o(1))\log y$. Dividing the above equality by $\log y$ on both sides hen yields the lemma. Now, suppose that $x < y/\log^2 y$. In this case, we have the following results:

$$\begin{aligned}
\vartheta(y) - \vartheta(x) &= \vartheta(y) - \vartheta\left(\frac{y}{\log^2 y}\right) + \vartheta\left(\frac{y}{\log^2 y}\right) - \vartheta(x) \\
&< \vartheta(y) - \vartheta\left(\frac{y}{\log^2 y}\right) + o(y) \\
&< 2\left(y - \frac{y}{\log^2 y}\right) + o(y) \\
&= 2(y-x) + o(y),
\end{aligned}$$

where we used the fact that $x = o(y)$ and where we applied the result of the first case to deduce the second-to-last equality. □

Let $\varepsilon > 0$, and take $c > C - \frac{\varepsilon}{2}$. As $x \to \infty$, running through the values where

$$\vartheta(x(1+c)) - \vartheta(x) = o(x),$$

we have by Lemma 5 that

$$\begin{aligned}
\vartheta(x(1+C)) - \vartheta(x) &= \vartheta(x(1+C)) - \vartheta(x(1+c)) + \vartheta(x(1+c)) - \vartheta(x) \\
&\leq 2(C-c)x + o(x) \\
&= \varepsilon x + o(x),
\end{aligned}$$

and since our choice of $\varepsilon > 0$ was arbitrary, we have that

$$\vartheta(x(1+C)) - \vartheta(x) = o(x),$$

as desired. Now, from the Selberg symmetry formula (9), we have that

$$\sum_{\substack{\text{prime } p \\ x<p\leq x(1+C)}} \log^2 p + \sum_{\substack{\text{prime } p,q \\ x<pq\leq x(1+C)}} \log p \log q = 2Cx\log x + o(x\log x).$$

Now, since for all primes $p \in (x, x(1+C)]$ we have that $\log p < \log(x(1+C))$, we have that

$$\sum_{\substack{\text{prime } p \\ x<p\leq x(1+C)}} \log^2 p \leq (\vartheta(x(1+C)) - \vartheta(x))\log(x(1+C)) = o(x\log x).$$

Combining the previous two results, we deduce that

$$\sum_{\substack{\text{prime } p \\ p\leq x(1+C)}} \left(\vartheta\left(\frac{x}{p}(1+C)\right) - \vartheta\left(\frac{x}{p}\right)\right)\log p = 2Cx\log x + o(x\log x). \tag{10}$$

We now require the following lemma:

Lemma 6. *Consider the limit as $x \to \infty$, running through the values where*

$$\vartheta(x(1+C)) - \vartheta(x) = o(x).$$

Then, we have that

$$\vartheta\left(\frac{x}{p}(1+C)\right)-\vartheta\left(\frac{x}{p}\right)=2C\frac{x}{p}+o\left(\frac{x}{p}\right)$$

for all primes $p \le x(1+C)$ outside a set $\mathcal{P}$ of primes $p \le x(1+C)$ satisfying

$$\sum_{p\in\mathcal{P}}\frac{\log p}{p}=o(\log x).$$

Proof. We proceed by contradiction. If the lemma is false, then there exist constants $b_1, b_2 > 0$ such that

$$\vartheta\left(\frac{x}{p}(1+C)\right)-\vartheta\left(\frac{x}{p}\right)<(2C-b_1)\frac{x}{p} \text{ for all } p\in\mathcal{P} \text{ and } \sum_{p\in\mathcal{P}}\frac{\log p}{p}\sim b_2\log x.$$

Using the well-known elementary estimate

$$\sum_{\substack{\text{prime } p\\ p\le x}}\frac{\log p}{p}=(1+o(1))\log x$$

along with the result of Lemma 5, we deduce that

$$\sum_{\substack{\text{prime } p\\ p\le x(1+C)}}\left(\vartheta\left(\frac{x}{p}(1+C)\right)-\vartheta\left(\frac{x}{p}\right)\right)\log p=$$

$$\sum_{p\in\mathcal{P}}\left(\vartheta\left(\frac{x}{p}(1+C)\right)-\vartheta\left(\frac{x}{p}\right)\right)\log p+\sum_{\substack{\text{prime } p\\ p\le x(1+C),p\notin\mathcal{P}}}\left(\vartheta\left(\frac{x}{p}(1+C)\right)-\vartheta\left(\frac{x}{p}\right)\right)\log p$$

$$\le b_2(2C-b_1)x\log x+2C(1-b_2)x\log x+o(x\log x)$$

$$=(2C-b_1b_2)x\log+o(x\log x).$$

But the above result is a contradiction to (10), so we conclude that the lemma is true. □

Given x, if a prime p satisfies

$$\vartheta\left(\frac{x}{p}(1+C)\right)-\vartheta\left(\frac{x}{p}\right)=2C\frac{x}{p}+o\left(\frac{x}{p}\right),$$

we will say that p is "good," and otherwise we will say that p is "bad." Now, assume that there exists a list of good primes $p_1 < \cdots < p_k$ such that

$$10p_1 < p_k < 100p_1 \text{ and } (1+C)(1+t)^2p_i > p_{i+1} > (1+t)p_i \text{ for all } i\in\{1,\ldots,k-1\},$$

where $t > 0$ is small compared to C. Consider the intervals I_i defined for each i by

$$I_i=\left[\frac{x}{p_i},\frac{x}{p_i}(1+C)\right].$$

If $I_i \cap I_{i+1} \neq \varnothing$ for some I, then by assumption we have that
$$\frac{x}{p_{i+1}}(1+t) < \frac{x}{p_i} < \frac{x}{p_{i+1}}(1+C).$$
We now claim that
$$\vartheta\left(\frac{x}{p_i}\right) - \vartheta\left(\frac{x}{p_{i+1}}\right) = 2\left(\frac{x}{p_i} - \frac{x}{p_{i+1}}\right) + o\left(\frac{x}{p_i}\right).$$
If the above equality does not hold, then
$$\vartheta\left(\frac{x}{p_i}\right) - \vartheta\left(\frac{x}{p_{i+1}}\right) < (2-c_1)\left(\frac{x}{p_i} - \frac{x}{p_{i+1}}\right)$$
for some constant $c_1 > 0$, and combining this result with the fact that p_{i+1} was taken to be good yields that
$$\vartheta\left(\frac{x}{p_{i+1}}(1+C)\right) - \vartheta\left(\frac{x}{p_i}\right) > (2+c_2)\left(\frac{x}{p_{i+1}}(1+C) - \frac{x}{p_i}\right)$$
for some constant c_2, which contradicts the result of Lemma 5. Thus, the claimed equality above must hold. Combining this equality with the fact that p_i was taken to be good yields that
$$\vartheta\left(\frac{x}{p_i}(1+C)\right) - \vartheta\left(\frac{x}{p_{i+1}}\right) = 2\left(\frac{x}{p_i}(1+C)\right) - \frac{x}{p_{i+1}} + o\left(\frac{x}{p_i}\right). \tag{11}$$
If $I_i \cap I_{i+1} = \varnothing$ for some i, then it is easy to see from the assumptions made on p_i, p_{i+1} and the fact that t is small compared to C that we have
$$\vartheta\left(\frac{x}{p_i}(1+C)\right) - \vartheta\left(\frac{x}{p_{i+1}}\right) > 1.9\left(\frac{x}{p_i}(1+C) - \frac{x}{p_{i+1}}\right). \tag{12}$$
Now, we run through the list $i = 1, \dots, k-1$. If $I_i \cap I_{i+1} \neq \varnothing$, we pick (11), and if $I_i \cap I_{i+1} = \varnothing$, we pick (12). Summing all of the picked (in)equalities yields
$$\vartheta\left(\frac{x}{p_1}(1+C)\right) - \vartheta\left(\frac{x}{p_k}\right) > 1.9\left(\frac{x}{p_1}(1+C) - \frac{x}{p_k}\right).$$
Throwing away the term $\vartheta\left(\frac{x}{p_k}\right)$ yields that
$$\vartheta\left(\frac{x}{p_1}(1+C)\right) > 1.9\left(\frac{x}{p_1}(1+C) - \frac{x}{p_k}\right).$$
We now claim that the subtracted term on the right-hand-side of the above equality is not very large. By assumption, we have that $p_k > 10p_1$, so
$$-1.9\frac{x}{p_k} > -0.19\frac{x}{p_1} > -0.19\frac{x}{p_1}(1+C).$$
Combining our results, we find that
$$\vartheta\left(\frac{x}{p_1}(1+C)\right) > 1.6\frac{x}{p_1}(1+C). \tag{13}$$

Since (13) holds for arbitrarily large values of x, we must have that $A \geq 1.6$. But, recalling the well-known elementary fact that $A \leq 1.5$, we find that we have a contradiction. Thus, the proposition is true.

It remains to show that there exists a list of good primes $p_1 < \cdots < p_k$ satisfying the following conditions:

$$10p_1 < p_k < 100p_1 \text{ and } (1+C)(1+t)^2 p_i > p_{i+1} > (1+t)p_i \text{ for all } i \in \{1, \ldots, k-1\},$$

where $t > 0$ is small compared to C. Fix B large, and for each

$$r \in \left\{0, 1, 2, \ldots, \left\lfloor \frac{\log x}{2 \log B} \right\rfloor - 1\right\},$$

let $I_r = (B^{2r}, B^{2r+1})$. It is clear from the construction that $I_r \subset (0, x)$ for all r. We will first show that I_r contains a good prime for all but $o(\log x)$-many values of r. Since $0 < a < A < \infty$, we have that by taking B sufficiently large, we can make $\vartheta(Bx) - \vartheta(x) > cx$, so

$$\sum_{\substack{\text{prime } p \\ p \in I_r}} \frac{\log p}{p} > c_1,$$

where c_1 does not depend on our choice of r. Now, if we had $(c_2 \log x)$-many intervals I_r not containing any good primes, we would have that

$$\sum_{\text{bad } p} \frac{\log p}{p} > c_1 c_2 \log x,$$

and this is a contradiction to the result of Lemma 6. Now if r is such that I_r contains good primes, let $p_1(r)$ be the smallest good prime in I_r, and suppose that we have a list $p_1(r) < p_2(r) < \cdots < p_i(r)$ of good primes such that

$$10p_1(r) < p_i(r) < 100p_1(r) \text{ and } (1+C)(1+t)^2 p_j(r) > p_{j+1}(r) > (1+t)p_j(r)$$

for all $j \in \{1, \ldots, i-1\}$ and such that this list cannot be extended to include any $p_{i+1}(r) > p_i(r)$. It follows that all primes in the interval

$$J_r(i) = \left[p_i(r)(1+t), p_i(r)(1+t)^2(1+C)\right] \subset I_r$$

are bad, so (just as we deduced before in the proof that most intervals I_r contain good primes) we have that

$$\sum_{p \in J_r(i)} \frac{\log p}{p} > \eta,$$

where η is an absolute constant. Now consider the collection of intervals $J_r(i)$, ranging over all values of r such that I_r contains a good prime. The number of such intervals is at least half the total number of values of r, since I_r contains no good primes for only $o(\log x)$-many values of r. Thus, the number of such intervals is at least $\frac{\log x}{4\log B}$. If all of the intervals $J_r(i)$ are disjoint, we would have

$$\sum_{\text{bad } p} \frac{\log p}{p} \geq \sum_{\substack{p\in J_r(i) \\ \text{for some } r}} \frac{\log p}{p} > \eta \cdot \frac{\log x}{4\log B},$$

which contradicts the result of Lemma 6. It therefore suffices to show that the intervals $J_r(i)$ are all distinct. By taking $B > 100$, we find that the upper endpoint of $J_r(i)$, namely $p_i(r)(1+t)^2(1+C)$, is bounded above by B^{2r+2}, which implies that the upper endpoint of $J_r(i)$ is less than the lower endpoint of I_{r+1}. Since the lower endpoint of $J_{r+1}(i')$ (we have to pick a different i' corresponding to $r+1$) is larger than the lower endpoint of I_{r+1}, we find that the intervals $J_r(i)$ are in fact disjoint, as desired. □

Within two days of learning about Erdős' result in Proposition 4, Selberg managed to prove the PNT ([Erd49]). We detail his proof as follows:

Proof of the PNT (from [Erd49]*).* We will require three lemmas:

Lemma 7. *We have that $a + A = 2$.*

Proof. By the definition of A, we may choose large x such that $\vartheta(x) = Ax + o(x)$. It follows easily that

$$\sum_{\substack{\text{prime } p \\ p\leq x}} \log^2 p \leq Ax\log x + o(x\log x).$$

Now, by subtracting the above inequality from the symmetry formula (9), we obtain the following result:

$$\sum_{\substack{\text{prime } p,q \\ pq\leq x}} \log p \log q = \sum_{\substack{\text{prime } p \\ p\leq x}} \vartheta\left(\frac{x}{p}\right)\log p = (2-A)x\log x + o(x\log x).$$

We now recall the following well-known elementary fact:

$$\sum_{\substack{\text{prime } p \\ p\leq x}} \frac{\log p}{p} = (1+o(1))\log x.$$

Combining the above fact with the definition of a yields that

$$\sum_{\substack{\text{prime } p \\ p\leq x}} \vartheta\left(\frac{x}{p}\right)\log p \geq \sum_{\substack{\text{prime } p \\ p\leq x}} \left(\frac{ax}{p}\right)\log p = ax\log x + o(x\log x).$$

We then have that

$$\begin{aligned} 0 &= \sum_{\substack{\text{prime } p \\ p\le x}} \vartheta\left(\frac{x}{p}\right)\log p - \sum_{\substack{\text{prime } p \\ p\le x}} \vartheta\left(\frac{x}{p}\right)\log p \\ &\le (2-A)x\log x + o(x\log x) - ax\log x + o(x\log x) \\ &= (2-a-A)x\log x + o(x\log x), \end{aligned}$$

from which we deduce that $2-a-A\ge 0$, so $a+A\le 2$. By repeating the above argument with large x chosen so that $\vartheta(x)=ax+o(x)$, we find that $a+A\ge 2$. It follows that $a+A=2$, as desired. □

Lemma 8. *Choose large x such that $\vartheta(x)=Ax+o(x)$. Then,*

$$\vartheta\left(\frac{x}{p'}\right) = a\frac{x}{p'} + o\left(\frac{x}{p'}\right)$$

for all primes p' outside a set $\mathcal{P}$ of primes $p\le x$ satisfying

$$\sum_{p\in\mathcal{P}} \frac{\log p}{p} = o(\log x).$$

Proof. We proceed by contradiction. If the lemma is false, then there exist $b_1, b_2 > 0$ and a set $\mathcal{P}$ of primes $p\le x$ such that for all $p\in\mathcal{P}$ we have

$$\vartheta\left(\frac{x}{p}\right) > (a+b_1)\frac{x}{p} \text{ and } \sum_{p\in\mathcal{P}}\frac{\log p}{p} > b_2\log x.$$

Now, as in the proof of Lemma 7, recall that for our choice of x we have

$$\sum_{\substack{\text{prime } p \\ p\le x}} \vartheta\left(\frac{x}{p}\right)\log p = (2-A)x\log x + o(x\log x) = ax\log x + o(x\log x),$$

where the last step is an application of Lemma 7. Combining our results, we have the following inequalities:

$$\begin{aligned} ax\log x + o(x\log x) &= \sum_{\substack{\text{prime } p \\ p\le x}} \vartheta\left(\frac{x}{p}\right)\log p \\ &= \sum_{p\in\mathcal{P}} \vartheta\left(\frac{x}{p}\right)\log p + \sum_{\substack{\text{prime } p \\ p\notin\mathcal{P}, p\le x}} \vartheta\left(\frac{x}{p}\right)\log p \\ &> b_2(a+b_1)x\log x + (1-b_2)ax\log x + o(x\log x) \\ &= ax\log x + b_1b_2x\log x + o(x\log x), \end{aligned}$$

where, in the second-to-last step, we used the aforementioned fact that

$$\sum_{\substack{\text{prime } p \\ p \leq x}} \frac{\log p}{p} = (1 + o(1)) \log x.$$

We now have that

$$ax \log x + o(x \log x) > ax \log x + b_1 b_2 x \log x + o(x \log x),$$

which is clearly a contradiction. It follows that the lemma is true. □

Lemma 9. *Consider the set of primes p' such that*

$$\vartheta\left(\frac{x}{p'}\right) = a\frac{x}{p'} + o\left(\frac{x}{p'}\right),$$

and let p_1 be the smallest member of this set. Then $p_1 < x^{\varepsilon}$ for every $\varepsilon > 0$, and

$$\vartheta\left(\frac{x}{p_1 p'}\right) = A\frac{x}{p_1 p'} + o\left(\frac{x}{p_1 p'}\right)$$

for all primes p' outside a set $\mathcal{P}$ of primes $p \leq x$ satisfying

$$\sum_{p \in \mathcal{P}} \frac{\log p}{p} = o(\log x).$$

Proof. Subtracting the estimates

$$\sum_{\substack{\text{prime } p \\ p \leq x}} \frac{\log p}{p} = (1 + o(1)) \log x \text{ and } \sum_{p \in \mathcal{P}} \frac{\log p}{p} = o(\log x)$$

yields that $p_1 < x^{\varepsilon}$ for every $\varepsilon > 0$. If we now take the proof of Lemma 8, replace x with x/p_1, and switch a with A, we obtain the second part of the lemma by the same argument. □

Having proven Lemmas 7, 8, and 9, we will now prove the PNT using Proposition 4. Take p_1 as in Lemma 9, let p' be any prime such that

$$\vartheta\left(\frac{x}{p'}\right) = a\frac{x}{p'} + o\left(\frac{x}{p'}\right),$$

and let $p'' < x/p_1$. Assume, for the sake of contradiction, that $\frac{x}{p_1 p''} < \frac{x}{p'}$. Furthermore, let $\delta \in (0, A/a - 1)$, and let the closed interval I be defined by

$$I = \left[\frac{p'}{p_1}, \frac{p'}{p_1}\left(\frac{A}{a} - \delta\right)\right].$$

Notice that if $p'' \in I$, then we have

$$a\frac{x}{p'} + o\left(\frac{x}{p'}\right) \leq A\frac{x}{p_1 p'} + o\left(\frac{x}{p_1 p'}\right).$$

But for any prime p'' satisfying the equality

$$\vartheta\left(\frac{x}{p_1p'}\right) = A\frac{x}{p_1p'} + o\left(\frac{x}{p_1p'}\right),$$

we have by the monotonicity of ϑ that

$$a\frac{x}{p'} + o\left(\frac{x}{p'}\right) = \vartheta\left(\frac{x}{p'}\right) \geq \vartheta\left(\frac{x}{p_1p''}\right) = A\frac{x}{p_1p'} + o\left(\frac{x}{p_1p'}\right).$$

It follows that none of the primes $p'' \in I$ satisfy the equality

$$\vartheta\left(\frac{x}{p_1p'}\right) = A\frac{x}{p_1p'} + o\left(\frac{x}{p_1p'}\right).$$

Now, observe that we have the following inequalities:

$$\begin{aligned}
\sum_{\substack{\text{prime } p \\ p\in I}} \frac{\log p}{p} &\geq \left[\pi\left(\frac{p'}{p_1}\left(\frac{A}{a} - \delta\right)\right) - \pi\left(\frac{p'}{p_1}\right)\right] \frac{\log\left(\frac{p'}{p_1}\right)}{\left(\frac{p'}{p_1}\right)} \\
&> \eta \cdot \frac{\left(\frac{p'}{p_1}\right)}{\log\left(\frac{p'}{p_1}\right)} \cdot \frac{\log\left(\frac{p'}{p_1}\right)}{\left(\frac{p'}{p_1}\right)} \\
&= \eta \ (\eta \text{ is an absolute constant})
\end{aligned}$$

where the first inequality is a trivial bound and the second inequality follows from Proposition 4. As was done in the proof of Proposition 4, it is possible to construct $(c\log x)$-many such disjoint intervals I, and if $\mathcal{P}$ denotes the set of all primes in the union of these intervals, we find that

$$\sum_{p\in\mathcal{P}} \frac{\log p}{p} > c' \log x$$

for some constant c'. This result contradicts the second part of Lemma 9, so we must have that $\frac{x}{p_1p''} \geq \frac{x}{p'}$. It is entirely possible to choose $p' = p_1$, from which we deduce that $p'' \leq 1$. It follows from the second part of Lemma 9 that

$$\vartheta\left(\frac{x}{p_1}\right) = A\frac{x}{p_1} + o\left(\frac{x}{p_1}\right),$$

and by assumption we have that

$$\vartheta\left(\frac{x}{p'}\right) = \vartheta\left(\frac{x}{p_1}\right) = a\frac{x}{p_1} + o\left(\frac{x}{p_1}\right).$$

Adding the above two equalities, we find that

$$2\vartheta\left(\frac{x}{p_1}\right) = (a+A)\frac{x}{p_1} + o\left(\frac{x}{p_1}\right) = 2\frac{x}{p_1} + o\left(\frac{x}{p_1}\right),$$

and dividing by 2 on both sides yields that

$$\vartheta\left(\frac{x}{p_1}\right) = \frac{x}{p_1} + o\left(\frac{x}{p_1}\right).$$

By the first part of Lemma 9, we have that $p_1 < x^\varepsilon$ for every $\varepsilon > 0$, so from the above equality, we deduce that

$$\lim_{x\to\infty} \frac{\vartheta(x)}{x} = 1 \Leftrightarrow \vartheta(x) \sim x.$$

Recalling that $\pi(x) \sim x/\log x \Leftrightarrow \vartheta(x) \sim x$, we observe that we have completed the proof of the PNT. □

The above proof of the PNT is the first elementary proof, but it is not the easiest elementary proof. Just days after the above proof was discovered, Selberg and Erdős jointly managed to simplify the above arguments to a large extent ([Erd49]). Nonetheless, these simplified arguments still depend on the two key ideas discussed in this article: the Selberg symmetry formula (2) and Erdős' result in Proposition 4. For Selberg's version of the first elementary proof detailed above, please refer to [Sel49].

References

[Bal06] Steve Balady. Annotation: A discussion of the fundamental ideas behind Selberg's "An elementary proof of the prime-number theorem". August 2006.

[Erd49] Paul Erdős. On a new method in elementary number theory which leads to an elementary proof of the prime number theorem. *Proceedings of the National Academy of Sciences*, 35(7):374–384, July 1949.

[Sel49] Atle Selberg. An elementary proof of the prime-number theorem. *Annals of Mathematics, Second Series*, 50(2):305–313, April 1949.

[Tao07] Terence Tao. Open question: the parity problem in sieve theory. Posted on "What's new" weblog, June 2007.

Department of Mathematics, Harvard College, 1 Oxford Street, Cambridge, MA 02138

E-mail address: `aaswaminathan@college.harvard.edu`