THE PRIME NUMBER THEOREM VIA THE LARGE SIEVE

ADOLF HILDEBRAND

§1. *Introduction.* In the last three decades there appeared a number of elementary proofs of the prime number theorem (PNT) in the literature (see [3] for a survey). Most of these proofs are based, at least in part, on ideas from the original proof by Erdős [5] and Selberg [12]. In particular, one of the main ingredients of the Erdős-Selberg proof, Selberg's formula

$$\sum_{p\leqslant x} \log^2 p + \sum_{pq\leqslant x} \log p \log q = 2x \log x + O(x) \tag{1}$$

(where p and q run through primes) appears, in some form, in almost all these proofs.

Several authors [**1, 2, 8, 10**] have given direct elementary proofs of the relation (μ being the Moebius function)

$$M(x) = \frac{1}{x} \sum_{n\leqslant x} \mu(n) = o(1) \qquad (x\to\infty), \tag{2}$$

which is known to be "equivalent" to the PNT. With the exception of a recent proof by Daboussi [**2**], these proofs are based on identities such as

$$M(x) \log x = -\sum_{p\leqslant x} \frac{\log p}{p} M\left(\frac{x}{p}\right) + O(1), \tag{3}$$

which are similar in structure of Selberg's formula (1) and play the same role in the proof of (2) as Selberg's formula does in the direct proof of the PNT. Daboussi's elementary proof of (2) does not use Selberg's formula (1) nor its analogue (3), and constitutes in fact the first elementary proof of the PNT, which is fundamentally different from the original Erdős-Selberg proof.

We shall give here a new elementary proof of (2), again without using Selberg type formulae. Its main feature is the application of a large sieve type inequality in order to show that $M(x)$ varies slowly over fairly large intervals. More precisely, we shall show in Lemma 4 below that $M(x)-M(x')\to 0$ as $x\to\infty$, uniformly for $x\leqslant x'\leqslant x^{1+\eta(x)}$, provided $\eta(x)\to 0$. Relation (2) is an almost immediate consequence of this result.

The central idea of the proof, the application of the large sieve, has already been exploited in [7] to obtain a new proof for Wirsing's mean value theorem. This result contains (2) as a special case, but since the proof in [7] made use of the PNT, it did not yield a new proof of the PNT. Our main task here, which will take up the largest part of the proof, is to eliminate the application of the PNT and replace it by the elementary prime number estimates from Mertens' theory, together with a sieve upper bound for primes in short intervals.

[MATHEMATIKA, **33** (1986), 23–30]

We remark that it is possible to prove, by the same method, a quantitative version of (2), namely

$$M(x) \ll (\log\log x)^{-1/2} \qquad (x \geqslant 3).$$

Since this result is relatively weak in comparison with the known elementary error term estimates for the PNT, and in order to keep the proof more transparent, we shall confine ourselves to the proof of the asymptotic relation (2).

§2. *Lemmas.* In the first lemma, we state the mentioned large sieve type inequality. It is due to Elliott [**4**, Lemma 4.7], who gave a simple and elementary proof for it via the Turán–Kubilius inequality.

LEMMA 1. *There exists an absolute constant $C > 0$ such that for all complex numbers a_n, $1 \leqslant n \leqslant x$, the inequality*

$$\sum_{p \leqslant x} \frac{1}{p} \left| \frac{p}{x} \sum_{\substack{n \leqslant x \\ p \mid n}} a_n - \frac{1}{x} \sum_{n \leqslant x} a_n \right|^2 \leqslant C \frac{1}{x} \sum_{n \leqslant x} |a_n|^2$$

holds.

The next lemma gives an upper bound for primes in short intervals. The result is, in a stronger form, a standard consequence of Selberg's upper bound sieve, see, *e.g.*, [**6**, Theorem 3.7]. It may also be deduced from Montgomery's version of the large sieve [**9**, Corollary 3.2] or from Selberg's formula (1).

LEMMA 2. *The estimate*

$$\pi(x+y) - \pi(x) \leqslant (2 + o(1)) \frac{y}{\log y}$$

holds, as $y \to \infty$, uniformly for $x \geqslant 1$.

COROLLARY. *The estimate*

$$\sum_{x < p \leqslant x+y} \frac{\log p}{p} \leqslant (2 + o(1)) \log \frac{x+y}{x}$$

holds, as $y \to \infty$, uniformly for $x \geqslant y$.

Proof. Dividing the interval $(x, x+y]$ into subintervals of length y/k, where $k = k(y)$ tends to infinity sufficiently slowly, as $y \to \infty$, the result follows from Lemma 2.

LEMMA 3. *Let $0 < \varepsilon \leqslant 1$ be given. For any $x' \geqslant x \geqslant 2$, there exists a number λ, $1 \leqslant \lambda \leqslant \lambda_0$ such that the inequality*

$$\sum_{y < p \leqslant y(1+\varepsilon)} \frac{\log p}{p} \geqslant \delta \tag{4}$$

holds both for $y=\lambda x$ and for $y=\lambda x'$. Here $\delta=\delta(\varepsilon)>0$ and $\lambda_0=\lambda_0(\varepsilon)\geqslant 1$ are constants depending only on ε.

Remark. In the course of his proof of the PNT, Erdős [**5**] established (4) (in an equivalent form) for *all* sufficiently large y. His proof relied on Selberg's formula (1). We need however only the weaker result stated in the lemma, the proof of which is simpler and does not require Selberg's formula.

Proof. Given $0<\varepsilon\leqslant 1$ and $x\geqslant 2$, put $\varepsilon_1=\varepsilon/3$ and

$$x_i=x(1+\varepsilon_1)^i \qquad (i\geqslant 0).$$

Fix a number $\delta>0$ and let I be the set of indices $i\geqslant 0$, for which

$$\sum_{x_i<p\leqslant x_{i+1}}\frac{\log p}{p}\geqslant\delta$$

holds. Moreover, put

$$\bar I=\{i\geqslant 0:\quad i\in I \quad\text{or}\quad i+1\in I\}.$$

Since

$$x_i<x_{i+1}<x_{i+2}=(1+\varepsilon_1)^2x_i\leqslant(1+\varepsilon)x_i \qquad (i\geqslant 0),$$

we see that, for every $i\in\bar I$, (4) holds with $y=x_i$. We shall show that if δ is sufficiently small in terms of ε, then for some i_0, depending only on ε, $\bar I$ contains more than $i_0/2$ indices $i<i_0$. From this assertion it follows that any two sets $\bar I$ and $\bar I'$, defined as before with respect to given numbers $x'\geqslant x\geqslant 2$, have a common element, say i, in the interval $[0, i_0)$, and in view of the above remark we obtain the conclusion of the lemma with $\lambda=(1+\varepsilon_1)^i$ and $\lambda_0=(1+\varepsilon_1)^{i_0}$.

To prove our claim, let $i_0\geqslant 1$ be for the moment unspecified and denote by N and $\bar N$ the cardinalities of $I\cap[0, i_0)$ and $\bar I\cap[0, i_0)$ respectively. From Mertens' formula

$$\sum_{p\leqslant y}\frac{\log p}{p}=\log y+O(1) \qquad (y\geqslant 2)$$

and the Corollary to Lemma 2 we get

$$\begin{aligned}\log(1+\varepsilon_1)^{i_0}+O(1)&=\sum_{x<p\leqslant x(1+\varepsilon_1)^{i_0}}\frac{\log p}{p}=\sum_{0\leqslant i<i_0}\sum_{x_i<p\leqslant x_{i+1}}\frac{\log p}{p}\\&\leqslant 2N\log(1+\varepsilon_1)+o(i_0)+(i_0-N)\delta,\end{aligned}$$

whence

$$N\geqslant i_0\left(\tfrac12+o(1)-\frac{\delta}{2\log(1+\varepsilon_1)}+O\left(\frac{1}{i_0\log(1+\varepsilon_1)}\right)\right). \tag{5}$$

(Here the "o"-notation refers to the limiting behaviour as $i_0\to\infty$, and is uniform with respect to $x\geqslant 2$). For $\delta\leqslant\varepsilon/10$ and sufficiently large i_0 the right-hand side of (6) becomes positive, and we now fix such an index $i_0=i_0(\varepsilon)$. Thus the interval $[0, i_0)$ contains an element of I, and replacing $x=x_0$ by x_i, $i\geqslant 1$, we see that every interval on the positive real axis of length i_0 contains an element of I.

Next, let $i'_0 > 6i_0$ and define N' and $\bar{N}'$, as before with respect to i'_0. If now

$$\bar{N}' - N' \leqslant \frac{i'_0}{3i_0},$$

then there are at most $(i'_0/3i_0)+1$ "blocks" of consecutive indices $i \leqslant i'_0$ not belonging to I. Since by the above remark each of these blocks has length $\leqslant i_0$, it follows that

$$\bar{N}' \geqslant N' \geqslant i'_0 - \tfrac{1}{3}i'_0 - i_0 > i'_0/2.$$

But if

$$\bar{N}' - N' > \frac{i'_0}{3i_0},$$

then we have by (5) (with i'_0 and N' in place of i_0 and N)

$$\bar{N}' \geqslant N' + \frac{i'_0}{3i_0} \geqslant i'_0\left(\tfrac{1}{2} + o(1) - \frac{\delta}{2\log(1+\varepsilon_1)} + \frac{1}{3i_0} + O\left(\frac{1}{i'_0 \log(1+\varepsilon_1)}\right)\right),$$

and choosing $\delta = \delta(\varepsilon)$ sufficiently small and i'_0 sufficiently large, we get again $\bar{N}' > i'_1/2$.

This proves our initial claim and hence the lemma.

§3. *Proof of the estimate* (2).

The proof is based on the following lemma.

LEMMA 4. *Let $\eta(x)$ be a non-negative function tending to zero as x tends to infinity. Then the relation*

$$M(x') = M(x) + o(1)$$

holds, as $x \to \infty$, uniformly for $x \leqslant x' \leqslant x^{1+\eta(x)}$.

The estimate (2) is an almost immediate consequence of this result. In fact, with $\eta = \eta(x) = (\log x)^{-1/2}$, the lemma yields, as $x \to \infty$,

$$\begin{aligned} M(x) &= \frac{1}{\eta \log x} \int_x^{x^{1+\eta}} \frac{M(x')}{x}\, dx' + o(1) \\ &= \frac{1}{\eta \log x} \sum_{n \leqslant x^{1+\eta}} \mu(n)\left(\frac{1}{\max(n,x)} - x^{-1-\eta}\right) + o(1) \\ &= \frac{1}{\sqrt{\log x}} \sum_{x < n \leqslant x^{1+\eta}} \frac{\mu(n)}{n} + o(1), \end{aligned}$$

and in view of the well-known elementary estimate

$$\sum_{n \leqslant y} \frac{\mu(n)}{n} = O(1) \qquad (y \geqslant 1)$$

(2) follows.

Proof of Lemma 4. Fix $0<\varepsilon\leqslant 1$. We shall show that if $\eta>0$ is sufficiently small in terms of ε, then we have

$$|M(x)-M(x')|\ll\varepsilon \tag{6}$$

uniformly for all sufficiently large x and $x\leqslant x'\leqslant x^{1+\eta}$. By letting $\varepsilon\to 0$, this implies the assertion of the lemma.

Let now $0<\eta\leqslant\frac{1}{2}$ and fix $2\leqslant x\leqslant x'\leqslant x^{1+\eta}$. We apply Lemma 1 with $a_n=\mu(n)$. Noting that, for all $p\leqslant x$,

$$\frac{p}{x}\sum_{\substack{n\leqslant x\\ p\mid n}}\mu(n)=\frac{p}{x}\mu(p)\sum_{\substack{n\leqslant x/p\\ p\nmid n}}\mu(n)+O\left(\frac{1}{p}\right)=-M\left(\frac{x}{p}\right)+O\left(\frac{1}{p}\right),$$

we obtain

$$\sum_{p\leqslant x}\frac{1}{p}\left|M(x)+M\left(\frac{x}{p}\right)\right|^2\ll 1.$$

By the Cauchy–Schwarz inequality it follows that, for any set $\mathscr{P}$ of primes $\leqslant x$,

$$\begin{aligned}\left(\sum_{p\in\mathscr{P}}\frac{1}{p}\right)M(x)&=-\sum_{p\in\mathscr{P}}\frac{M(x/p)}{p}+O\left(\sum_{p\in\mathscr{P}}\frac{1}{p}\left|M(x)+M\left(\frac{x}{p}\right)\right|\right)\\&=-\sum_{p\in\mathscr{P}}\frac{M(x/p)}{p}+O\left(\left(\sum_{p\in\mathscr{P}}\frac{1}{p}\right)^{1/2}\right).\end{aligned}\tag{7}$$

An analogous identity, $(7)'$ say, holds with x' instead of x and an arbitrary set $\mathscr{P}'$ of primes $\leqslant x'$. We shall show that, with an appropriate choice of the sets $\mathscr{P}$ and $\mathscr{P}'$, the right-hand sides of (7) and $(7)'$ nearly cancel each other, and $\mathscr{P}$ and $\mathscr{P}'$ have approximately the same (large) sum of reciprocals. This will lead to the desired estimate (6).

Put

$$x_0=x^{\sqrt{\eta}},\qquad x_0'=x_0(x'/x).$$

By a repeated application of Lemma 3 we obtain a sequence $(x_j)_{j\geqslant 1}$, satisfying

$$x_j(1+\varepsilon)\leqslant x_{j+1}\leqslant\lambda_0x_j(1+\varepsilon)\qquad(j\geqslant 0),$$

such that (4) holds for $y=x_j$ and $y=x_j'=x_j(x'/x)$ and all $j\geqslant 1$, where $\lambda_0=\lambda_0(\varepsilon)$ and $\delta=\delta(\varepsilon)$ are the constants of Lemma 3. Define j_0 by $x_{j_0}\leqslant x<x_{j_0}+1$; note that $j_0\geqslant 2$, if x is sufficiently large, as we may assume. Put

$$I_j=(x_j,x_j(1+\varepsilon)],\qquad I_j'=(x_j',x_j'(1+\varepsilon)].$$

By construction, the intervals I_j, $1\leqslant j<j_0$, and I_j', $1\leqslant j<j_0$, are pairwise disjoint and contained in $(x_0,x]$ and $(x_0',x']$, respectively. Now choose sets of primes

$$\mathscr{P}\subset\bigcup_{1\leqslant j<j_0}I_j\quad\text{and}\quad\mathscr{P}'\subset\bigcup_{1\leqslant j<j_0}I_j'$$

such that, for $1\leqslant j<j_0$,

$$\left|\sum_{p\in\mathscr{P}\cap I_j}\frac{\log p}{p}-\delta\right|\leqslant\frac{\log x_j(z+\varepsilon)}{x_j},\tag{8}$$

and

$$\left| \sum_{p \in \mathscr{P}' \cap I_j'} \frac{\log p}{p} - \delta \right| \leqslant \frac{\log x_j'(1+\varepsilon)}{x_j'}. \tag{8$'$}$$

Since (4) holds with $y = x_j$ and $y = x_j'$, this can be achieved by discarding from each I_j or I_j' a suitable number of primes.

Let

$$S = \sum_{p \in \mathscr{P}} \frac{1}{p}, \qquad S' = \sum_{p \in \mathscr{P}'} \frac{1}{p}.$$

From (8) and (8)$'$ we get

$$\begin{aligned} S &= \left(1 + O\left(\frac{1}{\log x_0}\right)\right) \sum_{1 \leqslant j < j_0} \frac{1}{\log x_j} \sum_{p \in \mathscr{P} \cap I_j} \frac{\log p}{p} \\ &= \left(1 + O\left(\frac{1}{\delta \log x_0}\right)\right) \sum_{1 \leqslant j < j_0} \frac{\delta}{\log x_j}, \\ S' &= \left(1 + O\left(\frac{1}{\delta \log x_0}\right)\right) \sum_{1 \leqslant j < j_0} \frac{\delta}{\log x_j'}. \end{aligned}$$

Since

$$0 < \frac{\log x_j' - \log x_j}{\log x_j} = \frac{\log (x'/x)}{\log x_j} \leqslant \frac{\eta \log x}{\log x_0} = \sqrt{\eta},$$

it follows that

$$S = S'\left(1 + O(\sqrt{\eta}) + O\left(\frac{1}{\delta \log x_0}\right)\right) \tag{9}$$

and, in particular,

$$S \ll S' \ll S, \tag{9$'$}$$

if η is sufficiently small and x sufficiently large, as we may assume. Moreover, we have

$$\begin{aligned} S' \gg S \gg_\varepsilon & \sum_{1 \leqslant j < j_0} \frac{1}{\log x_j} \\ \geqslant & \sum_{1 \leqslant j < j_0} \frac{1}{\log \{x_0(\lambda_0(1+\varepsilon))^j\}} \gg_\varepsilon \log \frac{\log x}{\log x_0} = \log \frac{1}{\sqrt{\eta}}. \end{aligned} \tag{10}$$

We now subtract the identity (7)$'$ from (7). Using (8), (8)$'$, (9)$'$ and the trivial bound

$$|M(y+z) - M(y)| \leqslant \frac{z+1}{y} \qquad (y \geqslant 1,\, z \geqslant 0),$$

we obtain

$$\begin{aligned}SM(x)-S'M(x') &= -\sum_{1\le j<j_0}\left\{\sum_{p\in\mathscr{P}\cap I_j}\frac{M(x/p)}{p}-\sum_{p\in\mathscr{P}'\cap I'_j}\frac{M(x'/p)}{p}\right\}+O(\sqrt{S})\\ &= -\sum_{1\le j<j_0}\frac{M(x/x_j)}{\log x_j}\left\{\sum_{p\in\mathscr{P}\cap I_j}\frac{\log p}{p}-\sum_{p\in\mathscr{P}'\cap I'_j}\frac{\log p}{p}\right\}\\ &\quad+O\left(\varepsilon S+\frac{S}{\log x_0}+\sqrt{S}\right)\\ &= O\left(\frac{j_0}{x_0}+\varepsilon S+\frac{S}{\log x_0}+\sqrt{S}\right),\end{aligned}$$

whence, by (9),

$$|M(x)-M(x')|\ll\varepsilon+\sqrt{\eta}+\frac{1}{\delta\log x_0}+\frac{j_0}{x_0 S}+\frac{1}{\sqrt{S}}.$$

In view of (10) and the definitions of x_0 and j_0, the right-hand side is of order $O(\varepsilon)$, provided x is sufficiently large and η sufficiently small. This proves (6) and hence the lemma.

§4. *Concluding remarks.* It is interesting to compare the identity (7), on which our proof is based, with the identity (3), which has been used in previous proofs of (2). In both instances, $M(x)$ is expressed as a certain average over values $-M(x/p)$, plus an error term. The essential difference between (3) and (7) lies in the fact that in (3) this average is extended over all primes $p\le x$, while in (7) the average is taken over primes belonging to an arbitrary set $\mathscr{P}$. The possibility of choosing this set freely was crucial for our proof. It enabled us to treat the sum on the right-hand side of (7) effectively without using the prime number theorem and thus avoid the problems arising from the "quadratic" terms in identities like (1) and (3).

We deduced (7) from the large sieve inequality given by Lemma 1. Elliott established this inequality by showing that it is the "dual" of the Turán-Kubilius inequality. This approach is probably the most elementary and at the same time one of the simplest proofs of a large sieve inequality. A slightly weaker inequality than that of Lemma 1, namely with the sum being extended over the smaller range $p\le x^{1/2}$ (which would be sufficient for our purposes), follows from the standard version of the large sieve, *viz.*

$$\sum_{p\le x^{1/2}}\frac{1}{p}\sum_{r=1}^{p}\left|\frac{p}{x}\sum_{\substack{n\le x\\ n\equiv r \bmod p}}a_n-\frac{1}{x}\sum_{n\le x}a_n\right|^2\le C\frac{1}{x}\sum_{n\le x}|a_n|^2. \tag{11}$$

However, the proof of (11), as given for example in [**9**, Chapter 3], involves exponential sums and can hardly be called elementary. Renyi [**11**] proved (11) with the exponent $\frac{1}{3}$ instead of $\frac{1}{2}$ and only for the case $a_n=0$ or 1. His proof is fairly simple and elementary. It can be generalized to arbitrary complex coefficients a_n and thus provides an alternative elementary access to the large sieve inequality of Lemma 1 (with the summation extended over $p\le x^{1/3}$).

As has been pointed out to the author by G. Halász, it is also possible to deduce (7) using the Turán–Kubilius inequality itself instead of its dual, the large sieve inequality of Lemma 1. The argument, which is essentially due to I. Z. Ruzsa, runs as follows: It is easily checked that, in the notation of (7),

$$M(x)\sum_{p\in\mathscr{P}}\frac{1}{p}+\sum_{p\in\mathscr{P}}\frac{1}{p}M(x/p)=-\frac{1}{x}\sum_{n\leqslant x}\mu(n)\left(\omega_{\mathscr{P}}(n)-\sum_{p\in\mathscr{P}}\frac{1}{p}\right)+O\left(\sum_{p\in\mathscr{P}}\frac{1}{p^2}\right)$$

where

$$\omega_{\mathscr{P}}(n)=\sum_{\substack{p\mid n\\ p\in\mathscr{P}}}1.$$

By the Cauchy-Schwarz inequality and the Turán–Kubilius inequality [**4**, Lemma 4.5], the last expression is, in absolute value,

$$\ll\left(\frac{1}{x}\sum_{n\leqslant x}\left(\omega_{\mathscr{P}}(n)-\sum_{p\in\mathscr{P}}\frac{1}{p}\right)^2\right)^{1/2}+\sum_{p\in\mathscr{P}}\frac{1}{p^2}\ll\left(\sum_{p\in\mathscr{P}}\frac{1}{p}\right)^{1/2},$$

and (7) follows.

References

1. K. Corrádi. A remark on the theory of multiplicative functions. *Acta Sci. Math.* (*Szeged*), 28 (1967), 83–92.

2. H. Daboussi. Sur le Théorème des Nombres Premiers. *C. R. Acad. Sc. Paris, Série I*, 298 (1984), 161–164.

3. H. Diamond. Elementary methods in the study of the distribution of prime numbers. *Bull. Amer. Math. Soc.*, 7 (1982), 553–589.

4. P. D. T. A. Elliott. *Probabilistic Number Theory I* (Springer, New York, 1979).

5. P. Erdős. On a new method, which leads to an elementary proof of the prime number theorem. *Proc. Nat. Acad. Sci. U.S.A.*, 35 (1949), 374–384.

6. H. Halberstam and H.-E. Richert. *Sieve Methods* (Academic Press, London, 1974).

7. A. Hildebrand. On Wirsing's mean value theorem for multiplicative functions. *Bull. London Math. Soc.*, 18 (1986), 147–152.

8. M. Kalecki. A simple elementary proof of $M(x)=\sum_{n\leqslant x}\mu(n)=o(x)$. *Acta Arith.*, 13 (1967), 1–7.

9. H. L. Montgomery. *Topics in Multiplicative Number Theory. Lecture Notes in Mathematics*, Vol. 227 (Springer, Berlin, 1971).

10. A. G. Postnikov and N. P. Romanov. A simplilcation of Selberg's elementary proof of the asymptotic law of distribution of primes. *Uspehi Mat. Nauk*, 10 (1955), 75–87.

11. A. Renyi. On the large sieve of Ju. V. Linnik. *Compositio Math.*, 8 (1950), 68–75.

12. A. Selberg. An elementary proof of the prime number theorem. *Ann. Math.*, 50 (1949), 305–313.

Prof. A. Hildebrand,
Department of Mathematics,
University of Illinois,
1409 West Green Street,
Urbana, Illinois 61801,
U.S.A.

10H15: *NUMBER THEORY; Multiplicative theory; Distribution of primes.*

Received on the 11*th of September*, 1985.