A MOTIVATED ACCOUNT OF AN ELEMENTARY PROOF OF THE PRIME NUMBER THEOREM

NORMAN LEVINSON, Massachusetts Institute of Technology

1. Introduction. One of the most striking results of mathematics is the prime number theorem first conjectured, independently, by Gauss and Legendre prior to 1800 and proved, independently, by Hadamard and de la Vallée Poussin in 1896. Among the many great mathematicians of the 19th century who did not succeed in proving the theorem were Chebychef and Riemann, both of whom obtained important partial results. Riemann indicated that the prime number theorem was related to the behavior of the zeta function in the complex plane and found many properties of this function which has since borne his name. Riemann's ideas were exploited and augmented in the proofs of Hadamard and de la Vallée Poussin.

In 1949, P. Erdös and A. Selberg, using a formula previously proved by Selberg in an elementary way, jointly succeeded in giving several elementary proofs of the prime number theorem, [3]. While elementary, neither these proofs, nor another one of Selberg [6], are simple.

With the tremendous proliferation of mathematics, many mathematicians no longer study number theory. Therefore it seems worthwhile to give a self-contained and motivated account of an elementary proof of the prime number theorem.

The prime numbers (2, 3, 5, 7, 11, 13, $\cdots$) were known to ancient man and in Euclid there is a proof that they are infinite in number. The number of primes not exceeding x is called $\pi(x)$ and can be represented by

$$\pi(x) = \sum_{p \le x} 1 \tag{1.1}$$

where the symbol p runs over the sequence of primes in increasing order. The simplest form of Legendre's conjecture was

$$\lim_{x\to\infty} \frac{\pi(x)}{x/\log x} = 1. \tag{1.2}$$

Gauss' conjecture has turned out to be more profound and was that $\pi(x)$, for large x, is close to $\int_2^x dt/\log t$. He arrived at this by observing from a tabulation of prime numbers that the primes seemed to have an asymptotic density which at

Professor Levinson studied with Norbert Wiener and received the Sc.D. degree in 1935. He was a travelling Fellow at Cambridge University in 1934–35 and an NRC Fellow in 1935–37. He has been on the MIT staff since 1937, except for a year as Guggenheim Fellow at the Mathematics Institute, Copenhagen and a year at the University of Tel Aviv. His main research interests are transforms, entire functions, probability, and differential equations. He is the author of the AMS Colloquium volume, *Gap and Density Theorems*, and (with E. Coddington) *Ordinary Differential Equations*. Professor Levinson received the AMS Bôcher Prize and he is a member of the National Academy of Sciences. *Editor*

x was $1/\log x$. Because of this, for some purposes a better way to find the asymptotic behavior of the primes is to weight each p with $\log p$. This is done in the function

$$\theta(x) = \sum_{p \leq x} \log p. \tag{1.3}$$

Actually it turns out to be even more convenient to use not $\theta(x)$ but a closely related function $\psi(x)$ as will be seen.

The account that follows begins with the factorization of an integer into the product of powers of primes and proceeds with motivated proofs of the relevant discoveries of the 19th century in sections 2 and 3. This approach is continued in section 4 to prove Selberg's formula, and finally in section 5 where an exposition of proof of Selberg [6] is given as simplified by Wright [4], [9], and further simplified by the author [5].

The elementary proof of the prime number theorem has been extended to give elementary proofs of sharper forms of the theorem with a remainder by Breusch [2], Bombieri [1], Wirsing [8] and others. I am indebted to George B. Thomas for a critical reading of the manuscript.

2. The Chebychef identity and its inversion. Our starting point is that a positive integer can be factored into a product of powers of distinct primes. Thus a positive integer

$$n = p_1^{k_1} p_2^{k_2} \cdots p_m^{k_m}, \tag{2.1}$$

where the p_j, $1 \leqq j \leqq m$, are distinct primes and each k_j is a positive integer. Because addition is simpler than multiplication a more useful form of (2.1) is

$$\log n = k_1 \log p_1 + k_2 \log p_2 + \cdots + k_m \log p_m. \tag{2.2}$$

The utility of this formula is very much enhanced by the use of the von Mangoldt symbol $\Lambda(n)$, introduced in 1895, which is defined by

$$\Lambda(n) = \log p \text{ for } n = p^j, \tag{2.3}$$

where p is a prime number and j is a positive integer, and $\Lambda(n)=0$ otherwise. Thus $\Lambda(n)\neq 0$ only if n is a power of a prime.

The symbol $\sum_{j|n}$ will be used to denote a sum on j where j runs through all of the positive divisors of the positive integer n. With this notation it will be shown that (2.2) can be written as

$$\log n = \sum_{j|n} \Lambda(j). \tag{2.4}$$

To prove (2.4) note that because of (2.1) and the definition of $\Lambda(j)$, the only nonzero terms that can appear on the right side of (2.4) are $\log p_1$, $\log p_2$, $\cdots$ log

p_m. Moreover log p_1 appears for $j=p_1$, for $j=p_1^2, \cdots$, and for $j=p_1^{k_1}$. Thus log p_1 appears exactly k_1 times; similarly log p_2 appears k_2 times, etc., which shows that (2.4) is a consequence of (2.2). The formula (2.4) is an extremely powerful variant of (2.1) and incorporates the properties of prime numbers which are needed here. The transformation of (2.2) into the form (2.4) is not obvious and historically came relatively late.

The formula (2.4) can be written in the equivalent form

$$\log n = \sum_{ij=n} \Lambda(j), \tag{2.5}$$

where i and j are positive integers each of which takes on all possible values satisfying $ij=n$, so that indeed j runs through all positive divisors of n (as does i also).

The number of primes up to x, $\pi(x)$, is closely related to the sum

$$\psi(x) = \sum_{j\leq x} \Lambda(j). \tag{2.6}$$

From the definition of $\Lambda(j)$,

$$\psi(x) = \sum_{p\leq x} \log p + \sum_{p^2\leq x} \log p + \sum_{p^3\leq x} \log p + \cdots = \theta(x) + \theta(x^{1/2}) + \theta(x^{1/3}) + \cdots$$

The function $\psi(x)$, expressed in the latter form, was already known to Chebychef, who gave a simple proof that the prime number theorem (1.2) is equivalent to

$$\lim_{x\to\infty} \frac{\psi(x)}{x} = 1. \tag{2.7}$$

This proof will be given in Lemma 3.4 and (2.7) will be proved in Section 5. (Roughly speaking, $\psi(x)$ acts like $\pi(x) \log x$ for large x because $\psi(x)$ counts each $p\leqq x$ with weight $\log p$, (2.3), and because $\log p$ is close to $\log x$ for "most" of the $p\leqq x$. True, $\psi(x)$ also counts $\log p$ again for $p\leqq x^{1/2}$, for $p\leqq x^{1/3}$, etc., but these last are very sparse as will be seen later in the proof.)

To use (2.5) to get information about $\psi(x)$, (2.5) is summed on $n\leqq x$ to get $\sum_{n\leqq x} \log n = \sum_{n\leqq x}\sum_{ij=n} \Lambda(j)$, so that if one defines

$$T(x) = \sum_{n\leq x} \log n, \tag{2.8}$$

then

$$T(x) = \sum_{ij\leq x} \Lambda(j). \tag{2.9}$$

Because the logarithm is a smooth function, $T(x)$ can be readily appraised for large x, and this will be done in (3.4).

The double sum in (2.9) is taken over those lattice points in the positive quadrant of the (i, j) plane which lie on or below the hyperbola $ij=x$. If the double sum (2.9) is treated as a repeated sum, summing first on j

$$T(x) = \sum_{i\leq x} \sum_{j\leq x/i} \Lambda(j) = \sum_{i\leq x}\psi(x/i).$$

This identity was discovered by Chebychef (1850) and will be rewritten as

$$T(x) = \sum_{n\leq x} \psi(x/n). \tag{2.10}$$

The Chebychef identity (2.10) is really a transform relationship. It suggests that given a function $F(x)$, defined for $x>1$, one defines a related function $G(x)$ for $x>1$ by

$$G(x) = \sum_{n\leq x} F(x/n) = F(x) + F(x/2) + F(x/3) + \cdots F(x/[x]), \tag{2.11}$$

where as usual $[x]$ is the largest integer not exceeding x. $G(x)$ may be regarded as a transform of $F(x)$. G is seen to be a linear homogeneous function of F. Transform relationships are among the most powerful tools of the mathematician and this one is no exception.

Since $T(x)$ is a comparatively simple function, it is of interest to try to invert the relationship (2.10) to express $\psi(x)$ in terms of T, or in the more general notation, to try to invert (2.11) to find F in terms of G. To solve for F in terms of G, a first modest step would be to eliminate $F(x/2)$ from the right side of (2.11). This is easily done by writing (2.11) with x replaced by $x/2$ to get

$$G\left(\frac{x}{2}\right) = F\left(\frac{x}{2}\right) + F\left(\frac{x}{4}\right) + F\left(\frac{x}{6}\right) + \cdots$$

Subtracting the above from (2.11) would eliminate $F(x/2)$. This process can be extended at once by writing (2.11) with x replaced by $x/2$, then by $x/3$, etc., to get

$$\begin{aligned}
G(x) &= F(x)+F\left(\frac{x}{2}\right)+F\left(\frac{x}{3}\right)+F\left(\frac{x}{4}\right)+F\left(\frac{x}{5}\right)+F\left(\frac{x}{6}\right)+\cdots \\
G\left(\frac{x}{2}\right) &= \qquad F\left(\frac{x}{2}\right) \qquad\qquad +F\left(\frac{x}{4}\right) \qquad\qquad +F\left(\frac{x}{6}\right)+\cdots \\
G\left(\frac{x}{3}\right) &= \qquad\qquad\qquad F\left(\frac{x}{3}\right) \qquad\qquad\qquad\qquad +F\left(\frac{x}{6}\right)+\cdots \\
G\left(\frac{x}{4}\right) &= \qquad\qquad\qquad\qquad\qquad F\left(\frac{x}{4}\right) \qquad\qquad\qquad\qquad +\cdots
\end{aligned} \tag{2.12}$$

$$(2.12)\quad \begin{aligned} G\left(\frac{x}{5}\right) &= \qquad\qquad\qquad\qquad F\left(\frac{x}{5}\right) \qquad + \cdots \\ G\left(\frac{x}{6}\right) &= \qquad\qquad\qquad\qquad\qquad\qquad F\left(\frac{x}{6}\right) + \cdots \end{aligned}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

If one uses the equations in sequence one can first eliminate $F(x/2)$, then $F(x/3)$, then $F(x/4)$, etc., from the right. For example up to $G(x/6)$ one gets

$$F(x) = G(x) - G\left(\frac{x}{2}\right) - G\left(\frac{x}{3}\right) - G\left(\frac{x}{5}\right) + G\left(\frac{x}{6}\right) + \cdots$$

This suggests, and indeed, because of the diagonal form of the right side of (2.12), actually proves that (2.11) can be inverted by a formula of the type

$$(2.13)\qquad F(x) = \sum_{k\leqq x} \mu(k) G\left(\frac{x}{k}\right),$$

where the $\mu(k)$ remain to be specified, (Möbius, 1832). To determine the $\mu(k)$ note that by (2.11) $G(x/k) = \sum_{j\leqq x/k} F(x/jk)$ which in (2.13) gives

$$F(x) = \sum_{k\leqq x} \mu(k) \sum_{j\leqq x/k} F(x/jk) = \sum_{jk\leqq x} \mu(k) F(x/jk).$$

If this double sum is summed first on the lattice points on the hyperbolas $jk = n$ and then for n, $1\leqq n\leqq x$,

$$(2.14)\qquad F(x) = \sum_{n\leqq x} F(x/n) \sum_{jk=n} \mu(k).$$

The equation (2.14) becomes an identity if $\mu(1) = 1$ and, replacing $jk = n$ by $k\mid n$, if

$$(2.15)\qquad \sum_{k\mid n} \mu(k) = 0, \qquad n \geqq 2.$$

Setting $n = 2, 3, 4$, etc. successively determines the $\mu(k)$ uniquely. To find the $\mu(k)$ explicitly try the case $n = p$ to get $k = 1$ and $k = p$ which gives $\mu(1) + \mu(p) = 0$ and hence $\mu(p) = -1$. The case $n = p_1p_2$ gives

$$\mu(1) + \mu(p_1) + \mu(p_2) + \mu(p_1p_2) = 0$$

and hence $\mu(p_1p_2) = 1$. Similarly it is easily found that

$$\mu(p_1p_2p_3) = -1,\ \mu(p^2) = 0,\ \mu(p^3) = 0, \cdots, \mu(p_1^2p_2) = 0.$$

This suggests that

(2.16) $$\mu(n) = (-1)^m, \qquad n = p_1 p_2 \cdots p_m,$$

where $p_1, p_2, \cdots, p_m$ are all distinct primes and

(2.17) $$\mu(n) = 0 \quad \text{if } p^2 \mid n,$$

where as usual p is a prime. The function $\mu(n)$ is known as the Möbius function.

It will now be proved that if $\mu(n)$ is defined as in (2.16) and (2.17) above, then (2.15) is indeed valid. We recall that the solution of (2.15) was unique. Because of (2.17) for $n = p_1^{k_1} p_2^{k_2} \cdots p_m^{k_m}$,

$$\sum_{j \mid n} \mu(j) = \sum_{j \mid p_1 p_2 \cdots p_m} \mu(j),$$

so only the right side need be treated to prove (2.15). If $m = 1$, (2.15) is true since $\mu(1) = 1$ and $\mu(p_1) = -1$. If $m \geqq 2$

(2.18) $$\sum_{j \mid p_1 \cdots p_m} \mu(j) = \sum_{k \mid p_1 \cdots p_{m-1}} (\mu(k) + \mu(kp_m)).$$

But from (2.16) if k in (2.18) is the product of r primes

$$\mu(kp_m) = (-1)^{r+1} = -\mu(k).$$

Hence each term on the right of (2.18) is zero and so (2.15) is proved. Moreover by (2.16) and (2.17)

(2.19) $$|\mu(n)| \leqq 1$$

(which is the only use we shall make of the material which begins after (2.15) and ends with (2.19)).

Applying the Möbius inversion formula (2.13) to Chebychef's identity (2.10) gives the *inversion formula*

(2.20) $$\psi(x) = \sum_{k \leqq x} \mu(k) T\left(\frac{x}{k}\right).$$

Using the definitions of ψ and T this can be written as

$$\begin{aligned}\sum_{n \leqq x} \Lambda(n) &= \sum_{k \leqq x} \mu(k) \sum_{j \leqq x/k} \log j = \sum_{jk \leqq x} \mu(k) \log j = \sum_{n \leqq x} \sum_{jk = n} \mu(k) \log j \\ &= \sum_{n \leqq x} \sum_{k \mid n} \mu(k) \log n/k.\end{aligned}$$

Used for $x = 1, 2, 3, \cdots$ the above proves that

(2.21) $$\Lambda(n) = \sum_{k \mid n} \mu(k) \log n/k, \qquad n \geqq 1,$$

which is the inversion formula for (2.4). The referee observes that one could also show (2.21) directly using (2.4) and (2.15).

3. Some elementary results. Although those results concerning prime numbers that follow were all discovered in the 19th century, some were not found until as much as 70 years after the prime number theorem was first conjectured by Legendre and Gauss.

It will be convenient to use the following well-known lemma in which, as usual, $[x]$ is the largest integer not exceeding x.

LEMMA 3.1. *Let $f(t)$ have a continuous derivative, $f'(t)$, for $t \geqq 1$. Let c_n, $n \geqq 1$, be constants and let $C(u) = \sum_{n \leqq u} c_n$. Then*

$$\sum_{n \leqq x} c_n f(n) = f(x)C(x) - \int_1^x f'(t)C(t)dt \tag{3.1}$$

and

$$\sum_{n \leqq x} f(n) = \int_1^x f(t)dt + \int_1^x (t - [t])f'(t)dt + f(1) - (x - [x])f(x). \tag{3.2}$$

Proof. $C(n) - C(n-1) = c_n$ and $C(u) = C([u])$ since $C(u)$ is a step function. Thus if $[x] = N$,

$$\begin{aligned} \sum_{n \leqq x} c_n f(n) &= \sum_{n \leqq x} (C(n) - C(n-1))f(n) \\ &= \sum_{n \leqq x-1} C(n)(f(n) - f(n+1)) + C(x)f(N) \\ &= - \sum_{n \leqq x-1} C(n) \int_n^{n+1} f'(t)dt + C(x)f(N) \\ &= - \int_1^N C(t)f'(t)dt + C(x)f(N). \end{aligned}$$

Since $C(t)$ is constant on $N \leqq t < x$,

$$\int_N^x C(t)f'(t)dt = C(x)(f(x) - f(N)).$$

Adding this to the previous equation and transposing the integral on the left side to the right proves (3.1).

In case $c_n = 1$, (3.1) becomes

$$\begin{aligned} \sum_{n \leqq x} f(n) &= [x]f(x) - \int_1^x f'(t)[t]dt \\ &= [x]f(x) - \int_1^x f'(t)tdt + \int_1^x f'(t)(t - [t])dt. \end{aligned}$$

Integrating the first integral on the right by parts proves (3.2).

It will be convenient to use the following notation. Suppose $f(x)$ is bounded for finite x and that there is a constant K and a $g(x)$ such that for large x

$$|f(x)| \leqq Kg(x);$$

then this will be denoted by

$$f(x) = O(g(x)) \tag{3.3}$$

and, where convenient, $f(x)$ will be replaced by the right side above.

Applying (3.2) to $f(t) = \log t$ and using $0 \leqq t - [t] < 1$ gives

$$T(x) = x \log x - x + O(\log x), \tag{3.4}$$

which is a weak form of Stirling's formula.

LEMMA 3.2. (Chebychef 1850). *For large* x

$$\psi(x) < \tfrac{3}{2}x. \tag{3.5}$$

Proof. Using Chebychef's identity (2.10)

$$T(x) - 2T(x/2) = \psi(x) - \psi(x/2) + \psi(x/3) - \psi(x/4) + \cdots \geqq \psi(x) - \psi(x/2)$$

because $\psi(x/(2n-1)) - \psi(x/2n) \geqq 0$ since ψ is monotone nondecreasing. Using (3.4), $\psi(x) - \psi(x/2) \leqq x \log 2 + K \log x$, $x \geqq 2$, for some constant K. Applying the above with x replaced by $x/2^j$,

$$\psi\left(\frac{x}{2^j}\right) - \psi\left(\frac{x}{2^{j+1}}\right) \leqq \frac{x}{2^j} \log 2 + K \log x \tag{3.6}$$

so long as $x/2^j \geqq 2$ which implies $j < \log x/\log 2$. Recalling that $\psi(t) = 0$, $t < 2$, and adding (3.6) for $0 \leqq j < \log x/\log 2$,

$$\begin{aligned} \psi(x) &\leqq x \log 2\left(1 + \frac{1}{2} + \frac{1}{2^2} + \cdots\right) + \frac{\log x}{\log 2} K \log x \\ &= 2x \log 2 + K \log^2 x/\log 2. \end{aligned}$$

Since $\log 2 < .7$ this proves (3.5).

LEMMA 3.3. (Proved 1874 by Mertens in a slightly different form.)

$$\sum_{n \leqq x} \frac{\Lambda(n)}{n} = \log x + O(1). \tag{3.7}$$

Proof. In the double sum (2.9), sum first on i and then on j, (the opposite of what was done in the derivation of (2.10)), to get

$$\begin{aligned} T(x) &= \sum_{j \leqq x} \Lambda(j) \sum_{i \leqq x/j} 1 = \sum_{j \leqq x} \Lambda(j) \left[\frac{x}{j}\right] \\ &= x \sum_{j \leqq x} \frac{\Lambda(j)}{j} - \sum_{j \leqq x} \Lambda(j) \left(\frac{x}{j} - \left[\frac{x}{j}\right]\right). \end{aligned} \tag{3.8}$$

Moreover

$$0 \leqq \sum_{j \leqq x} \Lambda(j)\left(\frac{x}{j} - \left[\frac{x}{j}\right]\right) \leqq \sum_{j \leqq x} \Lambda(j) = \psi(x) = O(x) \tag{3.9}$$

by (3.5). Using this and (3.4) in (3.8) proves (3.7).

LEMMA 3.4.

$$\psi(x) = \pi(x) \log x + O\left(\frac{x \log \log x}{\log x}\right), \tag{3.10}$$

so that (2.7) *is equivalent to the prime number theorem.*

Proof. From its definition (2.6) and from (2.3),

$$\psi(x) = \sum_{p \leqq x} \log p + \sum_{p \leqq x^{1/2}} \log p + \sum_{p \leqq x^{1/3}} \log p + \cdots, \tag{3.11}$$

where the sums $p \leqq x^{1/j}$ above are not zero only if $x^{1/j} \geqq 2$ or if $j \leqq \log x/\log 2$. Hence

$$\psi(x) \leqq \sum_{p \leqq x} \log p + \frac{\log x}{\log 2} \sum_{p \leqq x^{1/2}} \log p.$$

From the definition (1.1) of $\pi(x)$ this gives

$$\psi(x) \leqq \log x\, \pi(x) + \frac{\log x}{\log 2} \pi(x^{1/2}) \log x^{1/2}.$$

Since $\pi(y) \leqq y$, the above gives

$$\psi(x) \leqq \log x\, \pi(x) + \frac{x^{1/2} \log^2 x}{2 \log 2} \cdot \tag{3.12}$$

By (3.11)

$$\begin{aligned}\psi(x) &\geqq \sum_{x/\log^2 x < p \leqq x} \log p \geqq \log\left(\frac{x}{\log^2 x}\right) \sum_{x/\log^2 x < p \leqq x} 1 \\ &= \log\left(\frac{x}{\log^2 x}\right)\left(\pi(x) - \pi\left(\frac{x}{\log^2 x}\right)\right).\end{aligned}$$

Since $\pi(y) \leqq y$, this gives

$$\frac{\psi(x)}{\log x - 2 \log \log x} \geqq \pi(x) - \frac{x}{\log^2 x},$$

or

$$\pi(x) \log x \leqq \psi(x) \frac{\log x}{\log x - 2 \log \log x} + \frac{x}{\log x}$$

$$= \psi(x) + \psi(x)\frac{2\log\log x}{\log x - 2\log\log x} + \frac{x}{\log x}.$$

Using (3.5) and $2\log\log x < (\log x)/4$ for large x,

$$\pi(x)\log x \leqq \psi(x) + \frac{4x\log\log x}{\log x} + \frac{x}{\log x}, \tag{3.13}$$

which with (3.12) proves the lemma.

It will be useful later to apply (3.2) to $f(t) = 1/t$.

LEMMA 3.5.

$$\sum_{n\leq x} 1/n = \log x + \gamma + O(1/x), \tag{3.14}$$

where γ *is a constant* (Euler's constant).

Proof. Applying (3.2) to $f(t) = 1/t$

$$\sum_{n\leq x}\frac{1}{n} = \log x - \frac{x - [x]}{x} + 1 - \int_1^x \frac{t - [t]}{t^2}\,dt.$$

If

$$\gamma = 1 - \int_1^\infty \frac{t - [t]}{t^2}\,dt, \tag{3.15}$$

then $\sum_{n\leq x} 1/n = \log x + \gamma + H$, where

$$H = \int_x^\infty \frac{t - [t]}{t^2}\,dt - \frac{x - [x]}{x} = O\left(\frac{1}{x}\right)$$

since $0 \leqq t - [t] < 1$, which proves (3.14).

REMARK. From (3.15), $0 < \gamma < 1$.

4. Selberg's elementary inequality. The Möbius inversion formula (2.20) which expresses ψ in terms of T will now be used in an attempt to find how $\psi(x)$ behaves for large x. The computation will be simplified if it is possible to find a simple $F(x)$, say $\tilde{F}(x)$, with a transform $\tilde{G}(x)$ which is close to $T(x)$. In that case subtract the Möbius inversion formula for $\tilde{F}$ (2.13), from that for ψ:

$$\psi(x) - \tilde{F}(x) = \sum_{k\leq x}\mu(k)\left(T\left(\frac{x}{k}\right) - \tilde{G}\left(\frac{x}{k}\right)\right); \tag{4.1}$$

if the right side could be shown to be small, then $\psi(x)$ would be close to $\tilde{F}(x)$.

If it were proved that $\psi(x)/x \to 1$, then $\psi(x)$ would be close to x for large x. This suggests one try $\tilde{F}(x) = F_0(x) = x$. Hence $G_0(x) = \sum_{n\leq x} F_0(x/n) = x\sum_{n\leq x} n^{-1}$

which by (3.14) becomes $G_0(x)=x\log x+\gamma x+O(1)$. This is not close enough to $T(x)$ as given by (3.4). As a refinement let $\tilde{F}(x)=F_1(x)=x-C$ where C is a constant. (There are many choices other than C that would work here.) Then

$$\begin{aligned} G_1(x) &= x\sum_{n\leq x}\frac{1}{n} - C\sum_{n\leq x}1 = x\log x+\gamma x+O(1)-C[x] \\ &= x\log x-(C-\gamma)x+O(1). \end{aligned}$$

Hence if $C=1+\gamma$ then by (3.4)

$$T(x) - G_1(x) = O(\log x) \tag{4.2}$$

which is comparatively small. Using (4.1) with $\tilde{F}=x-C$

$$\psi(x) - x + C = \sum_{k\leq x}\mu(k)\left(T\left(\frac{x}{k}\right) - G_1\left(\frac{x}{k}\right)\right). \tag{4.3}$$

Even if the right side of (4.2) were in the stronger form $O(1)$ (which is false), the fact that (by (2.19)) $|\mu(k)|\leqq 1$ would imply only that the right side of (4.3) is $O(x)$. Thus the inversion formula (4.3) gives, *not* the prime number theorem, but at most the much weaker result

$$\psi(x) = O(x), \tag{4.4}$$

(already proved more simply in Lemma 3.2). Actually (4.3) does give (4.4) as the following crude argument shows.

Since the logarithm grows more slowly than any positive algebraic power, $\log x=O(x^{1/2})$. Thus, for example, (4.2) implies the much weaker result

$$T(x) - G_1(x) = O(x^{1/2}). \tag{4.5}$$

Using this and $|\mu(k)|\leqq 1$, there is a constant K such that the right side of (4.3) is dominated by

$$\begin{aligned} Kx^{1/2}\sum_{k\leq x}k^{-1/2} &< Kx^{1/2}\left(1+\sum_{2\leq k\leq x}\int_{k-1}^{k}u^{-1/2}du\right) \\ &\leqq Kx^{1/2}\left(1+\int_1^x u^{-1/2}du\right) = O(x) \end{aligned} \tag{4.6}$$

which does in fact prove (4.4).

Thus the Möbius inversion of Chebychef's formula yields only the crude result (4.4), and herein lies the reason for the long delay in the discovery of an elementary proof of the prime number theorem.

Note that *the crude result* (4.5) *serves just as well as the much more refined* (4.2) in appraising the right side of (4.3). This suggests the following idea.

In the Möbius inversion formula

$$F(x) = \sum_{k\leq x}\mu(k)G\left(\frac{x}{k}\right) \tag{4.7}$$

(where for us $F=\psi-x+C$ and $G=T-G_1$), we can increase the terms in the sum on the right side somewhat since doing so will not change the crude appraisal $O(x)$, (4.6), for this side. On the other hand, a judicious increase of the terms on the right side might possibly replace $F(x)$ (and hence $\psi(x)$) on the left side by some growing function multiplied by $F(x)$, which would then make the appraisal $O(x)$ for the right side useful.

A little experimentation shows that the simplest case to compute explicitly is where the right side of (4.7) is replaced by

$$J(x) = \sum_{k\leq x} \mu(k) \log \frac{x}{k} G\left(\frac{x}{k}\right). \tag{4.8}$$

This must now be computed in terms of F. From the definition of G,

$$\begin{aligned} J(x) &= \sum_{k\leq x} \mu(k) \log \frac{x}{k} \sum_{j\leq x/k} F\left(\frac{x}{jk}\right) \\ &= \sum_{jk\leq x} \mu(k) \log \frac{x}{k} F\left(\frac{x}{jk}\right) = \sum_{n\leq x} F\left(\frac{x}{n}\right) \sum_{jk=n} \mu(k) \log \frac{x}{k} \\ &= \sum_{n\leq x} F\left(\frac{x}{n}\right) \sum_{k|n} \mu(k) \log \frac{x}{k}\cdot \end{aligned}$$

Using $\log x/k = \log x/n + \log n/k$

$$J(x) = \sum_{n\leq x} F\left(\frac{x}{n}\right) \log \frac{x}{n} \sum_{k|n} \mu(k) + \sum_{n\leq x} F\left(\frac{x}{n}\right) \sum_{k|n} \mu(k) \log \frac{n}{k}\cdot$$

By (2.15) and (2.21) this becomes $J(x)=F(x)\log x+\sum_{n\leq x} F(x/n)\Lambda(n)$. With (4.8) this gives

$$F(x) \log x + \sum_{n\leq x} F\left(\frac{x}{n}\right) \Lambda(n) = \sum_{k\leq x} \mu(k) \log \frac{x}{k} G\left(\frac{x}{k}\right), \tag{4.9}$$

and this is the Tatuzawa-Iseki identity [7] which leads easily to the inequality of Atle Selberg. Indeed by (4.2)

$$\log x(T(x) - G_1(x)) = O(\log^2 x) = O(x^{1/2})$$

and hence as already shown in (4.6)

$$\sum_{k\leq x} \mu(k) \log \frac{x}{k}\left(T\left(\frac{x}{k}\right) - G_1\left(\frac{x}{k}\right)\right) = O(x).$$

Thus (4.9) with $F(x)=\psi(x)-x+C$ becomes

$$(\psi(x) - x) \log x + \sum_{n\leq x}\left(\psi\left(\frac{x}{n}\right) - \frac{x}{n}\right)\Lambda(n) = O(x), \tag{4.10}$$

where use is made of (4.4) to incorporate $C\psi(x)$ together with $C \log x$ in $O(x)$. (4.10) is a form of the famous inequality of Atle Selberg [6].

Because of Lemma 3.3, (4.10) can be written as

$$\psi(x) \log x + \sum_{n \leq x} \Lambda(n)\psi(x/n) = 2x \log x + O(x). \tag{4.11}$$

With $c_n = \Lambda(n)$, (3.1) and (3.5) yield

$$\sum_{n \leq x} \Lambda(n) \log n = \psi(x) \log x - \int_1^x \frac{\psi(t)}{t}\, dt = \psi(x) \log x + O(x). \tag{4.12}$$

Also

$$\sum_{j \leq x} \Lambda(j)\psi\left(\frac{x}{j}\right) = \sum_{j \leq x} \Lambda(j) \sum_{k \leq x/j} \Lambda(k) = \sum_{jk \leq x} \Lambda(j)\Lambda(k). \tag{4.13}$$

Thus if

$$\Lambda_2(n) = \Lambda(n) \log n + \sum_{jk=n} \Lambda(j)\Lambda(k), \tag{4.14}$$

then (4.12) and (4.13) in (4.11) yield $\sum_{n \leq x} \Lambda_2(n) = 2x \log x + O(x)$ as an equivalent to (4.11). By (3.4) $\sum_{n \leq x} \log n = x \log x + O(x)$. Combining the above two inequalities,

$$Q(n) = \sum_{k \leq n} (\Lambda_2(k) - 2 \log k) = O(n), \qquad n \geqq 2, \text{ and } Q(1) = 0. \tag{4.15}$$

5. Proof of the prime number theorem. If $R(x) = \psi(x) - x$, $x \geqq 2$, and $R(x) = 0$, $x < 2$, then (4.10) becomes

$$R(x) \log x + \sum \Lambda(n)R(x/n) = O(x), \tag{5.1}$$

where the summation is self terminating since $R(x/n) = 0$ for $n > x/2$. The goal (2.7) takes the form

$$\lim_{x \to \infty} \frac{R(x)}{x} = 0. \tag{5.2}$$

The derivation of (5.2) from (5.1) is complicated because the weights $\Lambda(n)$ in the weighted sum in (5.1) depend on the location of the prime numbers which is just what we are trying to find. Because of this complication no easy derivation of (5.2) from (5.1) has been found.

The proof that follows uses several smoothing operations on (5.1) to get a more tractable inequality. Most of these smoothings involve a loss of information, and the objective is to smooth for tractability but not to degrade (5.1) completely.

First $R(x)$ will be replaced by the smoother

(5.3) $$S(y) = \int_2^y \frac{R(x)}{x}\, dx, \qquad y \geqq ;2$$

$S(y)=0$, $y<2$. Fortunately it is easy to show, as will be done later, that (5.2) is implied if we can prove

(5.4) $$\lim_{y\to\infty} \frac{S(y)}{y} = 0.$$

LEMMA 5.1. *There exists a constant c such that*

(5.5) $$|S(y)| \leqq cy \qquad y \geqq 2$$

and

(5.6) $$|S(y_2) - S(y_1)| \leqq c|y_1 - y_2|.$$

Moreover a consequence of (5.1) *is*

(5.7) $$S(y)\log y + \sum \Lambda(j)S\left(\frac{y}{j}\right) = O(y).$$

Proof. From (3.5), $-x \leqq \psi(x) - x \leqq \frac{1}{2}x$ for large x. Hence

(5.8) $$\limsup_{x\to\infty} \frac{|R(x)|}{x} \leqq 1$$

and, since $|R(x)|$ is bounded for finite x, there must exist a constant c such that

(5.9) $$|R(x)| \leqq cx, \qquad x \geqq 2.$$

By (5.3) $S'(y) = R(y)/y$ except at $y = p^i$ where $R(y)$ is discontinuous. By (5.9) then

(5.10) $$|S'(y)| \leqq c, \qquad y \neq p^i.$$

Hence, first for the case where the interval $y_1 < y < y_2$ contains no p^i, (5.6) is true. However since $S(y)$ is continuous, the fact that the magnitude of a sum is less than or equal to the sum of the magnitudes, allows (5.6) to be extended for all y_1 and y_2. The condition (5.6) is known as a Lipschitz condition. The result (5.5) follows from (5.6) with $y_1 = 2$.

Since $||a| - |b|| \leqq |a - b|$, (5.6) yields

(5.11) $$||S(y_2)| - |S(y_1)|| \leqq c|y_2 - y_1|.$$

To prove (5.7), divide (5.1) by x and integrate to get

(5.12) $$\int_2^y \frac{R(x)}{x}\log x\, dx + \sum \Lambda(n)\int_2^y R\left(\frac{x}{n}\right)\frac{dx}{x} = O(y).$$

Integrating the first term by parts

$$\int_2^y \frac{R(x)}{x}\log x\,dx = \log y\,S(y) - \int_2^y \frac{S(x)}{x}\,dx = \log y\,S(y) + O(y)$$

by (5.5). Also if $\xi = x/n$

$$\int_2^y R\left(\frac{x}{n}\right)\frac{dx}{x} = \int_2^{y/n} \frac{R(\xi)}{\xi}\,d\xi = S\left(\frac{y}{n}\right).$$

These in (5.12) prove (5.7).

To make the weighted sum in (5.7) more tractable, the density of the set of points where $S(y/j)$ actually appears in the sum will be increased by iterating (5.7).

LEMMA 5.2. *With* $\Lambda_2(n) = \Lambda(n) + \sum_{ij=n}\Lambda(i)\Lambda(j)$ *as in* (4.14) *and* K_1 *a constant*

$$\log^2 y\,|S(y)| \leqq \sum \Lambda_2(m)\,|S(y/m)| + K_1 y\log y. \tag{5.13}$$

Proof. Replace y in (5.7) by y/k, multiply by $\Lambda(k)$, and sum for $k \leqq y$ to get

$$\sum \Lambda(k)S\left(\frac{y}{k}\right)\log\frac{y}{k} + \sum\sum \Lambda(k)\Lambda(j)S\left(\frac{y}{jk}\right) = O(y)\sum_{k\leqq y}\frac{\Lambda(k)}{k}\,.$$

Setting $jk = m$ in the second sum and summing on m, and setting $\log y/k = \log y - \log k$ in the first sum and replacing this latter k by m,

$$\log y\sum \Lambda(k)S\left(\frac{y}{k}\right) - \sum_{m\leqq y} S\left(\frac{y}{m}\right)\left\{\Lambda(m)\log m - \sum_{jk=m}\Lambda(j)\Lambda(k)\right\} = O(y\log y)$$

where (3.7) is used to get the right side. The first sum above is now replaced by use of (5.7) to give

$$S(y)\log^2 y = -\sum S\left(\frac{y}{m}\right)\left\{\Lambda(m)\log m - \sum_{jk=m}\Lambda(j)\Lambda(k)\right\} + O(y\log y).$$

Replacing all terms in the sum on the right by their magnitude gives (5.13).

The inequality (4.15) suggests that on the average $\Lambda_2(m)$ acts like 2 log m. A weighted sum with weights 2 log m is quite tractable and this suggests modifying (5.13) by replacing $\Lambda_2(m)$ by 2 log m.

LEMMA 5.3. *There is a constant* K_2 *such that*

$$\log^2 y\,|S(y)| \leqq 2\sum |S(y/m)|\log m + K_2 y\log y. \tag{5.14}$$

Proof.

$$\sum |S(y/m)|\,\Lambda_2(m) = 2\sum_{m\leq y}|S(y/m)|\log m + J(y) \tag{5.15}$$

where, since by (4.15), $\Lambda_2(m)-2\log m=Q(m)-Q(m-1)$,

$$\begin{aligned}J(y) &= \sum_{m\leqq y}(Q(m)-Q(m-1))\left|S(y/m)\right| \\ &= \sum_{m\leqq y}Q(m)\left|S(y/m)\right| - \sum_{m\leqq y}Q(m)\left|S(y/(m+1))\right| \\ &= \sum_{2\leqq m\leqq y}Q(m)(\left|S(y/m)\right| - \left|S(y/(m+1))\right|)\end{aligned}$$

since $S(y)=0$, $y<2$. Using (4.15) and (5.11) there is a constant K_3 such that

$$\begin{aligned}J(y) &\leqq K_3\sum_{2\leqq m\leqq y} m\left(\frac{y}{m}-\frac{y}{m+1}\right) \\ &= K_3 y\sum_{2\leqq m\leqq y}\frac{1}{m+1} < K_3 y\int_1^y\frac{dv}{v} = K_3 y\log y.\end{aligned}$$

This and (5.15) now prove that (5.14) is a consequence of (5.13).

There is a further simplification in replacing the sum in (5.14) by an integral.

LEMMA 5.4. *There is a constant K_4 such that*

(5.16) $$\log^2 y\left|S(y)\right| \leqq 2\int_2^y\left|S(y/u)\right|\log u\,du + K_4 y\log y.$$

Proof. Since $\log u$ is increasing

$$\log m\left|S(y/m)\right| \leqq \int_m^{m+1}\log u\left|S(y/m)\right|du.$$

On the right use $\left|S(y/m)\right|\leqq\left|S(y/u)\right|+\left|S(y/m)-S(y/u)\right|$ to get

(5.17) $$\begin{aligned}\log m\left|S(y/m)\right| &\leqq \int_m^{m+1}\log u\left|S(y/u)\right|du + J_m \\ J_m &= \int_m^{m+1}\log u\left|S(y/m)-S(y/u)\right|du.\end{aligned}$$

Using (5.6)

$$J_m \leqq c\left(\frac{y}{m}-\frac{y}{m+1}\right)\int_m^{m+1}\log u\,du \leqq \frac{cy\log(m+1)}{m(m+1)}.$$

Since $\log(m+1)\leqq m$, the above in (5.17) gives

$$\log m\left|S\left(\frac{y}{m}\right)\right| \leqq \int_m^{m+1}\log u\left|S\left(\frac{y}{u}\right)\right|dy + \frac{cy}{m+1}.$$

Using this in (5.14) now gives (5.16) with $K_4 = K_2 + c$.

The inequality (5.16) assumes a simpler form with an exponential change of variable. Replace u by $v = \log y/u$. Also let $x = \log y$. Then (5.16) becomes

$$x^2 \left| S(e^x) \right| \leqq 2 \int_0^{x-\log 2} \left| S(e^v) \right| (x - v) e^{(x-v)} dv + K_4 x e^x. \tag{5.18}$$

If

$$W(x) = e^{-x} S(e^x) \tag{5.19}$$

then (5.18) becomes

$$\left| W(x) \right| \leqq \frac{2}{x^2} \int_0^x (x - v) \left| W(v) \right| dv + \frac{K_4}{x} \cdot \tag{5 20}$$

This inequality contains valuable information since it says in effect that $|W(x)|$ is dominated by a weighted average of $|W|$. This has as a consequence the following lemma. (Note that γ below is not Euler's constant.)

LEMMA 5.5. *Let*

$$\alpha = \limsup_{x \to \infty} \left| W(x) \right|, \qquad \gamma = \limsup_{x \to \infty} \frac{1}{x} \int_0^x \left| W(\xi) \right| d\xi; \tag{5.21}$$

then $\alpha \leqq 1$ *and*

$$\alpha \leqq \gamma. \tag{5.22}$$

REMARK. Recalling (5.4) and (5.19), our goal now is $\alpha = 0$.

Proof. That $\alpha \leqq 1$ follows from (5.19) and the fact that (5.8) and (5.3) imply that

$$\limsup_{y \to \infty} \frac{\left| S(y) \right|}{y} \leqq 1. \tag{5.23}$$

The key result $\gamma \geqq \alpha$ will be proved by use of (5.20) and *this is the only use that is made of* Lemmas 5.2, 5.3 and 5.4. Note that (5.20) can be written as

$$\left| W(x) \right| \leqq \frac{2}{x^2} \int_0^x u \, du \left(\frac{1}{u} \int_0^u \left| W(v) \right| dv \right) + \frac{K_4}{x} \tag{5.24}$$

as can be verified by inverting the order of integration. But

$$\frac{2}{x^2} \int_0^x u \, du = 1$$

and hence the integral on the right of (5.24) is simply a weighted average of $(1/u) \int_0^u \left| W(v) \right| dv = (1/u) \int_0^u e^{-v} \left| S(e^v) \right| dv \leqq c$ by (5.5). Hence for any fixed x_1 and $x > x_1$,

$$I(x) = \frac{2}{x^2}\int_0^x u\,du\left(\frac{1}{u}\int_0^u |W(v)|\,dv\right) \tag{5.25}$$
$$\leqq \frac{2c}{x^2}\int_0^{x_1} u\,du + \frac{2}{x^2}\int_{x_1}^x u\,du\left(\frac{1}{u}\int_0^u |W(v)|\,dv\right).$$

Given $\epsilon > 0$, for sufficiently large x_1,

$$\frac{1}{u}\int_0^u |W(v)|\,dv < \gamma + \epsilon \qquad u \geqq x_1$$

from the definition of γ. Hence (5.25) gives

$$I(x) \leqq \frac{cx_1^2}{x^2} + (\gamma + \epsilon)\left(1 - \frac{x_1^2}{x^2}\right).$$

Thus for large x, (5.24) yields

$$|W(x)| \leqq \gamma + \epsilon + \frac{cx_1^2}{x^2} + \frac{K_4}{x}\cdot$$

Letting $x \to \infty$, $\alpha \leqq \gamma + \epsilon$, and since this is true for all $\epsilon > 0$ it implies (5.22).

Two more facts are required about W to prove that $\alpha = 0$.

LEMMA 5.5. *If $k = 2c$ then*

$$|W(x_2) - W(x_1)| \leqq k|x_2 - x_1|, \tag{5.26}$$

and hence

$$||W(x_2)| - |W(x_1)|| \leqq k|x_2 - x_1|. \tag{5.27}$$

Proof. Since $W(x) = e^{-x}S(e^x)$,

$$|W'(x)| \leqq e^{-x}|S(e^x)| + |S'(e^x)| \qquad x \neq j \log p.$$

Hence by (5.5) and (5.10), $|W'(x)| \leqq 2c = k$ for $x \neq j \log p$. This leads to (5.26) just as (5.10) led to (5.6).

LEMMA 5.7. *If $W(v) \neq 0$ for $v_1 < v < v_2$, then there exists a number M such that*

$$\int_{v_1}^{v_2} |W(v)|\,dv \leqq M, \qquad W(v) \neq 0, \qquad v_1 < v < v_2. \tag{5.28}$$

Proof. From (3.1) letting $c_n = \Lambda(n)$ and $f(n) = 1/n$, (3.7) implies

$$\int_2^x \frac{\psi(t)}{t^2}\,dt = \log x + O(1),$$

or since $R(t) = \psi(t) - t$,

$$\int_2^x \frac{R(t)}{t^2}\,dt = O(1). \tag{5.29}$$

But

$$\int_2^x \frac{S(y)}{y^2}\,dy = \int_2^x \frac{dy}{y^2}\int_2^y \frac{R(t)}{t}\,dt = \int_2^x \frac{R(t)}{t}\left(\int_t^x \frac{dy}{y^2}\right)dt$$
$$= \int_2^x \frac{R(t)}{t^2}\,dt - \frac{1}{x}\int_2^x \frac{R(t)}{t}\,dt.$$

Using (5.29) and (5.9), $\int_2^x (S(y)/y^2)dy = O(1)$, or letting $y = e^u$, $x = e^v$,

$$\int_{\log 2}^v W(u)du = O(1).$$

Writing this for $v = v_1$ and $v = v_2$ and subtracting, the resulting integral is bounded and hence there is a constant M such that

$$\left| \int_{v_1}^{v_2} W(u)du \right| \leqq M.$$

But if $W(u) \neq 0$, $v_1 < u < v_2$, this can be written as (5.28). Since M can be increased if convenient it can be assumed $Mk > 1$.

LEMMA 5.8. *A function $W(x)$ subject to the three conditions* (5.22), (5.27) *and* (5.28) *must in fact have* $\alpha = 0$.

Proof. Choose $\beta > \alpha$. Then from the definition of α there exists an x_β such that

$$| W(x) | \leqq \beta \qquad x \geqq x_\beta. \tag{5.30}$$

If $W(x) \neq 0$ for all large x it follows from (5.28) that $\gamma = 0$ and hence that $\alpha = 0$. Suppose then that $W(x)$ has arbitrarily large zeros. Let a and b be successive zeros of $W(x)$ for $x > x_\beta$.

CASE 1. $b - a \geqq 2M/\beta$. By (5.28), since $W(x) \neq 0$, $a < x < b$,

$$\int_a^b | W(x) | \leqq M \leqq \tfrac{1}{2}(b-a)\beta.$$

(Hence the average of $|W|$ on (a, b) is less than $\frac{1}{2}\beta$.)

CASE 2. $b - a \leqq 2\beta/k$. In this case it follows from (5.27) that if the graph of $|W(x)|$ rises as rapidly as possible going right from $x = a$ and left from $x = b$, it cannot lie above a triangle with altitude $k(b-a)/2 \leqq \beta$ and hence

$$\int_a^b | W(x) | \, dx \leqq \tfrac{1}{2}(b-a)\beta.$$

CASE 3. $2\beta/k < b - a < 2M/\beta$. Reasoning as in Case 2 for a distance β/k from each endpoint and using (5.30) otherwise,

$$\int_a^b |W(x)| \leqq \frac{\beta^2}{k} + \left(b - a - \frac{2\beta}{k}\right)\beta$$

$$(5.31) \qquad = (b-a)\beta\left(1 - \frac{\beta}{k(b-a)}\right) \leqq (b-a)\beta\left(1 - \frac{\beta^2}{2Mk}\right)$$

$$< (b-a)\beta\left(1 - \frac{\alpha^2}{2Mk}\right).$$

Since $Mk > 1$ and since $\alpha \leqq 1$, (5.31) is valid in Cases 1 and 2 also. If x_1 is the first zero of $W(x)$ to the right of x_β and $\bar{x}$ the largest zero to the left of y, then (5.31) and (5.28) imply that

$$\int_0^y |W(x)| \, dx \leqq \int_0^{x_1} |W(x)| \, dx + (\bar{x} - x_1)\beta\left(1 - \frac{\alpha^2}{2Mk}\right) + M.$$

Dividing by y and noting that $\bar{x} \leqq y$,

$$\frac{1}{y}\int_0^y |W(x)| \, dx \leqq \frac{1}{y}\int_0^{x_1} |W(x)| \, dx + \beta\left(1 - \frac{\alpha^2}{2Mk}\right) + \frac{M}{y} \cdot$$

Letting $y \to \infty$, $\gamma \leqq \beta(1 - \alpha^2/2Mk)$, and since $\gamma \geqq \alpha$,

$$\alpha \leqq \beta\left(1 - \frac{\alpha^2}{2Mk}\right).$$

Since this holds for all $\beta > \alpha$ it must hold for $\beta = \alpha$. Hence $\alpha^3 \leqq 0$, and since $\alpha \geqq 0$, this implies $\alpha = 0$. Since $W(x) = e^{-x}S(e^x)$, this implies that $|S(y)|/y \to 0$ as $y \to \infty$. Hence if given $\epsilon > 0$, if y is large enough,

$$|S(y)| \leqq \tfrac{1}{3}\epsilon^2 y.$$

Thus $S(y(1+\epsilon)) - S(y) \leqq \frac{1}{3}\epsilon^2(y(1+\epsilon) + y) < \epsilon^2 y$, or

$$\int_y^{y(1+\epsilon)} \frac{R(u)}{u} \, du \leqq \epsilon^2 y.$$

Since $R(u) = \psi(u) - u$ and ψ is nondecreasing,

$$\frac{\psi(y)}{y(1+\epsilon)}\int_y^{y(1+\epsilon)} du - \int_y^{y(1+\epsilon)} du \leqq \epsilon^2 y.$$

Hence $\psi(y)/y \leqq (1+\epsilon)^2$. Similarly $S(y) - S(y(1-\epsilon)) \geqq -\epsilon^2 y$ for large enough y leads to $\psi(y)/y \geqq (1-\epsilon)^2$. Since ϵ is arbitrary this proves (2.7).

Supported in part by the Office of Naval Research and by the National Science Foundation, NSF GP 7477.

References

1. E. Bombieri, Sulle formule di A. Selberg generalizzate per classi di funzioni aritmetiche e le applicazioni al problema del resto nel "Primzahlsatz", Riv. Mat. Univ. Parma, (2) 3 (1962) 393–440.
2. R. Breusch, An elementary proof of the prime number theorem with remainder term, Pacific J. Math., 10 (1960) 487–497.
3. P. Erdös, On a new method in elementary number theory which leads to an elementary proof of the prime number theorem, Proc. Nat. Acad. Sci. U. S. A., 35 (1949) 374–384.
4. G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, Oxford University Press, New York, 1960.
5. N. Levinson, On the elementary proof of the prime number theorem, Proc. Edinburgh Math. Soc., (2) 15 (1966) 141–146.
6. A. Selberg, An elementary proof of the prime number theorem, Ann. of Math., (2) 50 (1949) 305–313.
7. T. Tatuzawa and K. Iseki, On Selberg's elementary proof of the prime number theorem, Proc. Japan Acad., 27 (1951) 340–342.
8. E. Wirsing, Elementare Beweise des Primzahlsatzes mit Restglied, I, J. Reine Angew. Math., 211 (1962) 205–214.
9. E. M. Wright, The elementary proof of the prime number theorem, Proc. Roy. Soc. Edinburgh. Sect. A., 63 (1951) 257–267.

AN INTRODUCTION TO HESTENES TERNARY RINGS

M. F. SMILEY, State University of New York at Albany

Dedicated to Marian and Harry

Introduction. Spectral theory for rectangular matrices goes back to the general reciprocal of E. H. Moore [9] and has been studied in some detail by Penrose [10], M. R. Hestenes [2] and Lanczos [7]. In two papers [3, 4] Hestenes has cast this theory in the framework of a theory of a ternary operation based on the observation that if A, B and C are complex m by n matrices, then so is AB^*C. The purpose of this brief expository note is to indicate the possibility that Hestenes's idea extends to structure theory in the spirit of N. Jacobson [5, 6]. R. A. Stephenson [12] has already verified the rudiments of this extension. Nonetheless, many interesting questions remain.

We begin with a quick derivation of Moore's general reciprocal since this is the generic idea on which our algebra depends. In order to present an extension of the Chevalley-Jacobson Density Theorem (which we obtained jointly with R. A. Stephenson), we have taken the liberty of presenting Jacobson's original proof [5]. A well-tempered proof due to Tate [Artin, 1] is available. One should also mention Jacobson's proof in [6] which is based on a theorem very close to our extended version even though it involves only ordinary rings.

1. E. H. Moore's general inverse. Let A be an m by n complex matrix of rank r. Because $Ax=0$ if and only if $A^*Ax=0$, A^*A is an n by n nonnegative hermitian matrix of rank r. Let $x_1, \cdots, x_n$ be an orthonormal set of eigenvectors for