Formalizing an Analytic Proof of the Prime Number Theorem

Dedicated to Mike Gordon

John Harrison

Intel Corporation

TTVSI (Mikefest)

Royal Society, London

Wed 26th March 2008 (10:00 – 10:45)

What is the prime number theorem?

Let $\pi(x)$ be the number of primes $\leq x$.

The Prime Number Theorem asserts that

$$\pi(x) \sim \frac{x}{\log(x)}$$

or in other words $\frac{\pi(x)}{x/\log(x)} \to 1$ as $x \to \infty$.

See G. J. O. Jameson *The Prime Number Theorem*, LMS Student Texts 53, Cambridge University Press 2003.

The convergence is slow, and $\int_2^x \log(t)dt$ is a better approximation, but the ratio does tend to 1.

Small values

x	$\pi(x)$	$\frac{x}{\log(x)}$	Ratio
10^2	25	21.71	1.1515
10^3	168	144.76	1.1605
10^4	1229	1085.74	1.1319
10^5	9592	8685.89	1.1043
10^6	78498	72382.41	1.0845
10^7	664579	620420.69	1.0712
10^8	5761455	5428681.02	1.0613
10^9	50847534	48254942.43	1.0537
10^{10}	455052511	434294481.90	1.0478
10^{11}	4118054813	3948131653.67	1.0430
10^{12}	37607912018	36191206825.27	1.0391
10^{13}	346065536839	334072678387.12	1.0359
10^{14}	3204941750802	3102103442166.08	1.0331
10^{15}	29844570422669	28952965460216.79	1.0308
10^{16}	279238341033925	271434051189532.38	1.0288
10^{17}	2623557157654233	2554673422960304.87	1.0270

History of the Prime Number Theorem

ca. 1800 Conjectured by several mathematicians including Gauss, but no real progress towards a proof

1847 Chebyshev proves $\frac{\pi(x)}{x/\log(x)}$ is bounded quite close to 1, and that *if* it tends to a limit, that limit is 1

1859 Riemann points out deep relationship between distribution of primes and the complex zeta function $\zeta(s)$

1896 PNT proved independently by Hadamard and de la Valée Poussin using nonvanishing of $\zeta(s)$ for $\Re s \geq 1$ and complex contour integration

1949 Elementary proof by Erdös and Selberg, using very intricate manipulations but not relying on any complex analysis

1980 More streamlined version of complex-analytic proof by Newman

Formalizing the PNT

The elementary Erdös-Selberg proof has already been formalized by a team led by Jeremy Avigad.

> Jeremy Avigad, Kevin Donnelly, David Gray, Paul Raff
> *A formally verified proof of the prime number theorem*
> ACM Transactions on Computational Logic, vol. 9, 2007

We describe the formalization of Newman’s relatively slick analytic proof.

The analytic proof is simpler and clearer, but at the cost of requiring much more mathematical machinery.

Analytic functions, contour integrals, Cauchy’s integral formula, the Riemann ζ-function, Euler’s product formula, . . .

The Solovay challenge

From Freek Wiedijk’s *The Seventeen Provers of the World*, Springer LNCS vol. 3600, 2006, p. 12:

> Bob Solovay has challenged the proof assistant community to do a formalization of the analytic proof of the Prime Number Theorem. (He claims that proof assistant technology will not be up to this challenge for decades.) This challenge is still open, as the proof of the Prime Number Theorem that Jeremy Avigad formalized was the ‘elementary’ proof by Atle Selberg.

The Solovay challenge

From Freek Wiedijk’s *The Seventeen Provers of the World*, Springer LNCS vol. 3600, 2006, p. 12:

> Bob Solovay has challenged the proof assistant community to do a formalization of the analytic proof of the Prime Number Theorem. (He claims that proof assistant technology will not be up to this challenge for decades.[a]) This challenge is still open, as the proof of the Prime Number Theorem that Jeremy Avigad formalized was the ‘elementary’ proof by Atle Selberg.

[a] Others who are more optimistic about this asked me to add this footnote in which I encourage the formalization community to prove Bob Solovay wrong.

Mathematical machinery vs. brute force (1)

In many cases, the way a human and a machine prove a theorem are very different.

> Simply put, differences abound between the way a person reasons and the way a program of the type featured here reasons. Those differences may in part explain why OTTER has succeeded in answering questions that were unanswered for decades, and also explain why its use has produced proofs far more elegant than those previously known. [Larry Wos]

And indeed, in McCune's solution of the Robbins conjecture, the theorem prover actually does better than the human.

Mathematical machinery vs. brute force (2)

There has been some work on formalizing elliptic curves in theorem provers

> Laurent Théry and Guillaume Hanrot, *Primality Proving with Elliptic Curves*, TPHOLs 2007
>
> Joe Hurd, *Formally Verified Elliptic Curve Cryptography*, presentation from 2007.

In both cases, associativity of addition led to huge algebraic computations that taxed or even defeated the capacity of the theorem provers’ algebraic decision procedures.

Mathematical machinery vs. brute force (3)

From Dan Grayson:

> But why not enter one of the usual human-understandable proofs that + is associative? Too many prerequisites from algebraic geometry? [. . .] The proof I like most is to use the Riemann-Roch theorem to set up a bijection between the rational points of an elliptic curve and the elements of the group of isomorphism classes of invertible sheaves of degree 0. That's a lot of background theory, probably too much for this stage of development, but then the "real" reason for associativity is that tensor product of R-modules is an associative operation up to isomorphism.

Wouldn't it be nicer to formalize that instead?

HOL as a general mathematical framework

As well as HOL itself, Mike Gordon pioneered the use of the theorem prover as a universal framework to replace ad hoc extensions.

- Programming logics: define the underlying semantics of a programming language inside HOL and *derive* the usual programming rules. (Deep and shallow embeddings, . . .)
- Theory development: construct mathematical structures like natural numbers, lists, datatypes etc. instead of adding new axioms. (The advantages of honest toil over theft.)

For applications in floating-point arithmetic, nice to bring all these together with a construction of the reals.

Foundations for this work

Our development of ‘mathematical machinery’ leading up to Cauchy’s integral formula:

- *Constructing the Real Numbers in HOL*, TPHOLs 1992
- *Complex quantifier elimination in HOL*, TPHOLs 2001 (Category B)
- *A HOL theory of Euclidean space*, TPHOLs 2005
- *Formalizing Basic Complex Analysis*, Festschrift for Andrzej Trybulec, 2007

Source text

‘Second proof’ from *Analytic Number Theory* by Newman himself (Springer, 1998). Can be divided into five components:

1. The Newman-Ingham “Tauberian” analytical lemma.
2. Basic properties of the Riemann ζ-function and its derivative, including the Euler product.
3. Chebyshev’s elementary proof that $\sum_{p\leq n} \frac{\log p}{p} - \log n$ is bounded.
4. Application of analytic lemma to get summability of $\sum_n (\sum_{p\leq n} \frac{\log p}{p} - \log n - c)/n$ for some constant c.
5. From summability, deduce $\sum_{p\leq n} \frac{\log p}{p} - \log n$ tends to a limit.
6. Derivation of the PNT from that limit using partial summation.

The de Bruijn factor

An interesting way of measuring the difficulty of a formalization is the *de Bruijn factor*.

`http://www.cs.ru.nl/~freek/factor/`

> I defined the de Bruijn factor to be the quotient of the size of a formalization of a mathematical text and the size of its informal original. To make this specific, I made the following choices:
>
> - The size of the informal text is the size of a TeX encoding.
> - To be independent of arbitrary factors like lengths of identifiers and amount of whitespace, the files should be compared compressed with the Unix utility gzip.

Experience shows that the dB factor in many cases is around 4.

De Bruijn factors

De Bruijn factors for various parts of this PNT proof:

	Part of proof	dB factor
1	Analytical lemma	8.2
2	ζ-function	81.3
3	Chebyshev bound	28.2
4	Summability	11.0
5	Limit	5.4
6	PNT	30.4

By normal standards these range from high-ish to outrageously high.

Newman’s inexplicit style

However, some of the parts in Newman are not really ‘proofs’:

2. Let us begin with the well-known fact about the ζ-function: $(z-1)\zeta(z)$ is analytic and zero free throughout $\Re z \geq 1$.
3. In this section we begin with Tchebyshev’s observation that $\sum_{p\leq n} \frac{\log p}{p} - \log n$ is bounded, which he derived in a direct elementary way from the prime factorization on $n!$
6. The point is that the Prime Number Theorem is easily derived from ‘$\sum_{p\leq n} \frac{\log p}{p} - \log n$ converges to a limit’ by a simple summation by parts which we leave to the reader.

For parts 1, 4 and 5, the de Bruijn factor is ‘only’ about 8.

Moreover, part 2 vs. actual source (Bak & Newman) has a dB factor of only 3.1.

Why the dB factor is large (1)

The following short passage:

- determine $\delta = \delta(R) > 0$, $\delta \leq \frac{1}{2}$ and an $M = M(R)$ so that $F(z + w)$ is analytic and bounded by M in $-\delta \leq \Re z$, $|z| \leq R$.

takes 164 lines to formalize:

```
SUBGOAL_THEN
 ‘?d. &0 < d /\ d <= R /\
      (\z. f(w + z)) holomorphic_on {z | Re(z) >= --d /\ abs(Im z) <= R}‘
...
SUBGOAL_THEN
 ‘?M. &0 < M /\
      !z. Re z >= --d /\ abs (Im z) <= R /\ Re(z) <= R
          ==> norm(f(w + z):complex) <= M‘
...
```

Why the dB factor is large (2)

The second equation here:

- $f(z) = \sum_{n=1}^{\infty} \frac{1}{n^z} \left(\sum_{p \leq n} \frac{\log p}{p} \right) = \sum_p \frac{\log p}{p} \left[\sum_{n \geq p} \frac{1}{n^z} \right]$

takes 116 lines to formalize:

```
let lemma = prove
 (‘vsum (1..n) (\m. vsum {p | prime p /\ p <= m} (\p. f p m)) =
   vsum {p | prime p /\ p <= n} (\p. vsum (p..n) (\m. f p m))‘,
  SIMP_TAC[VSUM_VSUM_PRODUCT; FINITE_NUMSEG; FINITE_ATMOST] THEN
  REWRITE_TAC[IN_ELIM_THM; IN_NUMSEG; GSYM CONJ_ASSOC] THEN
  MATCH_MP_TAC VSUM_EQ_GENERAL_INVERSES THEN
...
MATCH_MP_TAC LOG_MONO_LE_IMP THEN
ASM_REWRITE_TAC[GSYM REAL_OF_NUM_ADD; REAL_OF_NUM_LT; LT_NZ] THEN
REAL_ARITH_TAC]);;
```

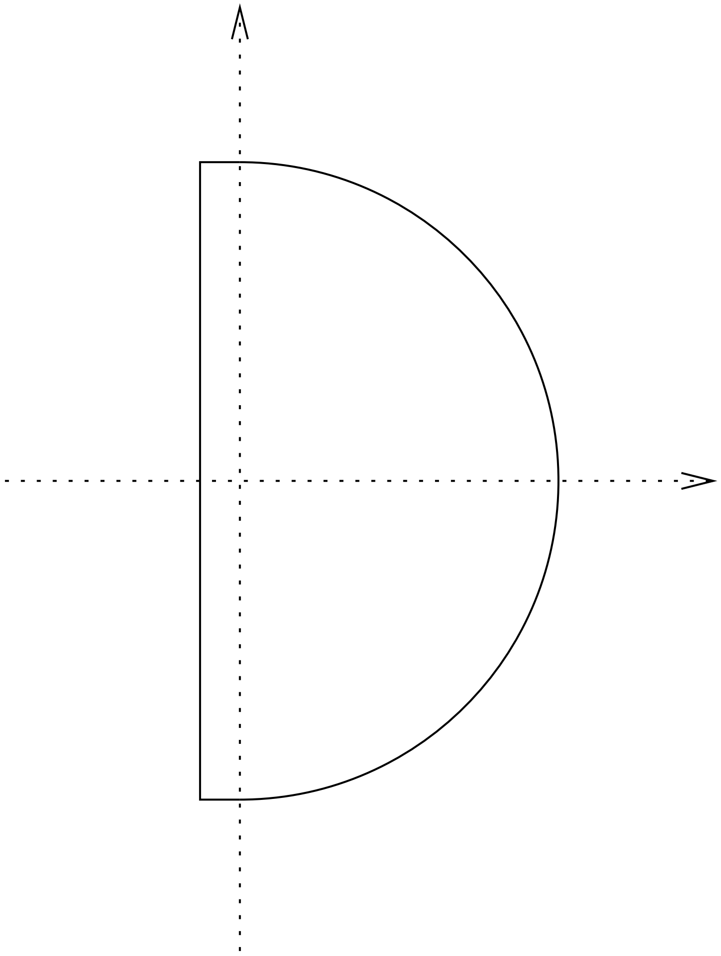
The analytic lemma

The heart of Newman’s proof is the following analytic lemma.

> **Theorem.** *Suppose* $|a_n| \leq 1$, *and form the series* $\sum a_n n^{-z}$ *which clearly converges to an analytic function* $F(z)$ *for* $\Re z > 1$. *If, in fact,* $F(z)$ *is analytic throughout* $\Re z \geq 1$, *then* $\sum a_n n^{-z}$ *converges throughout* $\Re z \geq 1$.

This is where the analytical machinery comes in, and its dB factor seems representative of the proof as a whole.

The contour



The key contour integral

$$2\pi i F(w) = \int_{\Gamma} F(z+w) N^{z} \left(\frac{1}{z} + \frac{z}{R^2} \right) dz$$

This is the application of Cauchy’s integral formula

$$2\pi i \mathsf{WN}(\Gamma, z) f(z) = \int_{\Gamma} \frac{f(w)}{w-z}$$

where the *winding number* is defined by:

$$\mathsf{WN}(\Gamma, z) = \frac{1}{2\pi i} \int_{\Gamma} \frac{1}{w-z}$$

HOL version of Cauchy’s integral formula

We have Cauchy’s integral for a region with a convex shape:

```
|- convex s /\ f holomorphic_on s /\
   z IN interior(s) /\
   valid_path g /\ (path_image g) SUBSET (s DELETE z) /\
   pathfinish g = pathstart g
   ==> ((\w. f(w) / (w - z)) has_path_integral
        (Cx(&2) * Cx(pi) * ii * winding_number(g,z) *
         f(z))) g
```

where the winding number is defined as:

```
|- winding_number(g,z) =
        Cx(&1) / (Cx(&2) * Cx(pi) * ii) *
        path_integral g (\w. Cx(&1) / (w - z))
```

Applying the integral formula

The only problem in applying the integral formula is proving that the winding number of our contour is 1

- Could do the integral, which is possible but messy and too ‘special’
- Could appeal to the fact that a closed curve has WN in the set $\{-1, 0, +1\}$ and exclude the first two
- Can piece together WN for two parts of the contour knowing they’re strictly between 0 and 1.

Yet in the informal proof, this is just obvious from the intuitive meaning of WN.

Results about winding numbers (1)

We deal with winding numbers of paths that are not closed; they are additive when we join two paths:

```
|- valid_path g1 /\ valid_path g2 /\
   ~(z IN path_image g1) /\ ~(z IN path_image g2)
   ==> winding_number(g1 ++ g2,z) =
       winding_number(g1,z) + winding_number(g2,z)
```

The WN is an integer *if and only if* the path is closed:

```
|- valid_path g /\ ~(z IN path_image g)
   ==> (complex_integer(winding_number(g,z)) <=>
       pathfinish g = pathstart g)
```

Results about winding numbers (2)

We can package this up conveniently to deduce several key properties from components using additivity:

```
|- (valid_path g1 /\
     ~(z IN path_image g1) /\
     &0 < Re(winding_number(g1,z))) /\
   (valid_path g2 /\
     ~(z IN path_image g2) /\
     &0 < Re(winding_number(g2,z))) /\
   pathfinish g1 = pathstart g2
   ==> valid_path(g1 ++ g2) /\
       ~(z IN path_image(g1 ++ g2)) /\
       &0 < Re(winding_number(g1 ++ g2,z))
```

Results about winding numbers (3)

We can prove the WN for each component is positive very easily, e.g. for line segments:

```
|- &0 < Im((b - a) * cnj(b - z))
   ==> &0 < Re(winding_number(linepath(a,b),z))
```

and prove that WN for a component is < 1 by showing that there's some ray it doesn't cross:

```
|- valid_path g /\ ~(z IN path_image g) /\ ~(w = z) /\
   (!a. &0 < a
        ==> ~(z + (Cx a * (w - z)) IN path_image g))
   ==> Re(winding_number(g,z)) < &1
```

Hence we can deduce that the WN for our contour is 1, as required.

The end result

The Prime Number Theorem formalized:

```
|- ((\n. &(CARD {p | prime p /\ p <= n}) /
         (&n / log(&n)))
    ---> &1) sequentially
```

Conclusion

We really are able to formalize a proof that needs a bit of mathematical machinery.

The analytic proof, even though not trivial, still seems less work than the Erdös-Selberg proof.

If we add all the work taken in developing the theory of analytic functions, it’s almost certainly more.

However that is independently interesting and useful for other applications too.