Analytic Approaches to the PNT

Daniel Rui

UCLA

March 6, 2024

Table of Contents

1 Pseudo-History/Motivation
- The Zeta Function
- Cheap Mertens, von Mangoldt
- Dirichlet Ring, Mobius Inversion, Divisor Sums, Mobius Cancellation
- Dyadic Decomposition, Mertens1
- Mertens2, Prime-Supported Measures
- Selberg Symmetry, Brun-Titchmarsh, Parity Problem

2 Banach Algebra Proof of PNT
- Setup/Outline
- Constructing Banach Algebra Norm
- Non-Trivial Spectrum
- Breaking Parity Barrier

Notation

For f, g defined for $x \in I$, we say $f \ll g$ or $f = O(g)$ if there exists an absolute constant $C > 0$ s.t. $|f(x)| \leqslant Cg(x)$ for all $x \in I$. Note that $f \ll g \iff |f| \ll g$.

If we add subscripts $f \ll_{a,b,c} g$ or $f = O_{a,b,c}(g)$, then C is allowed to depend on the parameters a, b, c, and on nothing else.

The notation $f \gg g$ is defined analogously, and $f \asymp g$ means the conjunction of both $f \ll g$ and $f \gg g$. A stronger version is $f \sim g$, where not only is $\frac{f}{g}$ bounded away from 0 and ∞, but in fact converges to 1 as x tends to some limit L ($L = \infty$ unless otherwise indicated).

If I go to the trouble of putting an explicit number (or explicitly the letter C) in the big-O notation, e.g. $f = O(3g)$ or $f = O(Cg)$, then one should interpret this as me explicitly providing the absolute constant, i.e. $|f(x)| \leqslant 3g(x)\ \forall x \in I$.

Table of Contents

1 Pseudo-History/Motivation
- The Zeta Function
- Cheap Mertens, von Mangoldt
- Dirichlet Ring, Mobius Inversion, Divisor Sums, Mobius Cancellation
- Dyadic Decomposition, Mertens1
- Mertens2, Prime-Supported Measures
- Selberg Symmetry, Brun-Titchmarsh, Parity Problem

2 Banach Algebra Proof of PNT
- Setup/Outline
- Constructing Banach Algebra Norm
- Non-Trivial Spectrum
- Breaking Parity Barrier

In middle school, people learn about the difference of squares formula $a^2 - b^2 = (a-b)(a+b)$ (i.e. the factorization of the simples nontrivial quadratic $x^2 - 1$). One might see that this formula is amenable to recursion: $x^4 - 1 = (x^2-1)(x^2+1) = (x-1)(x+1)(x^2+1)$, or more generally

$$(x^{2^n} - 1) = (x-1)\prod_{k=1}^{n-1}(x^{2^k} + 1).$$

From the study of geometric series (themselves extremely natural — c.f. Zeno's paradox ca. 300BC on successive divisions of time; or these nice tilings, i.e. successive divisions of space), we know that

$$\frac{(x^{2^n}-1)}{(x-1)} = 1 + x + \ldots + x^{2^n-1}.$$

So, we get the following functional identity:

$$1 + x + \ldots + x^{2^n-1} = (1+x)(1+x^2)\cdots(1+x^{2^{n-1}}),$$

which we can think of as encoding the *additive decomposition* result that all numbers in $\{0, \ldots, 2^n - 1\}$ have a unique n-bit binary representation. (By going to infinite series/products, can get formally at least the same result for arbitrary bit binary representations.)

By having the variable in the exponent instead (resulting in the parallel theory of Dirichlet series instead of power series), we can encode *multiplicative decomposition* results in terms of functional identities.

The most famous multiplicative decomposition theorem is of course the *fundamental theorem of arithmetic*, on the unique factorization of integers into primes, and can be encoded as the (formal) Dirichlet series functional identity:

$$\sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_p \left(\sum_{m=0}^{\infty} \frac{1}{p^{ms}} \right) = \prod_p \left(1 - \frac{1}{p^s} \right)^{-1}.$$

It is clear the fundamental ideas behind the Riemann zeta function $\zeta(s)$ are very natural, and from a practical point of view, inevitable to be discovered.

For the more theoretically minded however, Tate's thesis reveals that $\zeta(s)$ (its Euler product form, as well as deeper functional equations than the one above) fall out naturally once one has a sufficiently-developed theory of Fourier analysis on the adeles, so indeed all the zeta functions are inevitable from a deep mathematical perspective (as deep and inevitable as Fourier analysis), not just from pattern-spotting in basic mathematical formulas like the first motivation I presented.

Table of Contents

1 Pseudo-History/Motivation
- The Zeta Function
- Cheap Mertens, von Mangoldt
- Dirichlet Ring, Mobius Inversion, Divisor Sums, Mobius Cancellation
- Dyadic Decomposition, Mertens1
- Mertens2, Prime-Supported Measures
- Selberg Symmetry, Brun-Titchmarsh, Parity Problem

2 Banach Algebra Proof of PNT
- Setup/Outline
- Constructing Banach Algebra Norm
- Non-Trivial Spectrum
- Breaking Parity Barrier

Anyways, if one has acquired some basic analytic instincts, the Euler product formula above provides immediate non-trivial insight into the primes: taking logs (because we like sums better than products),

$$\log \zeta(s) := \log\left(\sum_{n=1}^{\infty} \frac{1}{n^s}\right) = -\sum_p \log\left(1 - \frac{1}{p^s}\right)$$

and taking advantage of the fact (Taylor's theorem) that

$$f(x) := \log(1+x) = x + O\left(\max_{c \in [-\frac{1}{2}, \frac{1}{2}]} |f''(c)| \cdot \frac{x^2}{2!}\right) = x + O(2x^2),$$

we get that

$$-\log\left(1 - \frac{1}{p^s}\right) = \frac{1}{p^s} + O\left(2 \cdot \frac{1}{p^{2s}}\right)$$

When summed over p, the error term $\sum_p \frac{1}{p^{2s}} \leqslant \sum_n \frac{1}{n^2} < 2$, and so we get for all $s > 1$:

$$\sum_p \frac{1}{p^s} = \log\left(\sum_n \frac{1}{n^s}\right) + O(4 \cdot 1).$$

And by the *quantitative integral test for monotone functions* (MonoQuaInt for short),

Lemma 1.1: MonoQuaInt

For $a < b$ in $\mathbb{R}$ and $f : [a, b] \to \mathbb{R}$ a monotone function,

$$\sum_{n \in \mathbb{Z} \cap [a,b]} = \int_a^b f(t)\, dt + O(|f(a)| + |f(b)|).$$

we get that

$$\zeta(s) := \sum_{n=1}^{\infty} \frac{1}{n^s} = \frac{1}{s-1} + O(1),$$

and hence we arrive at

$$\sum_p \frac{1}{p^s} = \log\left(\frac{1}{s-1}\right) + O(1). \qquad (⚘)$$

Taking $s \searrow 1$ (monotone convergence yada yada), we get Euler's 1737 theorem $\sum_p \frac{1}{p} = +\infty$.

This is an analytic proof of the Euclid's ca. 300BC theorem on the infinitude of primes, and arguably the first showcase of analytic techniques to prove things about the primes. Lest one discount these ideas due to their seeming triviality, Dirichlet proved in 1837 the infinitude of primes in admissible arithmetic progressions, using the very same ideas (divergence of reciprocal sums) plus the Fourier theory of finite abelian groups to decompose the non-multiplicative function $[n \mapsto 1_{n \equiv a\ (q)}]$ into completely multiplicative functions (Dirichlet characters).

Even better, I haven't even finished discussing the consequences of the above formula (⚘)!

The above formula (†) is for $s > 1$, but amazingly, we can use it to derive some information AT $s = 1$! We do this via the *Rankin trick*: for fixed $x \geqslant 1$, we have that

$$1 \leqslant n \leqslant x \iff 0 \leqslant \log n \leqslant \log x \iff 0 \leqslant \frac{\log n}{\log x} \leqslant 1 \iff 1 \leqslant n^{1/\log x} \leqslant e,$$

and so again for any fixed x, $\sum_p \frac{1}{p}$ and $\sum_p \frac{1}{p^s}$ for $s = s_x := 1 + \frac{1}{\log x}$ are within a factor of e of each other. Plugging in this s_x value in the above formula (†) yields

$$\sum_{p \leqslant x} \frac{1}{p} \asymp \sum_{p \leqslant x} \frac{1}{p^{s_x}} \leqslant \sum_p \frac{1}{p^{s_x}} = \log\log x + O(1).$$

Proposition 1.1: *Cheap Mertens 2*

For $x \geqslant 10$, $\sum_{p \leqslant x} \frac{1}{p} \ll \log\log x$. (In fact $\sum_{p \leqslant x} \frac{1}{p} \leqslant e \cdot \log\log x + O(1)$.)

Compare $\sum_{n \leqslant x} \frac{1}{n \log n} = \log\log x + O(1)$, so Mertens2 is some kind of quantification of the heuristic $p_n \approx n \log n$.

Returning to

$$-\log\zeta(s) = \sum_p \log\left(1 - \frac{1}{p^s}\right) = \log\left(\frac{1}{s-1}\right) + O(1),$$

naively differentiating both sides (we like $\frac{1}{s-1}$ — i.e. meromorphic functions/behavior — much better than $\log(\frac{1}{s-1})$!) produces

$$-\frac{\zeta'(s)}{\zeta(s)} = \sum_p \frac{\log p}{p^s - 1} = \frac{1}{s-1} + O(1).$$

Note that we can massage

$$\sum_p \frac{\log p}{p^s - 1} = \sum_p \frac{\log p}{p^s}\frac{1}{1 - (\frac{1}{p^s})} = \sum_p \frac{\log p}{p^s}\left(1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \dots\right) =: \sum_n \frac{\Lambda(n)}{n^s}.$$

To justify the LHS differentiation, one can do it from "first principles", using the DCT and FTC: using the DCT, first show that the term-by-term *integral* of the above Dirichlet series for $-\frac{\zeta'(s)}{\zeta(s)}$ is that of $\log\zeta(s)$, and then appeal to the FTC to go backward; OR, one can check that the series converges locally uniformly (using Weierstrass M-test), and then use that the locally uniform limit f of holomorphic functions f_n is holomorphic, and moreover have the derivatives $f_n^{(k)}$ converge locally uniformly to $f^{(k)}$.

To justify the RHS (of course, $f = g + O(1)$ does NOT imply $f' = g' + O(1)$ in general!!!), again we can appeal to regularity: $\zeta(s) - \frac{1}{s-1}$ and $\log\zeta(s) - \log(\frac{1}{s-1})$ are smooth approaching from $s \searrow 1$ (indeed they will eventually be revealed to holomorphically extend to even left of $s = 1$), so the $O(1)$-error is a smooth function on $s \in [1, 2]$ and its derivative is bounded there too.

Again using the *Rankin trick* on (🍃) (costing a factor of e), one obtains

Proposition 1.2: Cheap Mertens 1

For $x \geqslant 2$, $\sum_{p \leqslant x} \frac{\log p}{p} \leqslant \sum_{n \leqslant x} \frac{\Lambda(n)}{n} \ll \log x$.

Heuristic 1: von Mangoldt $\Lambda(n) \approx \log n \cdot 1_{Primes}(n)$

Mertens1's 2 sums differ by $O(1)$: their difference is $\sum_{j=2}^{\infty} \sum_{p \leqslant x^{1/j}} \frac{\log p}{p^j}$, and $j \geqslant 2 \implies \frac{\log p}{p^j} \leqslant \frac{1}{2^{j-1.5}} \cdot \frac{\log p}{p^{1.5}}$, where $\sum_n \frac{\log n}{n^{1.5}} = O(1)$ by p-test, and $\sum_j \frac{1}{2^{j-1.5}} = O(1)$ by geometric series.

Another realization: $\sum_{n \leqslant x} \Lambda(n)$ and $\sum_{p \leqslant x} \log p$ differ by $O(\sqrt{x} \log^2 x)$, because to have $n = p^j$ with $j \geqslant 2$ and $n \leqslant x$, we must have $p \leqslant \sqrt{x}$, so we have at most $\sqrt{x}$ such primes p, which can each have at most $\ll \log x$ powers $p^j = n \leqslant x$, meaning we have $O(\sqrt{x} \log x)$ terms coming from $n = p^{\geqslant 2}$, each of which contributes $\log n \leqslant \log x$.

Allow me to recall the following two asymptotics:

$$\zeta(s) = \frac{1}{s-1} + O(1).$$

$$-\frac{\zeta'(s)}{\zeta(s)} = \frac{1}{s-1} + O(1).$$

Heuristic 2: von Mangoldt $\Lambda(n) \approx 1$ on average

(9/28/24 254A notes) This is the first indication that $\Lambda(n)$ and 1 behave similarly/have similar estimates. For instance, famously

$$\sum_{n \leqslant x} 1 = x + O(1),$$

and ultimately, the prime number theorem is (equivalent to) the statement that

$$\sum_{n \leqslant x} \Lambda(n) = x + o(x).$$

Table of Contents

1 Pseudo-History/Motivation
- The Zeta Function
- Cheap Mertens, von Mangoldt
- Dirichlet Ring, Mobius Inversion, Divisor Sums, Mobius Cancellation
- Dyadic Decomposition, Mertens1
- Mertens2, Prime-Supported Measures
- Selberg Symmetry, Brun-Titchmarsh, Parity Problem

2 Banach Algebra Proof of PNT
- Setup/Outline
- Constructing Banach Algebra Norm
- Non-Trivial Spectrum
- Breaking Parity Barrier

I now prove the full Mertens 1 and 2 theorems, because their proofs contain some really important ideas (and they happen to be useful, e.g. in most of Terry's later 254A notes!). The key identity pops out naturally (see my 9/30/22 254A notes) by considering the Dirichlet series for $\zeta, -\frac{\zeta'}{\zeta}$ whom we've met before, and ζ' whom we've not met yet. By term-by-term differentiation (see above slide for justifications), we have that

$$\zeta'(s) = \sum_{n=1}^{\infty} \frac{-\log n}{n^s},$$

but also $-\zeta'(s) = -\frac{\zeta'(s)}{\zeta(s)} \cdot \zeta(s)$, which equals

$$-\zeta'(s) = \left(\sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^s}\right) \cdot \left(\sum_{m=1}^{\infty} \frac{1}{m^s}\right) = \sum_{n,m=1}^{\infty} \frac{\Lambda(n)}{(nm)^s} = \sum_{M=1}^{\infty} \frac{\left(\sum_{d|M} \Lambda(d)\right)}{M^s}.$$

By uniqueness of Dirichlet series, we get that $\log n = \sum_{d|n} \Lambda(n)$.

2nd perspective: of course one can prove the identity without mention of Dirichlet series (I just felt it was a more natural way of presenting it since I was doing Dirichlet series anyways); in fact the identity is equivalent to the fundamental theorem of arithmetic:

$$n = \prod_p p^{\nu_p(n)} \iff \log n = \sum_p \nu_p(n) \log p = \sum_p \sum_{j \geqslant 1 : p^j | n} \log n = \sum_{d|n} \Lambda(d).$$

3rd perspective: returning to $-\zeta' = (-\frac{\zeta'}{\zeta}) \cdot \zeta$, the corresponding identity on arithmetic functions (the correspondence being the homomorphism of commutative rings $([f : \mathbb{N} \to \mathbb{C}], +, \star) \leftrightarrow (\sum_n \frac{f(n)}{n^s}, +, \cdot)$ where $\delta := 1_{n=1} : \mathbb{N} \to \mathbb{C}$ is the multiplicative unit, and $\zeta \cdot \zeta^{-1} = 1 \Leftrightarrow \mathbf{1} \star \mu = \delta$ "Mobius inversion") is

$$\log = \Lambda \star \mathbf{1}.$$

Mobius inversion (i.e. multiplying by μ) yields $\Lambda = \mu \star \log$.

Dirichlet convolution identities involving log

Writing this out we get

$$\Lambda(n) = \sum_{d|n} \mu(d) \log(\tfrac{n}{d}) = (\log n) \sum_{d|n} \mu(d) - \sum_{d|n} \mu(d) \log(d)$$
$$= (\log n) \cdot (\mu \star \mathbf{1}) - ([\log \cdot \mu] \star \mathbf{1}) = 0 - ([\log \cdot \mu] \star \mathbf{1}).$$

Conceptually, what's happening in this calculation is that the homomorphism property $\log(mn) = \log(m) + \log(n)$ leads to a "Leibniz formula" $\log \cdot [f \star g] = [\log \cdot f] \star g + f \star [\log \cdot g]$, which we then apply to $f = \mu$ and $g = \mathbf{1}$:

$$0 = \log \cdot \delta = \log \cdot [\mu \star 1] = [\log \cdot \mu] \star \mathbf{1} + \underbrace{\mu \star [\log \cdot 1]}_{=[\mu \star \log] = \Lambda}.$$

One more application of Mobius inversion produces $[\log \cdot \mu] = -\mu \star \Lambda$.

On divisor sums $\sum_{d|n}$, like $\log n = \sum_{d|n} \Lambda(d)$

Unfortunately, divisors behave in an extremely irregular way — there seem to be little relating one row of the below table to the next.

$F(n) \setminus d$	1	2	3	4	5	6	7	8	9	10	11	12
$F(1) =$	$f(1)$											
$F(2) =$	$f(1)$	$f(2)$										
$F(3) =$	$f(1)$		$f(3)$									
$F(4) =$	$f(1)$	$f(2)$		$f(4)$								
$F(5) =$	$f(1)$				$f(5)$							
$F(6) =$	$f(1)$	$f(2)$	$f(3)$			$f(6)$						
$F(7) =$	$f(1)$						$f(7)$					
$F(8) =$	$f(1)$	$f(2)$		$f(4)$				$f(8)$				
$F(9) =$	$f(1)$		$f(3)$						$f(9)$			
$F(10) =$	$f(1)$	$f(2)$			$f(5)$					$f(10)$		
$F(11) =$	$f(1)$										$f(11)$	
$F(12) =$	$f(1)$	$f(2)$	$f(3)$	$f(4)$		$f(6)$						$f(12)$

However, the *columns* of the above table behave *extremely* regularly, in that every n is divisible by 1, every 2nd n is divisible by 2, every 3rd n is divisible by 3, and so on. Thus, although for just **single** values of n it is difficult to understand the behavior of the divisors, over **multiple** values of n the regularity of the rows might be able to help.

In other words, the **individualized** behavior of divisors for any given n may be hard to understand, but the **average** behavior over the divisors of n over all $n \in [1..N]$ is approximately that 1 will contribute all the time, 2 will contribute about half the time, 3 will contribute about a third of the time, and so on. More rigorously, the previous sentence says that the average $\frac{1}{N}\sum_{n=1}^{N} F(n)$, although difficult to analyze when summed over the rows first and then the columns, becomes much easier when summed over the columns first and then the rows, yielding

$$\frac{1}{N}\sum_{n=1}^{N} F(n) = \frac{1}{N}\sum_{n=1}^{N}\sum_{d|n} f(d) = \frac{1}{N}\sum_{d=1}^{N}\sum_{n\leqslant N:d|n} f(d) = \frac{1}{N}\sum_{d=1}^{N} f(d)\left\lfloor\frac{N}{d}\right\rfloor$$
$$= \frac{1}{N}\sum_{d=1}^{N} f(d)\left(\frac{N}{d} + O(1)\right) = \sum_{d=1}^{N}\frac{f(d)}{d} + O\left(\frac{1}{N}\sum_{d=1}^{N} f(d)\right).$$

Then for $f := \Lambda$, the 1st term on the RHS is exactly the sum in Mertens1!

Side Quest (not the main storyline)

The version of previous slides (all taken from my late2021-early2022 dirichletPrimes.tex) without all my commentary: (this is how Terry explained it in 254A) if $F(n) = \sum_{d|n} f(d)$, then

$$\sum_{n\leqslant x} F(n) = \sum_{n\leqslant x}\sum_{d|n} f(d) = \sum_{d\leqslant x}\sum_{m\leqslant \frac{x}{d}} f(d)$$

$$= \sum_{d\leqslant x} f(d) \underbrace{\left(\sum_{m\leqslant \frac{x}{d}} 1\right)}_{=\frac{x}{d}+O(1)} = \sum_{d\leqslant x} f(d)\cdot\frac{x}{d} + O\left(\sum_{d\leqslant x} f(d)\right) \qquad (\Sigma)$$

Terry: "Standard tip: Fubini! Analytic number theorists would rather see a double sum than a single sum." And more specifically: "We always want the innermost sum variable to behave best/inner sum to be understood best", e.g. in this case going from $d \mid n$ to the simple range $m \leqslant \frac{x}{d}$.

Mobius function cancellation

Before we move on, I'd like to say a little more on the Mobius inversion formula. Although trivial/elementary, and besides its usefulness/practicality, it expresses something quite deep about the Mobius function μ: by multiplying (not Dirichlet convolving!) any compactly supported arithmetic function $f : \mathbb{N} \to \mathbb{C}$ with $\mu \star \mathbf{1} = \delta$ and them summing $\sum_{n \leqslant x}$ (using the Fubini/divisor sum trick), we get

$$f(1) = \sum_{n \leqslant x} f(1) 1_{n=1} = \sum_{n \leqslant x} \sum_{d|n} \mu(d) f(n) = \sum_{d \leqslant x} \mu(d) \sum_{m \leqslant x/d} f(dm). \qquad (\lightning)$$

One can think of ($\lightning$) as expressing that μ must exhibit a significant amount of **cancellation** (i.e. *oscillation* information beyond just L^1-bound $|\mu(n)| \leqslant 1$). Indeed, the PNT is equivalent to the assertion that asymptotically, μ is $+1$ just as often as it is -1, i.e. $\sum_{n \leqslant x} \mu(n) = o(x)$.

So, we have understand well the sums $\sum_{d\leqslant x}\mu(d)F(d)$ for functions of the form $F(d)=\sum_{m\leqslant x/d}f(dm)$ for some f. We can get a decent understanding of such F by using MonoQuaInt: $F(d)=\int_1^{x/d}f(dt)\,dt+O(f(d)+f(x))$. We like the concrete endpoint 1 of the integral, but don't like that the integrand depends on d. Variable change $u=\frac{d}{x}t \iff dt=xu$ and define $g(u):=f(xu)$ to get $F(d)=\frac{x}{d}\int_{d/x}^1 g(u)\,du+O(g(\frac{d}{x})+g(1))$. So, from ↯ $g(\frac{1}{x})=f(1)=\sum_{d\leqslant x}\mu(d)F(d)$, we get

Proposition 1.3: Mobius cancellation

For any function G and its derivative $g:=G'$,

$$g(\tfrac{1}{x})=x\sum_{d\leqslant x}\mu(d)\frac{(G(1)-G(\frac{d}{x}))}{d}+O\left(\sum_{d\leqslant x}\mu(d)(g(\tfrac{d}{x})+g(1))\right).$$

Table of Contents

1 Pseudo-History/Motivation
- The Zeta Function
- Cheap Mertens, von Mangoldt
- Dirichlet Ring, Mobius Inversion, Divisor Sums, Mobius Cancellation
- Dyadic Decomposition, Mertens1
- Mertens2, Prime-Supported Measures
- Selberg Symmetry, Brun-Titchmarsh, Parity Problem

2 Banach Algebra Proof of PNT
- Setup/Outline
- Constructing Banach Algebra Norm
- Non-Trivial Spectrum
- Breaking Parity Barrier

Anyways, let me rewrite the divisor sum average formula ($\overline{\Sigma}$) for $f := \Lambda$:

$$\frac{1}{x}\sum_{n\leqslant x}\log n = \sum_{d\leqslant x}\frac{\Lambda(d)}{d} + O\left(\frac{1}{x}\sum_{d\leqslant x}\Lambda(d)\right).$$

Computation 1.1: Log Sum

Using MonoQuaInt, $\sum_{n\leqslant x}\log n = x\log x - x + O(1+\log x)$ for $x \geqslant 1$.

We'd be done if we could get something like

Proposition 1.4: Chebyshev upper bound

$\sum_{d\leqslant x}\Lambda(d) = O(x)$.

By Heuristic 1, this is equivalent to showing $\sum_{p\leqslant x}\log p \ll x$.

Dyadic Decomposition

Here we come to one of the most important (Zeno-esque) philosophies in analysis (especially harmonic analysis): *to study a sum of some things, first study the sum of only half of those things*. This is the principle of **dyadic decomposition**. In our case, dyadic decomposition ("glorified geometric series") tells us that our desired bound $\sum_{p\leqslant x} \log p \ll x$ is equivalent to the bound $\sum_{x<p\leqslant 2x} \log p \ll x$.

And this in turn can be done using just the most delightful "rabbit out of a hat" magic trick: the binomial coefficient $\binom{2N}{N}$ is an integer whose prime factorizations contains every prime $N < p \leqslant 2N$. Thus $\prod_{N<p\leqslant 2N} p \leqslant \binom{2N}{N} \leqslant 2^{2N}$, and taking logarithms produces $\sum_{N<p\leqslant 2N} p \leqslant (2\log 2)N$. (It is miraculous how close this simple argument gets to the true value: $2\log 2 \approx 1.386294$)

Side quest (not main storyline)

Let me just make a nice box for Mertens1 (recall Heuristic 1):

Theorem 1.1: Mertens1

$$\sum_{p \leqslant x} \frac{\log p}{p} = \sum_{n \leqslant x} \frac{\Lambda(n)}{n} + O(1) = \log x + O(1).$$

(10/5/22 254A notes) By the same idea as in dyadic decomposition of having a lower bound comparable to x, Mertens1 also gives Chebyshev lower bound:

$$\sum_{n \leqslant x} \Lambda(n) \geqslant \sum_{\epsilon x \leqslant n \leqslant x} \Lambda(n) \geqslant \epsilon x \cdot \sum_{\epsilon x \leqslant n \leqslant x} \frac{\Lambda(n)}{n}$$
$$= \epsilon x \cdot (\log x - \log \epsilon x + O(1)) \gg_{\epsilon} x.$$

Another application of **dyadic decomposition**: the following two asymptotics are equivalent:

$$\sum_{n \leqslant x} \Lambda(x) \sim x \iff \pi(x) := \sum_{n \leqslant x} 1_{n=p}(n) \sim \frac{x}{\log x}$$

Indeed, if you believe me (32, 45) that the LHS is $\iff \sum_{x<n\leqslant 2x} \Lambda(n) \sim x$ and the RHS is $\iff \sum_{x<p\leqslant 2x} 1 \sim \frac{x}{\log x}$, then the equivalence is obvious, just based on Heuristic 1 $\sum_{x<n\leqslant 2x} \Lambda(n) = \sum_{x<p\leqslant 2x} \log p + O(\sqrt{x}\log^2 x)$ and

$$\sum_{x<n\leqslant 2x} \log x \cdot 1_{n=p} \leqslant \underbrace{\sum_{x<n\leqslant 2x} \log n \cdot 1_{n=p}}_{\sim \sum_{x<n\leqslant 2x} \Lambda(n)} \leqslant \sum_{x<n\leqslant 2x} \log(2x) \cdot 1_{n=p},$$

where $\log x \sim \log(2x)$ implies that everything collapses into $\log x \cdot \sum_{x<p\leqslant 2x} 1 \sim \sum_{x<n\leqslant 2x} \Lambda(n)$.

Intracacies of little-o (not main storyline)

(I do remember Terry discussing this intricacy, and searching a while, found 11/4/22 254A notes discussing *exactly!* the equivalence that I'm discussing here between $\sum_{n\leqslant x}\Lambda(n)=x+o(x)$ and $\sum_{x/2<n\leqslant x}=\frac{x}{2}+o(x)$!)

Sadly, little-o terms are quite a bit trickier to work with than big-O terms: if we know $\sum_{x<n\leqslant 2x}\Lambda(n)=x+E(x)$ with $E(x)=o(x)$, then $\sum_{x/2<n\leqslant x}\Lambda(n)=\frac{x}{2}+E(\frac{x}{2})$. So by telescoping we get $\sum_{x/2^j<n\leqslant 2x}\Lambda(n)=2x-\frac{x}{2^j}+\sum_{i=0}^{j}E(\frac{x}{2^i})$. The error term (which I'll denote E_j) is still $o(x)$, but depending on j!

The idea is that $f=o(x)\iff\forall\epsilon>0, f=O(\epsilon x)$ for $x\geqslant x_\epsilon$ (implicit constant does NOT depend on ϵ!). The telescope above, plus the Chebyshev bound $\sum_{n\leqslant x/2^j}\Lambda(n)=O(C\cdot\frac{x}{2^j})$ gives us that $\sum_{n\leqslant 2x}\Lambda(n)=2x+O(3C\cdot\frac{x}{2^j})$ for all $j\geqslant 0$ (for $x\gg_j 1$ — i.e. x suff. large depending on j — because $E_j(x)=o_j(x)$ is $\leqslant C\cdot\frac{x}{2^j}$ for $x\gg_j 1$), which is exactly $2x+o(x)$ by the previous sentence!

Table of Contents

1 Pseudo-History/Motivation
- The Zeta Function
- Cheap Mertens, von Mangoldt
- Dirichlet Ring, Mobius Inversion, Divisor Sums, Mobius Cancellation
- Dyadic Decomposition, Mertens1
- Mertens2, Prime-Supported Measures
- Selberg Symmetry, Brun-Titchmarsh, Parity Problem

2 Banach Algebra Proof of PNT
- Setup/Outline
- Constructing Banach Algebra Norm
- Non-Trivial Spectrum
- Breaking Parity Barrier

(End of 9/30/22, Terry 254A Lec. 4) We now tackle Merten's 2nd theorem, i.e. the sum $\sum_{y \leqslant n \leqslant x} \frac{1}{p}$. The idea is as follows: $\frac{1}{p} = \frac{\log p}{p} \frac{1}{\log p}$ where the 1st factor is understood by Mertens1, and the 2nd factor is very slowly varying; if it were constant, then we could pull it out of the sum and directly use Mertens1. We can *approximately* do exactly this, by utilizing the FTC:

$$\frac{1}{\log p} = \frac{1}{\log x} + \int_y^x \frac{1_{t \geqslant p}}{t \log^2 t}\, dt.$$

Then

$$\sum_{y \leqslant p \leqslant x} \frac{1}{p} = \frac{1}{\log x} \sum_{y \leqslant p \leqslant x} \frac{\log p}{p} + \sum_{y \leqslant p \leqslant x} \frac{\log p}{p} \left(\int_y^x \frac{1_{t \geqslant p}}{t \log^2 t}\, dt \right),$$

and using Tonelli, the 2nd term becomes

$$\int_y^x \frac{1}{t \log^2 t} \left(\sum_{y \leqslant p \leqslant x} \frac{\log p}{p} 1_{t \geqslant p} \right) dt = \int_y^x \frac{1}{t \log^2 t} \left(\sum_{y \leqslant p \leqslant t} \frac{\log p}{p} \right).$$

Using Mertens1 in both terms, and doing a few lines of algebra/calculus one arrives at:

Theorem 1.2: Mertens2

We have for $2 \leqslant y \leqslant x$,

$$\sum_{y \leqslant p \leqslant x} \frac{1}{p} = \log\log x - \log\log y + O(\tfrac{1}{\log y})$$

and

$$\sum_{y \leqslant n \leqslant x} \frac{\Lambda(n)}{n \log n} = \log\log x - \log\log y + O(\tfrac{1}{\log y}).$$

(recall the 2nd LHS is the Dirichlet series of $\log\zeta$ at $s = 1$, and it was the relationship between $\sum_p \frac{1}{p^s}$ and $\log\zeta(s)$ that we used to get CheapMertens2. The two statements are equivalent by the same ideas as Heuristic 1 but a bit more complicated since here we want the error term to decay; see slide (36).)

It is very important that the error terms here, *unlike in Mertens1*, decay as y gets large.

Side quest on $\frac{\Lambda(n)}{n\log n}$ and $\frac{1}{p}$ (not main storyline)

Like in Heuristic 1, the difference between the 1st and 2nd LHS of Mertens2 is

$$\sum_{j=2}^{\infty}\sum_{y^{1/j}\leqslant p\leqslant x^{1/j}}\frac{\log p}{p^j\log(p^j)}\leqslant\sum_{j=2}^{\infty}\frac{1}{j}\sum_{p\geqslant y^{1/j}}\frac{1}{2^{2j/3}\cdot p^{j/3}}\leqslant\sum_{j=2}^{\infty}\frac{1}{j}\cdot\frac{1}{2^{j/3}}\sum_{n\geqslant y^{1/j}}\frac{1}{n^{2j/3}}.$$

The inner sum on the RHS by MonoQuaInt is

$$\sum_{n\geqslant y^{1/j}}\frac{1}{n^{2j/3}}=\int_{y^{1/j}}^{\infty}\frac{1}{t^{2j/3}}\,dt+O\left(\frac{1}{(y^{1/j})^{2j/3}}\right)=\frac{1}{\frac{2j}{3}-1}(y^{1/j})^{1-\frac{2j}{3}}+O\left(y^{-2/3}\right),$$

and so the whole RHS is

$$\sum_{j=2}^{\infty}\frac{1}{j}\cdot\frac{1}{2^{j/3}}\sum_{n\geqslant y^{1/j}}\frac{1}{n^{2j/3}}\leqslant\left(\sum_{j=2}^{\infty}\frac{1}{j}\cdot\frac{1}{2^{j/3}}\left(\frac{y^{1/j}}{\frac{2j}{3}-1}\right)\right)\cdot O\left(y^{-2/3}\right).$$

Finally $j\geqslant 2\implies y^{1/j}\leqslant y^{1/2}$, and the series is essentially geometric, hence we get $\leqslant O(y^{1/2})\cdot O(y^{-2/3})=O(y^{-1/6})\leqslant O(\frac{1}{\log y})$. (Just came across 10/5/22 254A notes, Terry discuss Merten3, did above calculation even better!)

It's a nice result, and the proof teaches the nice FTC trick, but there's a much better reason I'm presenting it to you. We can rephrase Mertens2 as saying: for any fixed $0 < a < b < \infty$,

$$\sum_{x^a \leqslant p \leqslant x^b} \frac{1}{p} = \log b - \log a + O_a\left(\frac{1}{\log x}\right) = \int_0^\infty 1_{[a,b]}(t) \frac{dt}{t} + O_a\left(\frac{1}{\log x}\right)$$

$$||$$

$$\sum_{x^a \leqslant p \leqslant x^b} \frac{1}{p} = \sum_{p : a \leqslant \frac{\log p}{\log x} \leqslant b} \frac{1}{p} = \sum_p \frac{1}{p} 1_{[a,b]}\left(\frac{\log p}{\log x}\right)$$

Observe that both RHS expressions are linear in the function variable f (currently occupied by $f := 1_{[a,b]}$), and behave well under limits (as sums and integrals are wont to do), and so this generalizes to:

Theorem 1.3: Generalized Mertens2

For $f : (0,\infty) \to \mathbb{C}$ a fixed compactly supported Riemann integrable function,

$$\sum_p \frac{1}{p} f\left(\frac{\log p}{\log x}\right) = \int_0^\infty f(t)\frac{dt}{t} + o_{x\to\infty}(1).$$

Quoted from Terry: "An alternate way to phrase the above ... is that the Radon measures $\sum_p \frac{1}{p}\delta_{\frac{\log p}{\log x}}$ on $(0,+\infty)$ converge in the vague topology to the absolutely continuous measure $\frac{dt}{t}$ in the limit $x\to\infty$,where δ_t denotes the Dirac probability measure at t. Similarly for the ... measures $\sum_n \frac{\Lambda(n)}{n\log n}\delta_{\frac{\log n}{\log x}}$.

To put this another way, [by Mertens1!] the Radon measures $\sum_p \frac{\log p}{p}\delta_{\log p}$ or $\sum_n \frac{\Lambda(n)}{n}\delta_{\log n}$ behave like Lebesgue measure on dyadic intervals such as $[u,(1+\varepsilon)u]$ for fixed $\varepsilon > 0$ and large u. This is weaker than the PNT ... which basically asserts the same statement but on the much smaller intervals $[u,u+\varepsilon]$. (Statements such as the Riemann hypothesis make the same assertion on even finer intervals, such as $[u,u+e^{-(\frac{1}{2}-\varepsilon)u}]$.)"

Table of Contents

1 Pseudo-History/Motivation
- The Zeta Function
- Cheap Mertens, von Mangoldt
- Dirichlet Ring, Mobius Inversion, Divisor Sums, Mobius Cancellation
- Dyadic Decomposition, Mertens1
- Mertens2, Prime-Supported Measures
- Selberg Symmetry, Brun-Titchmarsh, Parity Problem

2 Banach Algebra Proof of PNT
- Setup/Outline
- Constructing Banach Algebra Norm
- Non-Trivial Spectrum
- Breaking Parity Barrier

I did not spend all that time discussing divisor sums $\sum_{d|n}$ just for the sake of the Mertens1; indeed, one can use those ideas to analyze $\Lambda = \mu \star \log$: by the same Fubini/averaging method we get

$$\sum_{n \leqslant x} \Lambda(n) = \sum_{d \leqslant x} \mu(d) \sum_{m \leqslant \frac{x}{d}} \log(m).$$

Using Computation 1.1 (MonoQuaInt on $\sum \log$), we get

$$\sum_{n \leqslant x} \Lambda(n) = x \sum_{d \leqslant x} \frac{\mu(d)}{d} \log\left(\frac{x}{d}\right) - x \sum_{d \leqslant x} \frac{\mu(d)}{d} + \sum_{d \leqslant x} \mu(d) O\left(1 + \log\left(\frac{x}{d}\right)\right).$$

Already the error term has $O(\sum_{d \leqslant x} \mu(d) \cdot 1)$, which we have no bound for besides the trivial one using $|\mu| \leqslant 1$, i.e. $O(x)$. This is already the Chebyshev bound, so the error term in this calculation is simply too big. The $O(\sum_{d \leqslant x} \log \frac{x}{d})$ doesn't make it any worse however. One could of course use MonoQuaInt on $\log(\frac{1}{t})$, but let's do something a little slicker: $\log(t) \leqslant \sqrt{t}$ for $t \geqslant 1$, and $\sum_{d \leqslant x} \frac{\sqrt{x}}{\sqrt{d}}$ and MonoQuaInt gives this is $= \sqrt{x} \cdot (2x^{1/2} + O(1 + \frac{1}{\sqrt{x}})) = O(x)$

The main terms can be analyzed with Prop. 1.3 with $G = \log$ and $G = \mathrm{id}$:

Computation 1.2: Mobius cancellation $G = \log \implies g = \frac{1}{t}$

$$\frac{1}{1/x} = x \sum_{d \leqslant x} \mu(d) \frac{-\log(\frac{d}{x})}{d} + O\left(\sum_{d \leqslant x} \mu(d) (\tfrac{1}{d/x} + 1)) \right).$$

Computation 1.3: Mobius cancellation $G = id \implies g = 1$

$$1 = x \sum_{d \leqslant x} \frac{\mu(d)}{d} (1 - \tfrac{d}{x}) + O\left(\sum_{d \leqslant x} \mu(d)(1 + 1)) \right).$$

So, indeed doing the calculations we recover the Chebyshev bound (Prop. 1.4), and nothing better.

The ingenious insight (of Selberg? or at least most famously popularized by Selberg) is that although it is difficult to shrink the error bound, we can **grow** the main term to **overpower** the error, by replacing all $\log^1$ with $\log^2$.

The resulting LHS $\Lambda_2 := \mu \star \log^2$ still has number theoretic meaning, largely due to the Dirichlet convolution identities involving log from slide (21). Applying them here, we get

$$\begin{aligned} \Lambda_2 := \mu \star \log^2 =: \underbrace{f \star [\log \cdot g]}_{f:=\mu, g:=\log} &= \underbrace{\log \cdot [f \star g]}_{=\log \cdot \Lambda} + \underbrace{-[\log \cdot f] \star g}_{=[\mu \star \Lambda] \star \log = \Lambda \star \Lambda} \\ &= \log \cdot \Lambda + \Lambda \star \Lambda, \end{aligned} \qquad (\Lambda 2)$$

so indeed (for much the same reasons as Heuristic 1), Λ_2 is essentially the "log weighted indicator of primes and semiprimes", or more precisely

$$\Lambda_2(n) \approx 1_{n=p}(n) \cdot (\log^2 p) + 1_{n=p_1 p_2}(n) \cdot (\log p_1 \log p_2).$$

Much of the analysis we did before for $\Lambda = \mu \star \log$ carries over: we just need (using MonoQuaInt)

Computation 1.4: Log Squared Sum

$\sum_{n \leqslant x} \log^2 n = x \log^2 x - 2x \log x + 2x + O(1 + \log^2 x)$ for $x \geqslant 1$.

The error term $O(\sum_{d \leqslant x} \mu(d)(1 + \log^2 d))$ is $O(x)$ by the square-root trick from slide (40), and the main term is given by

Computation 1.5: Mobius cancellation $G = \log^2 \implies g = \frac{2 \log t}{t}$

$$\frac{2\log(1/x)}{1/x} = x \sum_{d \leqslant x} \mu(d) \frac{-\log^2(\frac{d}{x})}{d} + O\Big(\sum_{d \leqslant x} \mu(d) \Big(\frac{\log(d/x)}{d/x} + 1 \Big) \Big)$$

One can do this without MonoQuaInt and get a bit better control in general, but in this case, Prop. 1.3 works.

We thus arrive at

Theorem 1.4: Selberg symmetry formula

$$\sum_{n \leqslant x} \Lambda_2(n) = 2x \log x + O(x).$$

Corollary 1.1: Cheap Brun-Titchmarsh

For any $1 \leqslant y \leqslant x$, we have $\sum_{y \leqslant n \leqslant x} \Lambda(n) \leqslant 2(x-y) + O(\frac{x}{\log x})$.

Proof: Selberg symmetry gives $\sum_{y \leqslant n \leqslant x} \Lambda_2(n) = 2(x-y)\log x + O(x)$ (using $y \log \frac{x}{y} = O(y\frac{x}{y}) = O(x)$). The convolution identity ($\Lambda 2$) for Λ_2 from slide (42) gives the pointwise bound $\Lambda(n)\log n \leqslant \Lambda_2(n)$, and so we get $\sum_{x/2 < n \leqslant x} \Lambda(n)\log(\frac{x}{2}) \leqslant \sum_{x/2 \leqslant n \leqslant x} \Lambda(n)\log n \leqslant 2(\frac{x}{2})\log x + O(x)$. Dividing by $\log(\frac{x}{2})$ and using that $\frac{\log x}{\log(x/2)} = 1 + O(\frac{1}{\log x})$, we get the corollary for $y = \frac{x}{2}$. We are done by dyadic decomposition; see next slide for further details.

Side quest on dyadic decomposition of $O(\frac{x}{\log x})$ (not main storyline)

Continuing proof of Cor. 1.1: it suffices to prove it for $y = 1$ (because one gets to the full generality statement by subtracting $\sum_{n \leqslant x} - \sum_{n \leqslant y}$). Or indeed say $y = 4$ or whatnot. Let j (depending on x) be s.t. $4 \geqslant \frac{x}{2^j} \geqslant 2$. Then $\sum_{x/2^j \leqslant n \leqslant x} \Lambda(n) = 2(x - \frac{x}{2^j}) + O(\sum_{i=0}^{j} \frac{x/2^i}{\log(x/2^i)})$. Then

$$\frac{x/2^i}{\log(x/2^i)} = \frac{x/2^i}{\log(x)} \cdot \frac{\log(x)}{\log(x/2^i)},$$

where the 2nd factor

$$\frac{\log(x)}{\log(x/2^i)} = 1 + \frac{\log(2^i)}{\log(x/2^i)} \leqslant 1 + \frac{i\log(2)}{\log(x/2^j)} \leqslant 1 + \frac{i\log 2}{\log 2} = (1+i),$$

and of course $\frac{1+i}{2^i} \ll \frac{1}{(1.5)^i}$, so $O(\sum_{i=0}^{j} \frac{x/2^i}{\log(x/2^i)}) \leqslant O(\frac{x}{\log x} \sum_i \frac{1}{(1.5)^i})$. ■

One sees why it is more preferably to work with asymptotics like $\sum_{n \leqslant x} \Lambda(n) \sim x$ than $\pi(x) \sim \frac{x}{\log x}$; the former is far easier to dyadically decompose, for one! See MSE for *precise* dyadic decomp. of the latter.

Side quest on parity problem (not main storyline)

Quoted from Terry: "Sieve theory methods can provide good upper bounds, lower bounds, and even asymptotics for almost primes, which lead to upper bounds for primes which tend to be off by a constant factor such as 2. Rather frustratingly, though, sieve methods have proven largely unable to count or even lower bound the primes themselves... The reason for this — the *parity problem* — was first clarified by Selberg. Roughly speaking, it asserts:

Heuristic 3: Parity problem

Parity problem. If A is a set whose elements are all products of an odd number of primes (or are all products of an even number of primes), then (without injecting additional ingredients), sieve theory is unable to provide non-trivial lower bounds on the size of A. Also, any upper bounds must be off from the truth by a factor of 2 or more.

Thus we can hope to count P_2 almost primes (because they can have either an odd or an even number of factors), ... but we cannot hope to use plain sieve theory to just count primes ..."

Side quest on sieve theory (not main storyline)

I must say: I can not recommend this blogpost of Terry Tao enough. Other excerpts: "Suffice to say that a set of integers in $[N, 2N]$ is "smooth" if membership in that set can be largely determined by its most significant digits in the Euclidean sense, and also in the p-adic senses for small p; roughly speaking, this means that this set is approximately the pullback of some "low complexity" set in the adele ring — a set which can be efficiently fashioned out of a few of basic sets which generate the topology and σ-algebra of that ring."

... "a divisor sum (which is a number-theoretic analogue of a smooth function)"

Side quest on parity problem (not main storyline)

More: "The Liouville function oscillates quite randomly between $+1$ and -1. Indeed, ... if the Riemann hypothesis is true then we have ... $\sum_{n \leqslant N} \lambda(n) = O_\epsilon(N^{1/2+\epsilon})$ for all $\epsilon > 0$. Assuming the generalized Riemann hypothesis, we have a similar claim for residue classes: $\sum_{n \leqslant N} 1_{n=a(\mathrm{mod} q)} \lambda(n) = O_\epsilon(N^{1/2+\epsilon})$ for all $\epsilon > 0$.

What this basically means is that the Liouville function is essentially orthogonal to all smooth sets, or all smooth functions. Since sieve theory attempts to estimate everything in terms of smooth sets and functions, it thus cannot eliminate an inherent ambiguity coming from the Liouville function. More concretely, let A be a set where λ is constant (e.g. λ is identically -1, which would be the case if A consisted of primes) and suppose we attempt to establish a lower bound for the size of a set A in, say, $[N, 2N]$ by setting up a divisor sum lower bound (6) : $1_A(n) \geqslant \sum_d c_d 1_{d|n}(n)$"
(cont. on next slide)

Side quest on parity problem (not main storyline)

(cont.) "... If we sum in n we obtain a lower bound of the form (7) : $|A| \geqslant \sum_d c_d \frac{N}{d} + \dots$ and we can hope that the main term $\sum_d c_d \frac{N}{d}$ will be strictly positive and the error term is of lesser order, thus giving a non-trivial lower bound on $|A|$.

Unfortunately, if we multiply both sides of (6) by the non-negative weight $1 + \lambda(n)$ and sum in n, we obtain $0 \geqslant \sum_d c_d 1_{d|n}(n)(1 + \lambda(n))$ since we are assuming λ to equal -1 on A.

If we sum this in n, and use the fact that λ is essentially orthogonal to divisor sums, we obtain $0 \geqslant \sum_d c_d \frac{N}{d} + \dots$ which basically means that the bound (7) cannot improve upon the trivial bound $|A| \geqslant 0$. A similar argument using the weight $1 - \lambda(n)$ also shows that any upper bound on $|A|$ obtained via sieve theory has to essentially be at least as large as $2|A|$."

More: “The P_2 almost prime number theorem establishes the prime number theorem “up to a factor of 2”. It is surprisingly difficult to improve upon this factor of 2 by elementary methods, though once one can replace 2 by $2 - \epsilon$ for some $\epsilon > 0$ (a fact which is roughly equivalent to the absence of zeroes of $\zeta(s)$ on the line $\mathfrak{Re}(s) = 1$), one can iterate the Selberg symmetry formula (together with the tautological fact that an P_2 almost prime is either a prime or the product of two primes) to get the prime number theorem; this is essentially the Erdős-Selberg elementary proof of that theorem.”

Side quest on $\mathbb{C}$-analytic POV on parity problem (not main storyline)

Quoting Terry, regarding the Selberg symmetry formula (in comparison with PNT):

Heuristic 4: Complex analytic perspective on why Λ_2 is easier

"This fact is much easier to prove than the prime number theorem itself. In terms of zeta functions, the reason why the prime number theorem is difficult is that the simple pole of $\frac{\zeta'(s)}{\zeta(s)}$ at $s = 1$ could conceivably be counteracted by other simple poles on the line $\mathfrak{Re}(s) = 1$. On the other hand, the P_2 almost prime number theorem is much easier because the effect of the double pole of $\frac{\zeta''(s)}{\zeta(s)} ats = 1$ is not counteracted by the other poles on the line $\mathfrak{Re}(s) = 1$, which are at most simple."

Following mainly https://terrytao.wordpress.com/2014/10/25/a-banach-algebra-proof-of-the-prime-number-theorem/ (seems to be the original appearance of the proof anywhere public...), but there are some rather inexplicable conventions taken (namely his $\int_{\mathbb{R}} G(t)\,d\tau_h\mu(t) = \int_{\mathbb{R}} G(t+h)\,d\mu(t)$ and his $\log(\frac{x}{n})$). Thankfully there's also a published version in Manfred Einsiedler and Thomas Ward's book *Functional Analysis, Spectral Theory, and Applications*, that I found out through Redmond McNamara's paper on a "A Dynamical Proof of the Prime Number Theorem", and they do it with the notation/convention that I expect.

Table of Contents

1 Pseudo-History/Motivation
- The Zeta Function
- Cheap Mertens, von Mangoldt
- Dirichlet Ring, Mobius Inversion, Divisor Sums, Mobius Cancellation
- Dyadic Decomposition, Mertens1
- Mertens2, Prime-Supported Measures
- Selberg Symmetry, Brun-Titchmarsh, Parity Problem

2 Banach Algebra Proof of PNT
- Setup/Outline
- Constructing Banach Algebra Norm
- Non-Trivial Spectrum
- Breaking Parity Barrier

Definition 2.1: Mertens1 measure

We denote

$$\mu := \sum_{n=1}^{\infty} \frac{\Lambda(n)}{n} \delta_{\log n}.$$

Picture this as a bunch of sticks supported on essentially $\{\log p : p \text{ prime}\} \subseteq (0, \infty)$, with the sticks getting shorter and shorter as you go right, but also denser and denser. Measuring $[u, u + \varepsilon]$ against this for $u \to \infty$ can be thought of as having a fixed interval of width ε and shifting μ (i.e. the sticks) leftward.

Definition 2.2: Left shift

For any $h \in \mathbb{R}$, let $\tau_h \mu$ denote the left translate of μ by h units. That is,

$$\int_{\mathbb{R}} G(t)\, d\tau_h \mu(t) = \int_{\mathbb{R}} G(t - h)\, d\mu(t) \quad \forall G \in C_c(\mathbb{R}).$$

Proposition 2.1: PNT in terms of μ

PNT holds $\iff$ the translates $\tau_h\mu$ converge in the vague topology to Lebesgue measure m as $h \to \infty$ (notated $\tau_h\mu \rightharpoonup m$).

Proof: [see right after Thm. 3 box in Tao, and Prop. 14.4 in E-W]. Observe that $\tau_h\mu \rightharpoonup m \iff [\sum_n \frac{\Lambda(n)}{n} G(\log n - \log x) = \int_{\mathbb{R}} G(t)\,dt + o_{G,x\to\infty}(1)$ for any $G \in C_c(\mathbb{R})]$ (the decay rate $o_G(1)$ may then depend on G). Specializing to functions of the form $G(t) := e^t\eta(e^t)$ for some $\eta \in C_c((0,\infty))$, we get $\sum_n \Lambda(n)\eta\left(\frac{n}{x}\right) = x\int_{\mathbb{R}} \eta(u)\,du + O_{x \geqslant x_{\epsilon,\eta}}(\epsilon \cdot x)$. (Recall (32) that $o_{x\to\infty}(x) \iff \forall \epsilon > 0, O_{x\geqslant x_\epsilon}(\epsilon \cdot x)$.)

Choose (smooth Urysohn) $0 \leqslant \eta_- \leqslant 1_{[\frac{1}{2},1]} \leqslant \eta_+$ with $\int_{\mathbb{R}}(\eta_+ - \eta_-) < \delta$, so

$$\left(\frac{1}{2} - \delta\right) \cdot x + O_{x\geqslant x_{\epsilon,\delta}}(\epsilon \cdot x) \leqslant \sum_{\frac{x}{2} < n \leqslant x} \Lambda(n) \leqslant \left(\frac{1}{2} + \delta\right) \cdot x + O_{x\geqslant x_{\epsilon,\delta}}(\epsilon \cdot x),$$

i.e. exactly $\sum_{x/2<n\leqslant x} \Lambda(n) = \frac{x}{2} + o_{x\to\infty}(x)$. Conclude by using slide (32). ■

So, we define (exactly same expressions as appear in above proof)

Definition 2.3: Seminorm

For $G \in C_c(\mathbb{R})$, set $\|G\|_\Lambda := \limsup_{x\to\infty} \left| \sum_n \frac{\Lambda(n)}{n} G\Big(\log \frac{n}{x}\Big) - \int_{\mathbb{R}} G(t)\, dt \right|$.

We very much hope (Prop. 2.1) that $\|G\| \equiv 0$. We prove this in 3 steps.

Theorem 2.1: Construction of Banach algebra norm

First, $\|\bullet\|_\Lambda$ is a seminorm on $C_c(\mathbb{R})$, satisfying the bound

$$\|G\|_\Lambda \leqslant \|G\|_{L^1(\mathbb{R})} := \int_{\mathbb{R}} G(t)\, dt \quad \forall G \in C_c(\mathbb{R}), \tag{L1}$$

and furthermore the Banach algebra bound

$$\|G * H\|_\Lambda \leqslant \|G\|_\Lambda \cdot \|H\|_\Lambda \quad \forall G, H \in C_c(\mathbb{R}). \tag{*}$$

Ingredient 2: " an application of the spectral radius formula and some basic Fourier analysis (in particular, the observation that $C_c(\mathbb{R})$ contains a plentiful supply of local units):"

Theorem 2.2: BanAlg $\neq 0$ with many local units has spectrum $\neq 0$

Let $\| \bullet \|$ be any seminorm on $C_c(\mathbb{R})$ obeying (L1), (*), and suppose that it is not identically 0. Then there exists $\xi \in \mathbb{R}$ s.t. the linear functional $[f \mapsto \hat{f}(\xi) := \int_{\mathbb{R}} f(t)e^{-it\xi}\,dt] : C_c(\mathbb{R}) \to \mathbb{C}$ is continuous w.r.t. $\| \bullet \|$, i.e.

$$\left| \int_{\mathbb{R}} G(t)e^{-it\xi}\,dt \right| \leqslant \|G\| \quad \forall G \in C_c(\mathbb{R}).$$

In particular, by (L1), one has $\|G\| = \|G\|_{L^1(\mathbb{R})}$ whenever $G(t)e^{it\xi}$ takes only values in $[0, \infty)$.

Ingredient 3: "a consequence of the Selberg symmetry formula and the fact that Λ is real (as well as Mertens' theorem, in the $\xi = 0$ case), ... closely related to the non-vanishing of the Riemann zeta function ζ on the line $\{1 + i\xi : \xi \in \mathbb{R}\}$:"

Theorem 2.3: Breaking the parity barrier

Let $\xi \in \mathbb{R}$. Then there exists $G \in C_c(\mathbb{R})$ s.t. $G(t)e^{-it\xi}$ is non-negative (takes values only in $[0, \infty)$), and $\|G\|_\Lambda < \|G\|_{L^1(\mathbb{R})}$.

It is clear that **Ingredient 2** and **Ingredient 3** force $\| \bullet \|_\Lambda \equiv 0$, as desired.

Table of Contents

1 Pseudo-History/Motivation
- The Zeta Function
- Cheap Mertens, von Mangoldt
- Dirichlet Ring, Mobius Inversion, Divisor Sums, Mobius Cancellation
- Dyadic Decomposition, Mertens1
- Mertens2, Prime-Supported Measures
- Selberg Symmetry, Brun-Titchmarsh, Parity Problem

2 Banach Algebra Proof of PNT
- Setup/Outline
- Constructing Banach Algebra Norm
- Non-Trivial Spectrum
- Breaking Parity Barrier

Proof sketch of Thm. 2.1 (L1): the idea is that SelbergSym is equivalent to a statement about vague convergence of measures $\tau_h \nu \rightharpoonup 2m$ (in much the same way that we saw Mertens2 was equivalent to GeneralizedMertens2), and the identity (Λ2): $\Lambda_2(n) = \log \cdot \Lambda + \Lambda \star \Lambda$ allows us to bound the shifted Λ-measures $\tau_h \mu$ in terms of the shifted SelbergSym Λ_2-measures $\tau_h \nu$, giving us a uniform (in h) bound on how much mass (in the eys of $\tau_h \mu$) can escape to infinity, at which point Prokhorov's theorem tells us that there's some limit point λ of the shifted measures $\tau_h \mu$ in the vague topology. This will give us (L1). Let's flesh this out.

Definition 2.4: SelbergSym Λ_2 measure

We denote

$$\nu := \sum_{n=1}^{\infty} \frac{\Lambda_2(n)}{n \log n} \delta_{\log n}.$$

Like I said above, SelbergSym $\iff \tau_h \nu \rightharpoonup 2m$.

Lemma 2.1: Convolutions of measures

For $\upsilon_f := \sum_n \frac{f(n)}{n}\delta_{\log n}$ with $f : \mathbb{N}^+ \to \mathbb{C}$, we have $\upsilon_{f_1} * \upsilon_{f_2} = \upsilon_{[f_1 \star f_2]}$, where $\star$ is Dirichlet convolution, and the convolution of 2 Radon measures υ_1, υ_2 on $[0, \infty)$ is given by

$$[\upsilon_1 * \upsilon_2](B) := \iint 1_B(t_1 + t_2)\, d\upsilon_1(t_1)\, d\upsilon_2(t_2)$$

for all Borel sets $B \in \mathscr{B}([0, \infty))$ (and is itself also a Radon measure).

Proof: [copied from Lemma 14.10 in E-W] The key calculation is

$$\begin{aligned}[\upsilon_{f_1} * \upsilon_{f_2}](B) &= \sum_{m_1=1}^{\infty} \sum_{m_2=1}^{\infty} 1_B(\log m_1 + \log m_2) \frac{f_1(m_1)}{m_1} \frac{f_2(m_2)}{m_2} \\ &= \sum_{n=1}^{\infty} 1_B(\log n) \frac{1}{n} \sum_{d|n} f_1(d) f_2\left(\frac{n}{d}\right) \\ &= \int 1_B \; \mathrm{d}\upsilon_{f_1 {}_D f_2} = \upsilon_{f_1 \star f_2}(B).\end{aligned}$$

Corollary 2.1: Measure reformulation of SelbergSym

$$d\nu(t) = d\mu(t) + \frac{1}{t} \cdot d[\mu \star \mu](t)$$

Proof: [copied from Corollary 14.13 in E-W] the previous Lemma 2.1 gives us $\mu * \mu = \upsilon_{\Lambda\star\Lambda}$, and so the identity (Λ2): $\Lambda_2(n) = \log \cdot \Lambda + \Lambda \star \Lambda$ yields

$$\begin{aligned}\nu(B) &:= \upsilon_{\frac{\Lambda_2}{\log}}(B) := \sum_{n\geqslant 1} 1_B(\log n)\frac{\Lambda_2(n)}{n\log n} \\ &= \sum_{n\geqslant 1} 1_B(\log n)\left(\frac{\Lambda(n)}{n} + \frac{[\Lambda\star\Lambda](n)}{n\log n}\right) \\ &= \mu(B) + \int 1_B(t)\frac{1}{t}\, d[\upsilon_{\Lambda\star\Lambda}](t),\end{aligned}$$

as desired. ■

Note that the identity (Λ2) also gives the bound $0 \leqslant \tau_h\mu \leqslant \tau_h\nu$.

Theorem 2.4: Prokhorov's theorem

Let $\mathscr{F}$ be a family of finite measures on $\mathbb{R}$ (can extend to general separable metric spaces, but can't use distribution functions as below, so have to construct the measure by hand). If the family $\mathscr{F}$ is *tight* (uniform control of mass escape to infinity: $\forall \epsilon > 0$, $\exists$ compact K s.t. $\forall \upsilon \in \mathscr{F}$, $\upsilon(\mathbb{R} \setminus K) < \epsilon$), then it is relatively compact, i.e. there is some subsequence that converges to some limit point λ.

Proof sketch: look at the cumulative distribution functions $F_n(x) := \upsilon((-\infty, x]) : \mathbb{R} \to [0, b]$ of these measures $\upsilon \in \mathscr{F}$. The Helly selection theorem (an Arzelá-Ascoli/diagonal subsequence argument to cook up a pointwise convergent subsequence of a sequence of monotone functions) allows us to get $F_n(x) \to F(x)$ for some monotone function $F : \mathbb{R} \to [0, b]$. The Lebesgue-Radon-Nikodym converts this c.d.f. into a measure Υ, and tightness verifies that the subsequence of $\upsilon \in \mathscr{F}$ vague converges to Υ. ■

Now fix $G \in C_c(\mathbb{R})$ and a subsequence $h_n \to \infty$ s.t.

$$\|G\|_\Lambda = \lim_{n\to\infty} \left| \sum_n \frac{\Lambda(n)}{n} G(\log n - h_n) - \int_{\mathbb{R}} G(t)\,dt \right|$$

Because $0 \leqslant \tau_h \mu \leqslant \tau_h \nu$ and $\tau_h \nu \rightharpoonup 2m$, we do get a uniform in h bound (given by the upper bound $\ll 2m$) on how much mass can escape to infinity of the family of measures $\{\tau_h \mu\}_h$, so Prokhorov gives us a limit point $0 \leqslant \lambda \leqslant 2m$ (perhaps depending on G!) s.t. $\tau_{h_n}\mu \rightharpoonup \lambda$ and so we get

$$\begin{aligned} \|G\|_\Lambda &= \left| \int_{\mathbb{R}} G(t)\,d\lambda(t) - \int_{\mathbb{R}} G(t)\,dm(t) \right| \\ &= \left| \int_{\mathbb{R}} G(t)\,d[\lambda - m](t) \right| \leqslant \int_{\mathbb{R}} G(t)\,d|\lambda - m|(t). \end{aligned}$$

But $0 \leqslant \lambda \leqslant 2m \iff -m \leqslant \lambda - m \leqslant m \iff |\lambda - m| \leqslant m$! (Hahn-Jordan decomposition of signed measures blah blah blah.) Thus, we've proven (L1).

Proof sketch of Thm. 2.1 (*): so we want to prove that for any $G, H \in C_c(\mathbb{R})$,

$$\left| \int_{\mathbb{R}} [G * H](t)\, d\tau_h\mu(t) - \int_{\mathbb{R}} [G * H](t)\, dm(t) \right| \leqslant \|G\|_\Lambda \|H\|_\Lambda + o_{G,H}(1),$$

where the decay $o_{G,H}(1)$ can depend on G, H. Since we already know $\tau_h\nu \rightharpoonup 2m$, the decomposition $\nu = \mu + \frac{1}{t} \cdot [\mu * \mu]$ from Cor. 2.1 gives us (after rearranging a little bit)

$$\int_{\mathbb{R}} G * H\, d\tau_h\mu - \int_{\mathbb{R}} G * H\, dm = \int_{\mathbb{R}} G * H\, dm - \int_{\mathbb{R}} [G * H](t) \cdot \frac{1}{t}\, d[\mu * \mu](t),$$

so we come to the equivalent formulation

$$\left| \int_{\mathbb{R}} [G * H](t) \cdot \frac{1}{t}\, d[\mu * \mu](t) - \int_{\mathbb{R}} G * H\, dm \right| \leqslant \|G\|_\Lambda \|H\|_\Lambda + o_{G,H}(1).$$

We now denote $m_+ := m|_{[0,\infty)}$ (restrict Lebesgue measure m to $[0,\infty)$). Then $d[m_+ * m_+] = t\,dm_+$ because (quick change of variables) equals

$$\int_{\mathbb{R}} 1_B(t)\,d[m_+ * m_+](t) := \int_0^\infty \int_0^\infty 1_B(t_1+t_2)\,dt_1\,dt_2$$
$$= \int_0^\infty 1_B(u) \int_0^u dt_1\,du = \int_{\mathbb{R}} 1_B(t)\cdot t\,dm_+(t).$$

So $\tau_h(\frac{1}{t}[m_+ * m_+]) = \tau_h m_+$ vague converges $\rightharpoonup m$. We also consider the measure $\frac{1}{t}\,d[m_+ * \mu]$: the calculation [copied from E-W ca. Eq. (14.20)]

$$\int_{\mathbb{R}} 1_B(t)\,d[m_+ * \mu] = \sum_{n\geqslant 1} \int_0^\infty 1_B(t+\log n)\frac{\Lambda(n)}{n}\,dt$$
$$= \sum_{n\geqslant 1} \int_0^\infty 1_B(s) 1_{[\log n,\infty)}(s)\frac{\Lambda(n)}{n}\,ds = \int_0^\infty 1_B(s)\cdot \sum_{n\leqslant e^s} \frac{\Lambda(n)}{n}\,ds$$

tells us that $\int_{\mathbb{R}} 1_B(t)\cdot\frac{1}{t}\,d[m_+ * \mu] = \int_{\mathbb{R}} 1_B(s)\cdot\frac{1}{t}\sum_{n\leqslant e^t}\frac{\Lambda(n)}{n}\,dt$, but Mertens1 tells us the measure on the RHS is $(1+o_{t\to\infty}(1))\,dm(t)$, so we get $\tau_h(\frac{1}{t}\,d[m_+ * \mu]) \rightharpoonup m$.

The previous slide allows us to conclude that

$$\tau_h\left(\frac{1}{t}\, d[\mu * \mu]\right) - \tau_h\left(\frac{1}{t}\, d[(\mu - m_+) * (\mu - m_+)]\right) \rightharpoonup m.$$

So we have a new equivalent statement of Thm. 2.1 (*): we want that

$$\left|\int_{\mathbb{R}} [G * H](t) \cdot \frac{1}{t}\, d\tau_h[(\mu - m_+) * (\mu - m_+)](t)\right| \leqslant \|G\|_\Lambda \|H\|_\Lambda + o_{G,H}(1)$$

The LHS is (Terry has typo. Follow E-W.)

$$\text{LHS} := \left|\int_{\mathbb{R}} \frac{1}{t}[G * H](t-h)\, d[(\mu - m_+) * (\mu - m_+)](t)\right|.$$

$[G * H]$ has compact support, so we may always treat $t - h = O(1)$ so

$$\frac{1}{t} = \frac{1}{h} \cdot (1 + o_{h\to\infty}(1)) = \frac{1}{h} + \frac{o_{h\to\infty}(1)}{t}$$

(i.e. on the set S_h of t s.t. $t - h \in \operatorname{supp}[G * H]$, the previous asymptotics hold).

The error term $E(h)$ is then

$$\begin{aligned} E(h) &:= \left| \int_{\mathbb{R}} o_{h\to\infty}(1) \frac{1}{t} f(t-h)\, d[(\mu - m_+) * (\mu - m_+)](t) \right| \\ &\leqslant o_{h\to\infty}(1) \int_{\mathbb{R}} f(t-h) \cdot \frac{1}{t}\, d[(\mu + m_+) * (\mu - m_+)] \end{aligned}$$

(pull absolute values in on the measures, and $|(\mu - m_+) * (\mu - m_+)| \leqslant [(\mu + m_+) * (\mu - m_+)]$). And we can bound

$$\begin{aligned} \frac{1}{t}\, d[(\mu + m_+) * (\mu - m_+)] &\leqslant \frac{1}{t}\, d[(\mu + m_+) * (\mu - m_+)] + \mu \\ &= \nu + \frac{2}{t}\, d[\mu + m_+] + m_+, \end{aligned}$$

which vague converges $\rightharpoonup 5m$ when we apply τ_h to both sides and send $h \to \infty$. Thus, $E(h) \leqslant o_{h\to\infty}(1) \cdot O_f(1)$ (where $O_f(1)$ is essentially $\|f\|_{L^1(\mathbb{R})}$).

The remainder of the proof is essentially just a long manipulation with Fubini's theorem ("Fubini autopilot" analogous to "algebra autopilot" that high-schoolers may be accustomed to). The LHS ends up equalling

$$\left|\frac{1}{h}\int_{\mathbb{R}} \underbrace{\left(\int_{\mathbb{R}} G(t)d\tau_r[\mu - m_+](t)\right)}_{=:I_1(r)} \underbrace{\left(\int_{\mathbb{R}} H(t)d\tau_{h-r}[\mu - m_+](t)\right)}_{=:I_2(h-r)} dr\right| + o_{h\to\infty}(1).$$

Depending on the support of G, H (compact in $\mathbb{R}$), there exists $R \in \mathbb{R}$ s.t. $r < R \implies I_1(r) = 0$ and $h - r < R \implies I_2(r) = 0$ (i.e. μ and m_+ are supported on the half-line $[0, \infty)$, and will not hit the support of G, H if the measures are too far on the right still).

On the other hand, by (limsup) definition of $\|\bullet\|_\Lambda$, there is some $S \in \mathbb{R}$ s.t. $r > S \implies I_1(r) \leqslant \|G\|_\Lambda + \varepsilon$ and $h - r > S \implies I_2(h - r) \leqslant \|H\|_\Lambda + \varepsilon$ (i.e. measures have moved far enough left that they see all the mass of G, H).

So ALTOGETHER! we get [copied from end of §14.2.6 in E-W]

$$\begin{aligned} \text{LHS} &\leqslant \frac{1}{h}\left|\int_R^{h-R} I_1(r) I_2(h-r)\,dr\right| + o_{G,H;h\to\infty}(1) \\ &\leqslant \frac{1}{h} O_{G,H}(1) + \frac{1}{h}\left|\int_S^{h-S} I_1(r) I_2(h-r)\,dr\right| + o_{G,H;h\to\infty}(1) \\ &\leqslant \frac{h-2S}{h}\left(\|G\|_\Lambda + \varepsilon\right)\left(\|G\|_\Lambda + \varepsilon\right) + o_{G,H;h\to\infty}(1). \end{aligned}$$

And sending $\varepsilon \searrow 0$, we WIN!

Table of Contents

1 Pseudo-History/Motivation
- The Zeta Function
- Cheap Mertens, von Mangoldt
- Dirichlet Ring, Mobius Inversion, Divisor Sums, Mobius Cancellation
- Dyadic Decomposition, Mertens1
- Mertens2, Prime-Supported Measures
- Selberg Symmetry, Brun-Titchmarsh, Parity Problem

2 Banach Algebra Proof of PNT
- Setup/Outline
- Constructing Banach Algebra Norm
- Non-Trivial Spectrum
- Breaking Parity Barrier

Restate:

Ingredient 2: " an application of the spectral radius formula and some basic Fourier analysis (in particular, the observation that $C_c(\mathbb{R})$ contains a plentiful supply of local units):"

Theorem 2.5: BanAlg $\neq 0$ with many local units has spectrum $\neq 0$

Let $\|\bullet\|$ be any seminorm on $C_c(\mathbb{R})$ obeying (L1), (*), and suppose that it is not identically 0. Then there exists $\xi \in \mathbb{R}$ s.t. the linear functional $[f \mapsto \hat{f}(\xi) := \int_{\mathbb{R}} f(t)e^{-it\xi}\,dt] : C_c(\mathbb{R}) \to \mathbb{C}$ is continuous w.r.t. $\|\bullet\|$, i.e.

$$\left|\int_{\mathbb{R}} G(t)e^{-it\xi}\,dt\right| \leqslant \|G\| \quad \forall G \in C_c(\mathbb{R}).$$

In particular, by (L1), one has $\|G\| = \|G\|_{L^1(\mathbb{R})}$ whenever $G(t)e^{it\xi}$ takes only values in $[0, \infty)$.

Proof sketch of Thm. 2.5: (Purely functional analysis/no number theory whatsoever.) [everything in this section copied from Tao §3, with minor edits]

Definition 2.5: The Banach algebra

Denote by B be the Banach algebra completion of $C_c(\mathbb{R})$ under the seminorm $\| \bullet \|$ (thus B is the space of Cauchy sequences in $C_c(\mathbb{R})$, quotiented out by the sequences that go to zero in the seminorm $\| \bullet \|$, [by (L1), we can think of these concretely as $L^1(\mathbb{R})$ functions, i.e. $\Phi : B \hookrightarrow L^1(\mathbb{R})$]). Since $\| \bullet \|$ is not identically zero, B is a non-trivial commutative Banach algebra (but it is not necessarily unital).

"It is convenient to adjoin a unit 1 to B to create a unital commutative Banach algebra $B' := \mathbb{C}1 + B$ with the extended norm

$$\|t1 + f\| := |t| + |f|$$

for $t \in \mathbb{C}$ and $f \in B$; one easily verifies that B' is a unital commutative Banach algebra." (I.e. just "add a Dirac delta δ_0" to $B \subseteq L^1(\mathbb{R})$?)

The KEY SOFT ANALYSIS FACT driving the whole approach.

Theorem 2.6: Spectral radius

For a unital commutative Banach algebra $(B, +, \cdot, \|\bullet\|)$, and any $f \in B$, $(\mathrm{Hom}(B, \mathbb{C})$ being continuous unital Banach algebra homomorphisms)

$$\lim_{n\to\infty} \|f^n\|^{1/n} = \sup_{\varrho \in \mathrm{Hom}(B,\mathbb{C})} |\varrho(f)|.$$

This quantity (really originally only the RHS) is called the *spectral radius*.

Proof: 2 paragraphs in §6 of Tao. ■

"Suppose [f.s.o.c.] that all elements of $f \in B$ [again, the space of Cauchy sequences in $C_c(\mathbb{R})$, quotiented out by the sequences that go to zero in the seminorm $\| \bullet \|$; by (L1), we can think of these concretely as $L^1(\mathbb{R})$ functions] have zero spectral radius, i.e. $\|f^n\|^{1/n} \to 0$.

Let $f \in L^1(\mathbb{R})$ be a Schwartz function with compactly supported Fourier transform. Then we can find another Schwartz function g with compactly supported Fourier transform such that $f * g = f$ (by ensuring that $\hat{g} = 1$ on the support of $\hat{f}$ which is compact; thus g is a "local unit" on the Fourier support of f). Thus $f * g^{*n} = f$ for all n.

But g has spectral radius zero, thus f is zero in B [if not, can divide the inequality $\|f\| \leqslant \|f\| \cdot \|g^{*n}\|$ by $\|f\|$, take nth root of both sides, and take $n \to \infty$ and get $1 \leqslant 0$; contradiction]. By density of Schwartz functions (with Fourier compact support) in $L^1(\mathbb{R})$, this implies that B is trivial, a contradiction."

“Thus there is an element of B with positive spectral radius. Then by the SpectralRadiusFormula, there is a [continuous unital Banach algebra hom.] $\varrho : B' \to \mathbb{C}$ that is does not vanish identically on B.

Suppose [f.s.o.c] that for each $\xi \in \mathbb{R}$, there exists $f [= f_\xi] \in C_c(\mathbb{R})$ in the $\ker(\varrho) \subseteq B$ whose [single] Fourier coefficient $\hat{f}(\xi) := \int_{\mathbb{R}} f(t) e^{-it\xi}\, dt$ is non-vanishing.”

Since $\ker(\rho) \subseteq B$ is a space [a vector space!] closed with respect to convolutions by $C_c(\mathbb{R})$ functions, some Fourier analysis and a smooth partition of unity then shows that the kernel of ϱ contains any Schwartz function with compactly supported Fourier transform, and thus by density ϱ is trivial; a contradiction.”

Theorem 2.7: Spectrum of $L^1(G)$ of LCA group G is $\widehat{G}$

[Proposition 11.38 in E-W] For a LCA (locally compact abelian) group G, the Gelfand dual/spectrum $\sigma(L^1(G)) := \mathrm{Hom}(L^1(G), \mathbb{C})$ of continuous Banach algebra homomorphisms, is a locally compact σ-compact metrizable space which can be identified with the Pontryagin dual $\widehat{G}$.

"Thus there must exist $\xi \in \mathbb{R}$ such that $\ker(\varrho)$ contains all test functions with Fourier coefficient vanishing at ξ. From this we conclude that ϱ on B is a constant multiple of the Fourier coefficient map $f \mapsto \hat{f}(\xi)$; being a non-trivial algebra homomorphism on B, we thus have $\varrho(f) = \hat{f}(\xi)$ for all $f \in C_c(\mathbb{R})$. Since characters have norm at most 1 (as can be seen for instance from SpectralRadiusFormula), we obtain the claim."

Table of Contents

1 Pseudo-History/Motivation
- The Zeta Function
- Cheap Mertens, von Mangoldt
- Dirichlet Ring, Mobius Inversion, Divisor Sums, Mobius Cancellation
- Dyadic Decomposition, Mertens1
- Mertens2, Prime-Supported Measures
- Selberg Symmetry, Brun-Titchmarsh, Parity Problem

2 Banach Algebra Proof of PNT
- Setup/Outline
- Constructing Banach Algebra Norm
- Non-Trivial Spectrum
- Breaking Parity Barrier

Restate:

Theorem: Breaking the parity barrier

Let $\xi \in \mathbb{R}$. Then there exists $G \in C_c(\mathbb{R})$ s.t. $G(t)e^{-it\xi}$ is non-negative (takes values only in $[0, \infty)$), and $\|G\|_\Lambda < \|G\|_{L^1(\mathbb{R})}$.

Proof sketch of Thm. 2.3: (The easiest of the 3 ingredients.) [everything in this section copied fromTao §4, with minor edits].

Case 1: $\xi = 0$. "Let $G = G_0 : \mathbb{R} \to [0, 1]$ be a continuous function that equals 1 on $[0, N]$ and is supported on $[-1, N+1]$ for some large N. From Mertens' theorem we have

$$\sum_n \frac{\Lambda(n)}{n} G_0(\log \frac{n}{x}) = N + O(1)$$

for x sufficiently large depending on N, and thus

$$\|G_0\|_\Lambda = O(1).$$

The claim then follows by taking N sufficiently large."

Case 2: $\xi \neq 0$. "In the language of [the (measure-theory heavy) section constructing the Banach algebra norm], we have

$$\|G\| = \left| \int_{\mathbb{R}} G(t)\, d[\lambda - m](t) \right|$$

for some limit point λ of the $\tau_h \mu$. We can write the right-hand side as

$$\mathfrak{Re}\left(e^{i\theta} \int_{\mathbb{R}} G(t)\, d[\lambda - m](t) \right)$$

for some phase $e^{i\theta}$ [θ constant]. From $0 \leqslant \lambda \leqslant 2m$, $\lambda - m$ is a real measure between $-m$ and m, so by [the fact that integration against a real measure and $\mathfrak{Re}$ can be interchanged, linearity of integral, and] the triangle inequality we have

$$\|G\| \leqslant \int_{\mathbb{R}} \left| \mathfrak{Re}(e^{i\theta} G(t)) \right| dt.$$

(cont. Terry) Now we set $G(t) := G_0(t)e^{it\xi}$, where G_0 is as before. Then

$$\int_{\mathbb{R}} \left| \mathfrak{Re}(e^{i\theta} G(t)) \right| dt = \int_{\mathbb{R}} |\cos(t\xi + \theta)| G_0(t)\, dt.$$

Since $t \mapsto |\cos(t\xi + \theta)|$ is periodic with period $2\pi/|\xi|$ and has mean value strictly less than 1 (in fact, it has mean $\frac{2}{\pi}$), we thus have

$$\int_{\mathbb{R}} \left| \mathfrak{Re}(e^{i\theta} G(t)) \right| dt = \int_{\mathbb{R}} |\cos(t\xi + \theta)| G_0(t)\, dt < \int_{\mathbb{R}} G_0(t)\, dt$$

if N is sufficiently large depending on ξ. The claim follows."