9th European Intensive Course on Complex Analysis

Fernando Chamizo

Coimbra 2003

A Journey through

The Prime Number Theorem

Notation Along these notes we shall employ extensively Landau's O-notation that we recall briefly here.

The symbols $O(g)$ and $o(g)$ mean respectively a function f such that

$$\limsup_{x\to\infty}\left|\frac{f(x)}{g(x)}\right| < \infty \qquad \text{and} \qquad \lim_{x\to\infty}\frac{f(x)}{g(x)} = 0.$$

Note that $f = O(g)$ is only a short way of saying $|f(x)| \le C|g(x)|$ for some positive constant C and x large enough.

If f and g has the same asymptotic behavior, *i.e.* $\lim f/g = 1$, we shall write $f \sim g$. Typically we shall consider the asymptotic behavior when $x \to \infty$, otherwise it will be explicitly indicated.

1 Warming up

The basic cornerstone in prime number distribution theory is the simple and beautiful *Euler's identity*:

$$\boxed{\sum_{n=1}^{\infty}\frac{1}{n^s} = \prod_p \left(1 - \frac{1}{p^s}\right)^{-1}}$$

where $s > 1$, to assure the convergence, and p runs over the prime numbers. This identity is equivalent to Fundamental Theorem of Arithmetic (unique factorization into primes), just noting that the right hand side is $\prod(1 + p^{-s} + p^{-2s} + p^{-3s} + \dots)$.

The importance of Euler's identity stems from establishing a link between an analytic object, the Riemann zeta function

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

and an arithmetical object, prime numbers. For instance, Euler himself in 1737 realized that when $s \to 1^+$ the divergence of harmonic series implies that there are infinitely many prime numbers.

The giant step toward the understanding of the distribution of primes was given by Riemann who, in his celebrated memoir of 1859, considered ζ as a function of complex variable and proved that it can be extendend to a meromorphic function on the whole complex plane. There are quite elementary proofs of this fact. For instance, the simple identity

$$\left(1 - \frac{2}{2^s}\right)\zeta(s) = \sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{n^s}$$

proves that ζ has a meromorphic continuation to $\{\Re s > 0\}$ (note that $|n^{-s} - (n+1)^{-s}| < C(s)|n^{-s-1}|$) with a simple pole at $s = 1$ with residue 1. More generally, Taylor expansion $(1-x)^{-s}-1 = \sum a_m x^m$ with a_m the generalized binomial coefficient $\binom{s+m-1}{m}$, implies

$$\begin{aligned}\sum_{m=1}^{\infty} 2^{-s-m} a_m \zeta(s+m) &= \sum_{n=1}^{\infty} (2n)^{-s} \sum_m a_m (2n)^{-m} \\ &= \sum_{n=1}^{\infty} (2n)^{-s} \left(\left(1 - \frac{1}{2n}\right)^{-s} - 1 \right) = \left(1 - \frac{2}{2^s}\right) \zeta(s).\end{aligned}$$

Hence meromorphic continuation to $\{\Re s > k\}$ implies meromorphic continuation to $\{\Re s > k-1\}$ and ζ gets extended (of course uniquely) to a holomorphic function on $\mathbb{C} - \{1\}$ with a simple pole at $s = 1$ with residue 1.

Primes appear in an involved way in Euler's identity. It would be desirable a relation between ζ and prime numbers counting function, *i.e.*

$$\pi(x) = \sum_{p \le x} 1 = |\{p \le x \ : \ p \text{ is a prime number}\}|.$$

But it is technically simpler to establish this connection through the, in principle unnatural, function

$$\psi(x) = \sum_{n \le x} \Lambda(n) \qquad \text{with} \qquad \Lambda(n) = \begin{cases} \log p & \text{if } n = p^k \text{ with } p \text{ prime} \\ 0 & \text{otherwise} \end{cases}$$

Proposition 1.1 *For* $\Re s > 1$

$$-\frac{\zeta'(s)}{\zeta(s)} = \sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^s} \qquad \textit{and} \qquad -\frac{\zeta'(s)}{\zeta(s)} - \frac{s}{s-1} = s \int_1^{\infty} (\psi(x) - x) x^{-s-1}\, dx.$$

Proof: By logarithmic differentiation (Euler's favorite trick)

$$\zeta(s) = \prod_p (1 - p^{-s})^{-1} \Rightarrow \frac{\zeta'(s)}{\zeta(s)} = -\sum_p \frac{\log p}{1 - p^{-s}} p^{-s} = -\sum_p \left(\frac{\log p}{p^s} + \frac{\log p}{p^{2s}} + \dots \right).$$

The convergence for $\Re s > 1$ is assured comparing with a geometric series and first formula follows.

Now it is not hard to obtain

$$-\frac{\zeta'(s)}{\zeta(s)} = \sum_{m=1}^{\infty} \psi(m) \left(\frac{1}{m^s} - \frac{1}{(m+1)^s} \right) = \sum_{m=1}^{\infty} s \int_m^{m+1} \frac{\psi(x)}{x^{s+1}}\, dx = s \int_1^{\infty} \frac{\psi(x)}{x^{s+1}}\, dx$$

and the second formula is proved noting that $s/(s-1) = s \int_1^\infty x^{-s}\, dx$. $\square$

We know that ζ is analytic and $\zeta(s) \sim (s-1)^{-1}$ as $s \to 1$, hence $-\zeta'(s)/\zeta(s) - s/(s-1) \to 0$, and the second identity of the proposition suggests that $\psi(x)$ should be well approximated by x. It will be the content of the prime number theorem, but firstly we want to know what does it mean in terms of $\pi(x)$.

Proposition 1.2 *Let $E = E(x)$ be an increasing function such that $\psi(x) = x + O(E(x))$, then $\pi(x) = \mathrm{li}(x) + O(x^{1/2} + E(x)/\log x)$ where $\mathrm{li}(x)$ is the* integral logarithm *$\int_2^x dt/\log t$.*

Proof: It is not difficult to prove that $\pi(x) = \sum_{2 \le n \le x} \Lambda(n)/\log n + O(x^{1/2})$ (use that the sum equals $\pi(x) + \frac{1}{2}\pi(x^{1/2}) + \frac{1}{3}\pi(x^{1/3}) + \dots$ and the bound $\pi(x^{1/n}) \le x^{1/n}$). On the other hand,

$$\sum_{2 \le n \le x} \frac{\Lambda(n)}{\log n} = \frac{\psi(x)}{\log x} + \sum_{2 \le n \le x} \Lambda(n) \int_n^x \frac{dt}{t \log^2 t} = \frac{\psi(x)}{\log x} + \int_2^x \frac{\psi(t)\, dt}{t \log^2 t},$$

where the second equality is just partial summation in the form $a_2 b_2 + a_3 b_3 + a_4 b_4 + \dots = a_2(b_2 - b_3) + a_3(b_3 - b_4) + \dots$ with $a_n = \Lambda(n)$ and $b_n = \int_n^x$. This gives

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

because integrating by parts $\int dt/\log t = x/\log x + \int dt/\log^2 t$. $\square$

According to these results, $\psi(x) \sim x$ translate into $\pi(x) \sim \mathrm{li}(x)$ or equivalently into $\pi(x) \sim x/\log x$ (l'Hôpital rule proves $\mathrm{li}(x) \sim x/\log x$). Any of these asymptotic formulas is called *Prime Number Theorem* (abbreviated as PNT in the following). Usually one wants to go beyond when estimating the size of the error term (the function $\mathcal{E}(x)$ below).

Theorem 1.3 (PNT with error term) *It holds*

$$\pi(x) = \mathrm{li}(x) + O(\mathcal{E}(x))$$

for some function $\mathcal{E}(x) = o(\mathrm{li}(x))$.

In this notes we shall prove this theorem with $\mathcal{E}(x) = xe^{-\frac{1}{6}\sqrt{\log x}}$. In fact, thanks to previous proposition we shall forget about $\pi(x)$, and prove $\psi(x) = x + O(\mathcal{E}(x))$. Even today it is not known a valid error term verifying $\mathcal{E}(x) = O(x^\alpha)$ for some $\alpha < 1$. As we shall see later, this is related to the so-called *Riemann hypothesis*.

2 PNT timelines

- 1849 Gauss conjetures that $\mathrm{li}(x)$ approximates $\pi(x)$.
- 1851 Chebyshev proves $C_1 x/\log x < \pi(x) < C_2 x/\log x$ with explicit C_i.
- 1859 Riemann writes his celebrated 8-paged memoir containing a proof of PNT with serious gaps, using complex analysis.
- 1896 Hadamard and de la Vallée Poussin prove (independently) PNT.
- 1948 Erdős and Selberg find the first "elementary proof" of PNT.
- 1958 Vinogradov and Korobov find the best known error term.
- ????≥2003 Somebody proves Riemann Hypothesis.

3 A proof (?) of PNT for dreamers

In this section we give a fake proof à la Riemann that is non rigorous but contains all the ingredients of the real proof. At first sight it seems that the missing points are of technical nature and not difficult to fill, but probably subsequent pages will show a different truth.

The starting point is the formula valid for $\Re s > 1$

$$-\frac{\zeta'(s)}{\zeta(s)} = \sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^s}. \tag{3.1}$$

Let L be the line $\{\Re s = c\}$, for some $1 < c < 2$, and R_T, S_T the "infinite rectangles" $\{\Im s \le T,\ \Re s \le c\}$, $\{\Im s \le T,\ \Re s \ge c\}$, respectively. For $x > 2$, $x \notin \mathbb{Z}$,

$$\begin{aligned}
-\int_L \frac{\zeta'(s)}{\zeta(s)} \frac{x^s}{s}\, ds &= \sum_{n<x} \Lambda(n) \int_L \left(\frac{x}{n}\right)^s \frac{ds}{s} + \sum_{n>x} \Lambda(n) \int_L \left(\frac{x}{n}\right)^s \frac{ds}{s} \\
&= \sum_{n<x} \Lambda(n) \lim_{T\to\infty} \int_{\partial R_T} \left(\frac{x}{n}\right)^s \frac{ds}{s} + \sum_{n>x} \Lambda(n) \lim_{T\to\infty} \int_{\partial S_T} \left(\frac{x}{n}\right)^s \frac{ds}{s}
\end{aligned}$$

By Cauchy's integral formula, the first integral equals $2\pi i$ (there is a simple pole at $s = 0$) and the second integral equals 0 (no poles in S_T). Hence we have a neat analytic formula for our favorite arithmetical function:

$$\psi(x) = \frac{1}{2\pi i}\int_L f(s)\,ds \qquad \text{where} \qquad f(s) = \frac{\zeta'(s)}{s\zeta(s)}x^s.$$

As ζ is meromorphic so is $f(s)$. Residue theorem gives

$$\psi(x) = -\lim_{T\to\infty}\frac{1}{2\pi i}\int_{\partial R_T}\frac{\zeta'(s)}{\zeta(s)}\frac{x^s}{s}\,ds = \sum_{s\in\mathcal{P}}\mathrm{Res}(f,s)$$

where $\mathcal{P}$ is the set of poles of f. Recalling that $s = 1$ is the unique pole of ζ, it follows $\mathcal{P} = \{0,1\}\cup\mathcal{Z}$ where $\mathcal{Z}$ is the set of zeros of ζ. Moreover $\mathrm{Res}(f,0) = -\zeta'(0)/\zeta(0)$, $\mathrm{Res}(f,1) = x$ (because $\zeta(s)\sim 1/(s-1)$ as $s\to 1$); and if $z\in\mathcal{Z}$ is a zero of multiplicity m, $\mathrm{Res}(f,z) = mx^z/z$. Therefore

$$\psi(x) = x - \frac{\zeta'(0)}{\zeta(0)} - \sum_{z\in\mathcal{Z}}\frac{x^z}{z} \tag{3.2}$$

where each zero is repeated in the summation according to its multiplicity.

After this amazing formula, one can claim that the answer to any question regarding to the distribution of prime numbers is embodied in the distribution of the zeros of Riemann's zeta function. In particular, if $\Re z < 1$ for every $z\in\mathcal{Z}$, then $|x^z| = x^{\Re z} = o(x)$ and, trusting on good convergence properties of the series, PNT follows in the form $\psi(x)\sim x$.

Let us finish with an unbelievably ingenious proof due to Mertens of the missing point $\mathcal{Z}\subset\{\Re s < 1\}$. The convergence for $\Re s > 1$ of the series in (3.1) implies that $\mathcal{Z}\subset\{\Re s\le 1\}$. Assume that there exists a zero z of ζ with $\Re z = 1$, say $z = 1 + Bi$ (note that after (3.2), the existence of this zero ruins PNT), and consider $g(s) = \zeta^3(s)\zeta^4(s+Bi)\zeta(s+2Bi)$. This is a meromorphic function with a zero at $s = 1$ (because $3 < 4$) and consequently, $\lim_{x\to 1^+}\log|g(x)| = -\infty$ for $x\in\mathbb{R}$. On the other hand, for $x > 1$, using the definition of ζ and Taylor expansion:

$$\begin{aligned}\Re\log g(x) &= -\Re\sum_p\big(3\log(1-p^{-x}) - 4\log(1-p^{-x-Bi}) - \log(1-p^{-x-2Bi})\big)\\ &= \Re\sum_p\sum_{n=1}^{\infty}\frac{1}{n}p^{-nx}(3 + 4p^{-Bni} + p^{-2Bni}).\end{aligned}$$

But the term between parenthesis is positive, because primer calculus course techniques prove $3 + 4\cos\alpha + \cos(2\alpha)\ge 0$ (or apply double-angle formulas to $2(1+\cos\alpha)^2 > 0$). Hence $\Re\log g(x) = \log|g(x)| > 0$ and it contradicts $\log|g(x)|\to-\infty$.

4 A nice symmetric function

We are going to prove at once that ζ extends to a meromorphic function on the whole complex plane (this is our second proof of this fact) and that it has a kind of symmetry with respect to the line $\Re s = 1/2$. The formula expressing this symmetry is the well known *functional equation* and is formally a consequence of Poisson summation formula applied to $f(x) = x^{-s}$ crossing out infinities. The actual proof given by Riemann establishes in general an interesting link between functional equations and modular relations. Modern Number Theory is plenty of underlying modular forms, and this explains why we have a lot of similar-looking functional equations.

Riemann's starting point was the integral representation for $\Re s > 0$ of Γ-function after a change of variable:

$$\Gamma(s/2) = \int_0^\infty t^{s/2-1} e^{-t}\, dt = \pi^{s/2} n^s \int_0^\infty t^{s/2-1} e^{-\pi n^2 t}\, dt.$$

Summing on n, for $\Re s > 1$,

$$\pi^{-s/2}\Gamma(s/2)\zeta(s) = \frac{1}{2}\int_0^\infty t^{s/2-1}(\theta(t)-1)\, dt \qquad \text{where} \quad \theta(t) = \sum_{n=-\infty}^{\infty} e^{-\pi n^2 t}.$$

Now we can apply comfortably Poisson summation formula through the closed (modular) relation* $\theta(t) = t^{-1/2}\theta(1/t)$. It allows to transform the part $\int_0^1$ of the integration which is reponsible of the lack of convergence for $\Re s \le 1$.

$$\int_0^1 t^{s/2-1}(\theta(t)-1)\, dt = \int_0^1 t^{s/2-1}(t^{-1/2}\theta(1/t)-1)\, dt = \int_1^\infty t^{-s/2-1}(t^{1/2}\theta(t)-1)\, dt.$$

Substituting, after some calculations, we obtain

$$\pi^{-s/2}\Gamma(s/2)\zeta(s) = \frac{1}{s(s-1)} + \frac{1}{2}\int_1^\infty (t^{s/2-1} + t^{-s/2-1/2})(\theta(t)-1)\, dt. \tag{4.1}$$

The right hand side defines a meromorphic function with $s = 0, 1$ as only poles. Moreover, it remains invariant under the change $s \mapsto 1-s$. Hence ζ is a meromorphic function and satisfies the functional equation

$$\boxed{\pi^{-s/2}\Gamma(s/2)\zeta(s) = \pi^{-(1-s)/2}\Gamma((1-s)/2)\zeta(1-s).} \tag{4.2}$$

*See the appendix.

Following (partially) Riemann, we can introduce the entire function

$$\xi(s) = \frac{1}{2}s(s-1)\pi^{-s/2}\Gamma(s/2)\zeta(s)$$

and functional equation (4.2) reduces to

$$\xi(s) = \xi(1-s).$$

From (4.1) we conclude (again) that ζ has a single pole at $s = 1$ with residue 1. Using that Γ is holomorphic up to simples poles at $0,\ -1,\ -2,\ -3\ldots$, (4.2) proves ζ is holomorphic on $\mathbb{C} - \{1\}$ and has simple zeros at $s = -2,\ -4,\ -6,\ldots$. These zeros are called *trivial zeros*. As $\pi^{-s/2}\Gamma(s/2)\zeta(s)$ does not vanish for $\Re s > 1$, there are not other zeros in $\Re s < 0$, hence *non-trivial zeros* are in the so called *critical strip* $0 \le \Re s \le 1$. Summarizing:

$$\{\text{Poles of } \zeta\} = \{1\} \qquad \{\text{Zeros of } \zeta\} = 2\mathbb{Z}^- \cup \{\text{Non-trivial zeros}\}$$
$$\{\text{Poles of } \xi\} = \emptyset \qquad \{\text{Zeros of } \xi\} = \{\text{Non-trivial zeros of } \zeta\}$$

5 Zeros here and there

The fake proof suggests that the core of prime number distribution theory is the study of the zeros of ζ. We have already separated the "trivial zeros" $2\mathbb{Z}^-$, and the whole problem is to understand the zeros in the critical strip $0 \le \Re s \le 1$. We shall denote with ρ each of these *non-trivial zeros*.

Some basic results in Complex Analysis play an important role in the subsequent study. From the historical point of view, it can be claimed that a part of the basis of Complex Analysis was created in connection with the proof of PNT.

Our first result is a neat relation between ζ'/ζ and the non-trivial zeros.

Theorem 5.1 *For a certain constant C_0, it holds*

$$\frac{\zeta'(s)}{\zeta(s)} = C_0 - \frac{1}{s-1} - \frac{\Gamma'(s/2+1)}{2\Gamma(s/2+1)} + \sum_\rho \left(\frac{1}{s-\rho} + \frac{1}{\rho}\right). \tag{5.1}$$

Proof: Comparing (4.1) with the definition of Γ (use $\theta(t) - 1 = O(e^{-t})$ and $|t^\alpha| \le t^{|\alpha|}$), it is deduced

$$|\xi(s)| = O((1+|s|^2)\Gamma(|s|+1)/2)) = O(e^{K|s|\log(|s|+1)}).$$

Hence ξ is an entire function of order one, and according to Hadamard finite order function theory*, there exists a factorization

$$\xi(s) = e^{A+Bs}\prod_{\rho}(1 - s/\rho)e^{s/\rho}$$

for certain constants A, B. Calculating the logarithmic derivative ξ'/ξ with this formula and the definition of ξ, the theorem follows. $\Box$

A superabundance of zeros could give a too large error term in PNT. The following result shows that it is not the case.

Proposition 5.2 *Let $N(T)$ be the number of non-trivial zeros ρ, counted with multiplicity, such that $|\Im\rho| \le T$. Then*

$$N(T+1) - N(T) = O(\log T) \qquad \textit{and} \qquad N(T) = O(T\log T).$$

Proof: Of course the latter formula is a straightforward consequence of the former. Using (5.1) with $s = 2 + iT$, $T \ge 2$, we obtain $\sum_\rho((s-\rho)^{-1} + \rho^{-1}) = O(\log T)$. Taking real parts:

$$O(\log T) = \Re\sum_{\rho=\beta+i\gamma}\left(\frac{1}{2-\beta+i(T-\gamma)} + \frac{1}{\beta+i\gamma}\right) \ge \sum_{\rho}\frac{1}{4+(T-\gamma)^2}.$$

And it implies $N(T+1) - N(T) = O(\log T)$. $\Box$

Although it is not necessary in our proof of PNT, with some extra effort (using argument principle) previous result can be sharpened as follows:

Theorem 5.3 *It holds the asymptotic formula*

$$N(T) = \frac{T}{2\pi}\log\frac{T}{2\pi} - \frac{T}{2\pi} + O(\log T).$$

The identity (5.1) show that the value of $\zeta'(s)/\zeta(s)$ is greatly influenced by the closest zeros to s. The next result quantifies this phenomenon.

*See the Appendix.

Proposition 5.4 *We have*

$$\frac{\zeta'(s)}{\zeta(s)} = \sum_{\rho\,:\,|s-\rho|<1} (s-\rho)^{-1} + O(\log|t|)$$

with a uniform O-constant for $s=\sigma+it$, $\sigma\ge -1$, $|t|\ge 2$ and $s\notin\mathcal{Z}$.

Proof: Substracting (5.1) for $s=\sigma+it$ and $s=2+it$,

$$\frac{\zeta'(s)}{\zeta(s)} = \sum_{\rho}\left(\frac{1}{s-\rho}-\frac{1}{2+it-\rho}\right) + O(\log|t|).$$

For $|s-\rho|>1$, say $\rho=\beta+i\gamma$,

$$\left|\frac{1}{s-\rho}-\frac{1}{2+it-\rho}\right| = \frac{2-\sigma}{|s-\rho||2+it-\rho|} \le C\frac{1}{4+(t-\gamma)^2}$$

with C an absolute constant. We have already seen that the last term contributes $O(\log|t|)$ when summing over ρ. On the other hand, $\sum_{|s-\rho|<1}(2+it-\rho)^{-1} = O(\log|t|)$ because there are $O(\log|t|)$ zeros $\rho=\beta+i\gamma$ with $\gamma\in[t-1,t+1]$. $\Box$

We have studied so far "vertical distribution" of the zeros, but in order to prove PNT we need some horizontal control, namely we want to separate the real part of the zeros from $\Re s=1$.

Theorem 5.5 *There exists a positive constant C such that ζ does not vanish for $s=\sigma+it$ in the region*

$$\sigma > 1-\frac{1}{35\log(|t|+C)}.$$

Proof: From (3.1) and Mertens' argument (see the end of the fake proof), if $\rho_0=A+Bi$ is a non-trivial zero, for $\sigma>1$

$$-3\frac{\zeta'(\sigma)}{\zeta(\sigma)} - 4\Re\frac{\zeta'(\sigma+Bi)}{\zeta(\sigma+Bi)} - \Re\frac{\zeta'(\sigma+2Bi)}{\zeta(\sigma+2Bi)} \ge 0. \tag{5.2}$$

Note that for $s=\sigma+iB$ with $\sigma>1$ and any non-trivial zero ρ, it holds $\Re(s-\rho)^{-1}$, $\Re\rho^{-1}>0$. Taking this into account, by (5.1), for C large enough

$$-\Re\frac{\zeta'(\sigma+Bi)}{\zeta(\sigma+Bi)} \le -\Re(s-\rho_0)^{-1}+\frac{1}{2}\log(|B|+C), \quad -\Re\frac{\zeta'(\sigma+2Bi)}{\zeta(\sigma+2Bi)} \le \frac{1}{2}\log(|B|+C).$$

On the other hand, $-\zeta(s)/\zeta(s) \sim (s-1)^{-1}$ as $s \to 1$, implies that $-\zeta(\sigma)/\zeta(\sigma) < 1.01/(\sigma-1)$ for σ in some interval $(1, 1+\epsilon]$. Substituting these inequalities in (5.2)

$$\frac{3.03}{\sigma-1} + \frac{4}{\sigma-A} + \frac{5}{2}\log(|B|+C) > 0.$$

Choose $\sigma = 1+2/(11\log(|B|+C))$ (suppose C large enough to assure $\sigma \in (1, 1+\epsilon]$). If $A > 1 - 1/(35\log(|B|+C))$, we get a contradiction. $\square$

6 The real (complex) proof

The first gap in our fake proof of PNT is the application of Cauchy's integral formula to some suspicious infinite regions. For each $T > 2$, let $L_T = L \cap \{\Im s \le T\}$, *i.e.* L_T is the segment connecting $c - iT$ with $c + iT$. Consider $x > 2$ far away from integers, say $\mathrm{Frac}(x) = 1/2$. Exponential decay of $(x/n)^s$ is enough to deduce:

$$\psi(x) = \frac{1}{2\pi i}\sum_{n<x}\Lambda(n)\int_{\partial R_T}\left(\frac{x}{n}\right)^s\frac{ds}{s} + \frac{1}{2\pi i}\sum_{n>x}\Lambda(n)\int_{\partial S_T}\left(\frac{x}{n}\right)^s\frac{ds}{s}.$$

(Remember that $R_T = \{\Im s \le T,\ \Re s \le c\}$ and $S_T = \{\Im s \le T,\ \Re s \ge c\}$). It is not hard to prove for $t > 0$, $t \ne 1$, that

$$\int_{c+iT}^{\pm\infty+iT}\frac{t^s}{s}\,ds = O\left(\frac{t^c}{T|\log t|}\right).$$

Using this bound, we obtain (note that $|n-x| < x/2 \Rightarrow (x/n)^c = O(1)$)

$$\psi(x) = \frac{1}{2\pi i}\int_{L_T} f(s)\,ds + O\left(\frac{1}{T}\sum_{|n-x|<x/2}\frac{\Lambda(n)}{|\log(x/n)|} + \frac{x^c}{T}\sum_{|n-x|\ge x/2}\frac{\Lambda(n)}{n^c|\log(x/n)|}\right)$$

(Remember that $f(s) = -\frac{\zeta'(s)}{\zeta(s)}\frac{x^s}{s} = \sum\Lambda(n)\left(\frac{x}{n}\right)^s\frac{1}{s}$). Taylor expansion leads to $|\log(x/n)|^{-1} = O(x/|n-x|)$ and the first term in the error contributes $O(\frac{x}{T}\log^2 x)$. On the other hand the second sum is $O(|\zeta'(c)/\zeta(c)|) = O((c-1)^{-1})$. If we choose $c = 1 + 1/\log x$ to clear up the calculations, we get

$$\psi(x) = \frac{1}{2\pi i}\int_{L_T} f(s)\,ds + O\left(\frac{x}{T}\log^2 x\right).$$

By Proposition 5.2 it is possible to choose lines $\Im s = \pm T$, for T on each fixed interval of length one, separated from the zeros at least $C/\log T$, for some absolute constant C. Proposition 5.3 implies $|\zeta'/\zeta| = O(\log^2 T)$ on these lines for $\Re s \geq -1$. Consequently

$$\int_{-\infty-iT}^{c-iT} + \int_{c+iT}^{-\infty+iT} f(s)\, ds = O\left(\frac{x}{T}\log^2 T\right)$$

and

$$\psi(x) = \frac{1}{2\pi i}\int_{\partial R_T} f(s)\, ds + O\left(\frac{x}{T}\log^2(xT)\right).$$

Residue theorem gives

$$\psi(x) = x - \frac{\zeta'(0)}{\zeta(0)} + \sum_{n=1}^{\infty}\frac{x^{-2n}}{2n} - \sum_{|\Im\rho|<T}\frac{x^\rho}{\rho} + O\left(\frac{x}{T}\log^2(xT)\right). \tag{6.1}$$

When $T \to \infty$ this gives the so called *explicit formula*, the analog of (3.2). It is useless (because of the conditional convergence) but wonderful:

$$\psi(x) = x - \frac{\zeta'(0)}{\zeta(0)} + \frac{1}{2}\log(1 - x^{-2}) - \sum_\rho \frac{x^\rho}{\rho}.$$

Note that $\psi(x+h) - \psi(x) \leq \log(x+1)$ for every $h \in [0,1]$, then adding an extra $O(\log x)$ term in (6.1), it holds true without restrictions on $\mathrm{Frac}(x)$. Cleaning negligible terms, we have

$$\psi(x) = x - \sum_{|\Im\rho|<T}\frac{x^\rho}{\rho} + O\left(\frac{x}{T}\log^2(xT) + \log x\right).$$

By Theorem 5.5 there are no zeros with $\Re\rho > 1 - 1/(35\log(T+C))$ and $|\Im\rho| \leq T$, and by Proposition 5.2 there are $O(\log N)$ with $N \leq |\Im\rho| \leq N+1$. Hence

$$\sum_{|\Im\rho|<T}\frac{x^\rho}{\rho} = O\left(\sum_{N\leq T}\frac{\log N}{N}x^{1-1/(35\log(T+C))}\right) = O\left(x^{1-1/(35\log(T+C))}\log^2 T\right).$$

Choosing $T = e^{0.17\sqrt{\log x}}$ and substituting, we obtain finally

$$\boxed{\psi(x) = x + O\left(x\, e^{-\frac{1}{6}\sqrt{\log x}}\right)}$$

indeed something slighty better. PNT is (at last) proved.

7 Epilogue: Riemann hypothesis

By the second formula in Proposition 1.1, if $\psi(x) = x + O(x^\beta)$ for every $\beta > \alpha$, then $-\zeta'(s)/\zeta(s) - s/(s-1)$ is holomorphic on $\Re s > \alpha$, in particular all the non-trivial zeros verify $1-\alpha \le \Re\rho \le \alpha$. Hence the best scenario occurs when $\alpha = 1/2$, *i.e.* when the non-trivial zeros keep in single file.

Riemann hypothesis: *If ρ is a non-trivial zero of ζ then $\Re s = 1/2$.*

From the anlytical point of view this is a very strange conjecture because there are no known reasons motivating the lining up of the zeros. And it is even more strange taking into account that in Number Theory there is a huge family of zeta-like complex analytical functions that apparently share the same property.

It is known by extensive numerical analysis that more than the first 10^{11} non-trivial zeros of the Riemann zeta function verify Riemann hypothesis, but so far we do not even know how to prove $\Re\rho \le \delta$ for some $\delta < 1$ and every non-trivial zero ρ.

Although, after more than 140 years, we are desperately far from Riemann hypothesis, in the meantime some theorems have sprung up about the distribution of the zeros that Riemann probably would like. We shall only mention three according to their strength:

(Hardy) There are infinitely many zeros of ζ with real part $1/2$.

(Bohr, Landau et al.) The "density" of zeros on $\Re s \ge \alpha$ for any $\alpha > 1/2$ is arbitrarily small in comparison with the density on $\Re s \ge 1/2$.

(Selberg) A positive proportion of the zeros lay on the line $\Re s = 1/2$.

8 Appendix

In this appendix we shall refresh some topics related to Complex Analysis that we have used in previous pages.

Gamma function:

Gamma function is a kind of natural analytic extension of factorials to complex plane. For $\Re s > 0$ it is defined by the integral formula

$$\Gamma(s) = \int_0^\infty t^{s-1} e^{-t}\, dt.$$

Integrating by parts, $\Gamma(n) = (n-1)!$ for $n \in \mathbb{Z}^+$ and $\Gamma(s+1) = s\Gamma(s)$ in general. This functional equation allows to extend Γ to a meromorphic function with simple poles at $s = 0, -1, -2, \ldots$. In particular $\Gamma(s) \sim s^{-1}$ when $s \to 0$. Far away from these poles, say for instance on $\mathbb{C} - \bigcup B_{0.1}(-n)$, $\Gamma(s) = O\big(e^{|s|\log|s|}\big)$ and $\Gamma'(s)/\Gamma(s) = O(\log|s|)$. Indeed it is possible to replace these bounds by asymptotic formulas (cf. Stirling's formula).

Finite order functions:

Hadamard's finite order function theory allows to factorize entire functions, under some growth condition, into something close to linear factors. Roughly speaking, it is a kind of Fundamental Theorem of Algebra for entire functions. In the order one case, it asserts that for an entire function f satisfying $|f(z)| = O\big(e^{|z|^\alpha}\big)$ for every $\alpha > 1$, it holds

$$f(z) = e^{A+Bz} \prod (1 - z/z_n) e^{z/z_n}$$

where A and B are constants and z_n runs over the zeros of f. Moreover the (possibly infinite) product is absolutely convergent.

Poisson summation:

If f is smooth enough (say for instance $|f|$, $|f'|$ and $|f''|$ integrable) then Poisson summation formula reads

$$\sum_{n=-\infty}^{\infty} f(n) = \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} f(x) e^{-2\pi i n x}\, dx.$$

Choosing $f(x) = e^{-\pi t x^2}$ where $t > 0$ is a parameter, one concludes the functional equation for θ-function, $\theta(t) = t^{-1/2}\theta(1/t)$.

Bibliography

[1] J. B. Conrey. The Riemann Hypothesis. *Notices of the AMS*, March 2003, 341-353.

[2] H. Davenport. *Multiplicative number theory*, 2nd ed. revised by H.L. Montgomery. Graduate texts in mathematics 74. Springer Verlag, 1980.

[3] H.M. Edwards. *Riemann's zeta function*. Academic Press, 1974.

[4] W.J. Ellison, M. Mendès-France. *Les nombres premiers*. Hermann, 1975.

[5] A. Ivić. *The Riemann zeta-function: the theory of the Riemann zeta-function with applications*. New York, wiley, 1985.

[6] B. Riemann. Ueber die Anzahl der Primzahlen unter einer gegebenen Grösse. *Monat. der Königl. Preuss. Akad. der Wissen. zu Berlin aus der Jahre 1859* (1860), 671-680. Also *Gesammelte math. Werke und wissensch. Nachlass* 2. Aufl.1892, 145- 155.

[7] E.C. Titchmarsh. *The theory of the Riemann zeta-function*, 2nd ed. (revised by D.R. Heath-Brown). Oxford, Clarendon, 1986.

The Circle Method

Basic ideas

1 The method

Some of the most famous problems in Number Theory are additive problems (Fermat's last theorem, Goldbach conjecture...). It is just asking whether a number can be expressed as a sum of other numbers, all of them belonging to some subsets of integers.

Suppose that we have sequence of integers $0 \le a_1 < a_2 < a_3 < \dots$ and we want to know if a large positive integer N is a sum of k terms of this sequence (repetitions are allowed). We can be even more ambitous and ask about a good (possibly asymptotic) approximation for

$$r_k(N) = \#\{(n_1, n_2, \dots, n_k) \ : \ N = a_{n_1} + a_{n_2} + \dots + a_{n_k}\}.$$

This is the kind of problems treated by circle method. The starting point is to consider the analytic function $F(z) = \sum z^{a_n}$ and note that

$$r_k(N) = \text{coeff. of } z^N \text{ in } (F(z))^k.$$

We can write it in a fancy way involving a complex integral:

$$r_k(N) = \frac{1}{2\pi i} \int_C (F(z))^k \frac{dz}{z^{N+1}} \tag{1.1}$$

where C is a circle $\{z \in \mathbb{C} \ : \ |z| = r\}$ with $0 < r < 1$. In principle this seems rather unnatural and useless, we can see the circle but not the method. The guidelines for success in this approach come from a general philosophy in Analytic Number Theory: *extract arithmetical information from the singularities*. For instance, the study of the distribution of prime numbers depends heavily on the poles of ζ'/ζ.

In (1.1) the only singularity, the high order pole at $z = 0$, has been introduced rather artificially and an application of residue theorem simply dismantles the formula recovering the definition of $r_k(N)$. We have to escape from $z = 0$. On the other hand, in the most of the practical examples, the unit circle is the natural boundary of the holomorphic function F, and hence it is impossible to push r beyond 1 in search of new singularities. In the circle method one takes r close enough to 1 in order to feel the influence of the "main singularities" on the boundary, but small enough to avoid uncontrolled "interferences".

The circle method appeared firstly in a paper by Hardy and Ramanujan about partitions [3], but it was developed by Hardy and Littlewood (it is sometimes called Hardy-Littlewood method). They introduced the nowadays standard terminology *major arcs* and *minor arcs* referring to a subdivision of C. In the former the influence of near singularities leads to a good approximated formula, while in the latter we have to content ourselves with a bound.

In practice the definition of major and minor arcs depends on diophantine approximation properties. This is not so strange, if z tends to 1 radially there is not any cancellation in F and this “big singularity” causes a big major arc. On the other hand, if z tends radially to $e^{2\pi i a/q}$, the size of the singularity, if it exists, depends on the distribution of a_n modulus q, but typically one hopes less cancellation and a bigger major arc when q is small (because of lower oscillation). In this scheme, minor arcs correspond to directions far apart from having small denominator slopes.

2 Sums of squares

One can ask about the number of representations of a (large) positive integer N as a sum of k squares. By technically reasons (write a simpler final formula) we shall assume that $8|k$ although this is not essential for the method. As square function is not injective, $(-n)^2 = n^2$, we shall forget about a_n giving an *ad hoc* definition

$$r_k(N) = \#\{(n_1, n_2, \dots, n_k) \in \mathbb{Z}^k \ : \ n_1^2 + n_2^2 + \cdots + n_k^2 = N\}.$$

Now (1.1) reads

$$r_k(N) = \frac{1}{2\pi i}\int_C (F(z))^k \frac{dz}{z^{N+1}} \qquad \text{with} \ \ F(z) = \sum_{n=-\infty}^{\infty} z^{n^2} = 1 + 2\sum_{n=1}^{\infty} z^{n^2}.$$

A change of variable $z \mapsto e^{2\pi i z}$ leads to a famous θ-function

$$r_k(N) = \int_L \theta^k(z) e^{-2N\pi i z}\, dz \qquad \text{with} \ \ \theta(z) = \sum_{n=-\infty}^{\infty} e^{2\pi i n^2 z} \tag{2.1}$$

and L the horizontal segment $\{0 \le \Re z < 1, \ \Im z = y\}$ where $r = e^{-2\pi y}$. We shall choose $y = 1/N$. This is the natural choice because it constitutes a penalty for the terms with $n^2 > N$ which are negligible in order to represent N as a sum of squares.

A fundamental property of θ^k is that it is automorphic. This means a kind of invariance by certain (Fuchsian) group of fractional linear transformations. Namely

$$\theta^k\left(\frac{az+b}{4cz+d}\right) = (4cz+d)^{k/2}\theta^k(z)$$

when $a, b, c, d \in \mathbb{Z}$ and $ad - 4bc = 1$. It allows to pass the information from some arcs to others. In fact, it can be proved that it is enough to study the arcs close to 0, 1/2 and 1/4 (for the *cognoscenti*, these are the inequivalent cusps). Without entering into details, it turns out that if a/q is an irreducible fraction it holds

$$\theta^k(z) \sim (qz-a)^{-k/2} \text{ if } 4|q, \quad \theta^k(z) \sim (2(qz-a))^{-k/2} \text{ if } 2\not| q \text{ and } \theta(z) \approx 0 \text{ otherwise,}$$

as $qz - a \to 0$ with $\Im(qz - a)^{-1} \to -\infty$. If $0 \le a < q \le \sqrt{N}$, this is the case for $z = a/q + (u+i)/N$ with $u/N = o\big((q\sqrt{N})^{-1}\big)$ when $N \to \infty$. Hence for N large, the contribution to the integral in (2.1) of this "arc" when $4|q$ is asymptotically equal to

$$\frac{1}{N}\int_{-\sqrt{N}/q}^{\sqrt{N}/q} \big(\frac{qu}{N} + \frac{qi}{N}\big)^{-k/2} e^{-2\pi i a N/q} e^{-2\pi i(u+i)}\, du$$
$$= q^{-k/2} N^{k/2-1} e^{-2\pi i a N/q} \int_{-\sqrt{N}/q}^{\sqrt{N}/q} (u+i)^{-k/2} e^{-2\pi i(u+i)}\, du.$$

If q is small enough in comparison with $\sqrt{N}$ this is $\sim C_k q^{-k/2} N^{k/2-1} e^{-2\pi i a N/q}$, otherwise the contribution is small and we can consider that we are dealing with a minor arc. Taking into account also the case $2\not| q$, and adding all the contributions, it follows

$$r_k(N) \sim C_k N^{k/2-1} \sum_{q=1}^{\infty} \sum_{\substack{a=0 \\ (a,q)=1}}^{q} \epsilon_q q^{-k/2} e^{-2\pi i a N/q} \qquad \text{with} \quad \epsilon_q = \begin{cases} 1 & \text{if } 4|q \\ 2^{-k/2} & \text{if } 2\not| q \\ 0 & \text{otherwise} \end{cases}$$

Some tricky (elementary but not easy) manipulations* allow to simplify enormously the summation. The final result is

$$r_k(N) \sim A_k N^{k/2-1} \sum_{d|N} (-1)^{N+N/d} d^{1-k/2}.$$

The value of A_k can be explicitly computed in terms of the $k/2$-th Bernoulli number.

3 Sums of primes

One of the most impressive approaches to Goldbach conjecture is Vinogradov's theorem asserting that every large enough odd integer is a sum of three primes.

*See the appendix

This is one of the highlights of the circle method and, although the details are rather involved, it is possible to sketch the proof according to the main lines mentioned before.

In this case $\{a_n\}$ is the sequence of prime numbers and we want to approximate $r_3(N)$ for N large and odd. Therefore (1.1) reads

$$r_3(N) = \frac{1}{2\pi i}\int_C (F(z))^3 \frac{dz}{z^{N+1}} \qquad \text{with} \quad F(z) = \sum_p z^p.$$

In this case F has not automorphic properties and Vinogradov considered that it is useless to preserve the whole series for F. We do not lose anything truncating F to $p \le N$ and pushing C to the unit circle. With a change of variable $z = e^{2\pi i x}$ we obtain a Fourier series version of the previous formula:

$$r_3(N) = \int_{-1/2}^{1/2} (S(x))^3 e^{-2\pi i N x}\, dx \qquad \text{with} \quad S(x) = \sum_{p\le N} e^{2\pi i p x}.$$

Instead of studying the radial behavior of F, we have to face with the trigonometrical sum $S(x)$. It is a completely equivalent procedure but technically simpler.

Using prime number theorem, we have $S(0) \sim N/\log N$. If x is smaller than $1/N$ then we have a similar approximation because $e^{2\pi i p x}$ does not oscillate (use Taylor expansion). And it seems that for x a little greater, the oscillation should cause some cancellation. In fact, using prime number theorem with error term and partial summation, it is not difficult to get an asymptotic formula reflecting this behavior in a "major arc" I_0 slighty greater that $[-1/N, 1/N]$. So we have

$$\int_{I_0} (S(x))^3 e^{-2\pi i N x}\, dx \sim C\frac{1}{N}\left(\frac{N}{\log N}\right)^3 = C\frac{N^2}{\log^3 N}.$$

In fact, it holds $C = 1/2$.

In the same way, if a/q is an irreducible fraction,

$$S(a/q) = e^{2\pi i/q} \sum_{\substack{p\le N\\ q|ap-1}} 1 + e^{4\pi i/q} \sum_{\substack{p\le N\\ q|ap-2}} 1 + e^{6\pi i/q} \sum_{\substack{p\le N\\ q|ap-3}} 1 + \dots$$

And we need an asymptotic formula for the number of primes in the arithmetic progression $qn + c$. If c and q are not coprime, of course there are finitely many. On the other hand, there are $\phi(q)$ values of $c \in [1, q]$ which are coprime with q, and prime number theorem for arithmetic progressions asserts that prime numbers are equidistributed in the corresponding $\phi(q)$ progressions. Hence

$$S(a/q) \sim \frac{N}{\phi(q)\log N} \sum_{\substack{c=0\\(c,q)=1}}^{q} e^{2\pi i c/q}$$

(if the sum does not vanish). Reasoning as before, we can find a "major arc" $I_{a/q}$ around a/q slighty greater than $[a/q - 1/N, a/q + 1/N]$.

The problem is that the error term in prime number theorem for arithmetic progressions is rather unknown when q varies, and we are forced to take q very small in comparison with N (something like a logarithm). This causes minor arcs to be really large and Vinogradov had to use very involved arguments to obtain acceptable non trivial bounds on them. If we skip this big problem and we consider only major arcs contribution, we get:

$$r_3(N) \sim \frac{1}{2}\frac{N^2}{\log^3 N} \sum_{q=1}^{\infty} \sum_{\substack{a=0\\(a,q)=1}}^{q} \left(\frac{1}{\phi(q)} \sum_{\substack{c=0\\(c,q)=1}}^{q} e^{2\pi i c/q} \right)^3 e^{-2\pi i N a/q}.$$

Again, with very tricky but elementary arguments, the sum can be evaluated explictly, giving

$$r_3(N) \sim \frac{1}{2}\frac{N^2}{\log^3 N} \prod_{p|N} \left(1 - \frac{1}{(p-1)^2}\right) \prod_{p\nmid N} \left(1 + \frac{1}{(p-1)^3}\right).$$

Note that for N even, the first product vanishes ruinning the asymptotic formula.

4 Appendix

The key observation to simplify the formulas obtained by circle method in the previous examples, is that the *Ramanujan sum*

$$c_q(-N) = \sum_{\substack{a=0\\(a,q)=1}}^{q} e^{-2\pi i a N/q}$$

is multiplicative in q, *i.e.* $c_{q_1}(-N)\cdot c_{q_2}(-N) = c_{q_1q_2}(-N)$ if q_1 and q_2 are coprimes. This is a simple consequence of chinese remainder theorem. For each prime number p, let l be a non-negative integer such that $p^l|N$ and $p^{l+1}\not|N$. The following elementary properties allow to evaluate $c_q(-N)$:

$$
\begin{aligned}
&0<l<m \Rightarrow c_{p^m}(-N) = p^l c_{p^{m-l}}(-N/p^l),\ 0<m\le l \Rightarrow c_{p^m}(-N) = p^m - p^{m-1}\\
&0=l<m \ \Rightarrow\ c_p(-N) = -1, \quad c_{p^{m+1}}(-N) = 0.
\end{aligned}
$$

Hence for any multiplicative arithmetical function f, under suitable convergence conditions,

$$
\sum_{q=1}^{\infty} f(q)c_q(-N) = \prod_p \left(1 + f(p)c_p(-N) + f(p^2)c_{p^2}(-N) + \dots\right) = \prod_p \mathcal{F}_p
$$

and $\mathcal{F}_p$ is actually a finite sum.

For instance, in the case of the sum of squares, we can take $f(q) = 2^{k/2}\epsilon_q q^{-k/2}$. Then for $p \neq 2$

$$
\begin{aligned}
\mathcal{F}_p &= 1 + p^{-k/2}(p-1) + (p^2)^{-k/2}(p^2-p) + \dots + (p^l)^{-k/2}(p^l - p^{l-1}) - (p^{l+1})^{-k/2}p^l\\
&= (1-p^{-k/2})\left(1 + p^{1-k/2} + p^{2(1-k/2)} + \dots + p^{l(1-k/2)}\right).
\end{aligned}
$$

Similar manipulations lead, for $p = 2$, to

$$
\mathcal{F}_2 = 1 + 2^{1-k/2} + 2^{2(1-k/2)} + \dots + 2^{l(1-k/2)} - 2\cdot 2^{l(1-k/2)}.
$$

Then the product $\prod \mathcal{F}_p$ is, up to a factor only depending on k, the sum of the $1-k/2$ powers of the divisors of N, substracting twice the divisors containing a maximal power of 2. Note that $N + N/d$ is even if and only if d is even and contains this maximal power, and the closed form $\sum_{d|N}(-1)^{N+N/d}d^{1-k/2}$ follows.

The case of sums of three primes is much easier. The multiplicative function is $f(q) = c_q^3(1)/\phi^3(q)$, hence $\mathcal{F}_p = 1 - c_p(-N)/(p-1)^3$, which is $1-(p-1)^{-2}$ if $p|N$, and $1+(p-1)^{-3}$ if $p\not|N$.

Bibliography

[1] H. Davenport. *Multiplicative number theory*, 2nd ed., revised by Hugh L. Montgomery. Graduate texts in mathematics 74. Springer Verlag, 1980.

[2] W.J. Ellison, M. Mendès-France. *Les nombres premiers*. Hermann, 1975.

[3] G.H. Hardy, S. Ramanujan. Asymptotic formulae in combinatory analysis. *Proc. London Math. Soc.* (2) **17** (1918), 75-115.

[4] R.C. Vaughan. *The Hardy-Littlewood method*. Cambridge tracts in Mathematics 80. Cambridge University Press, 1981.