Contents

1 The Prime Number Theorem **2**
1.1 Chebyshev's bounds 2
1.2 A proof of the PNT 7
1.3 Remarks 13

2 Shnirel'man's theorem **14**
2.1 Shnirel'man's density 15
2.2 Bounding $r(n)$ via Selberg's sieve 17
2.3 Conclusion of the proof 22
2.4 Lemma on numbers λ_d^* 24
2.5 Remarks 26

3 Roth's theorem on arithmetic progressions **27**
3.1 An analytic proof 28
3.2 Proof of the uniform bound on the unit circle 32
3.3 A graph-theoretical proof 36
3.4 Szemerédi's regularity lemma 38
3.5 Remarks 41

Bibliography **42**

Chapter 1

The Prime Number Theorem

1.1 Chebyshev's bounds

Euclid's proof of infinititude of primes. By Euclid, if p_n is the n-th prime then

$$p_{n+1} \leq 1 + \prod_{j=1}^{n} p_j$$

—the number on the right must be divisible by a prime but it cannot be either of $p_1, p_2, \ldots, p_n$. By induction,

$$p_n \leq 2^{2^n}.$$

Form this it folows that

$$\pi(x) \geq \left\lfloor \frac{\log(\log x / \log 2)}{\log 2} \right\rfloor \geq \frac{\log \log x}{\log 2} - 1 - \frac{\log \log 2}{\log 2}.$$

Erdős proof of infinititude of primes. By Erdős, for every $n \in \mathbf{N}$

$$\sqrt{n} \cdot 2^{\pi(n)} \geq n.$$

To see this inequality, write every $m \in \{1, 2, \ldots, n\}$ as $m = r^2 s$ where $r, s \in \mathbf{N}$ and s is a squarefree number. For r we have at most $\sqrt{n}$ values because $r^2 \leq n$. Every s is a product of distinct primes not exceeding x. Therefore the number of values of s is bounded by the number of subsets of the set $S = \{p \mid p \leq n\}$, which equals $2^{|S|} = 2^{\pi(n)}$. The product $\sqrt{n} \cdot 2^{\pi(n)}$ is

the upper bound on the number of pairs r, s. Since m attains n values and to distinct ms correspond distinct pairs r, s, the number of pairs and thus the upper bound on it are at least n.

The inequality gives

$$\pi(n) \geq \frac{\log n}{2 \log 2}.$$

This is an order of magnitude better than Euclid's bound but still very weak in comparison with truth.

The upper bound of Legendre. Legendre found an inclusion-exclusion type argument showing that $\pi(x) = o(x)$, in fact $\pi(x) = O(x/\log\log x)$, which we now present.

For every $x \geq 1$,

$$\sum_{n \leq x} \frac{1}{n} \geq \int_1^x \frac{dt}{t} = \log x.$$

Thus

$$\prod_{p \leq x} \left(1 - \frac{1}{p}\right)^{-1} = \prod_{p \leq x} \left(1 + \frac{1}{p} + \frac{1}{p^2} + \cdots\right) \geq \sum_{n \leq x} \frac{1}{n} \geq \log x$$

and, for $x \geq 2$,

$$\prod_{p \leq x} \left(1 - \frac{1}{p}\right) \leq \frac{1}{\log x}.$$

For $1 \leq y \leq x$ and $d \in \mathbf{N}$ we denote $[x] = \{1, 2, \ldots, \lfloor x \rfloor\}$, $M(y) = \{p \mid p \leq y\}$, $P(y) = \prod_{p \leq y} p$, $A(y, x) = \{n \in [x] \mid p|n \Rightarrow p > y\}$, and $A_d(x) = \{n \in [x] \mid d|n\}$. By the inclusion-exclusion principle,

$$\begin{aligned} |A(y,x)| &= \left|[x] \setminus \bigcup_{p \in M(y)} A_p(x)\right| = \sum_{I \subset M(y)} (-1)^{|I|} \left|\bigcap_{p \in I} A_p(x)\right| \\ &= \sum_{d | P(y)} \mu(d) \left\lfloor \frac{x}{d} \right\rfloor \end{aligned}$$

where $\mu(\cdot)$ is the Möbius function; $\mu(d) = \mu(p_1 p_2 \ldots p_k) = (-1)^k$ for square-free argument (which is always the case in the sum) and $\mu(d) = 0$ else (but $\mu(1) = 1$). We have used the fact that for $I = \{p_1, \ldots, p_k\}$ and $d = p_1 p_2 \ldots p_k$,

$$\bigcap_{p \in I} A_p(x) = A_d(x).$$

The sum has $2^{\pi(y)} \le 2^y$ summands and $\lfloor x/d \rfloor = x/d + \delta$ with $\delta \in [0,1)$. It folows that

$$|A(y,x)| = x \sum_{d|P(y)} \frac{\mu(d)}{d} + E = x \prod_{p \le y} \left(1 - \frac{1}{p}\right) + E$$

where $|E| \le 2^y$. On the other hand,

$$|A(y,x)| \ge \pi(x) - \pi(y) + 1$$

because $A(y,x)$ contains 1 as well as every prime in the interval $(y,x]$. Combining both relations for $|A(y,x)|$ and using the above estimate for the product over primes, we get the inequality

$$\pi(x) \le \frac{x}{\log y} + E + \pi(y) - 1 \le \frac{x}{\log y} + 2^y + y - 1.$$

Setting $y = \log x$, we get for every $x \ge 3$ the estimate

$$\pi(x) \le \frac{x}{\log\log x} + x^{\log 2} + \log x - 1 = \frac{x(1+o(1))}{\log\log x}$$

($\log 2 = 0.69314\ldots < 1$).

Chebyshev's bounds. These are bounds of the form ($x \ge 2$)

$$\frac{c_1 x}{\log x}(1+o(1)) < \pi(x) < \frac{c_2 x}{\log x}(1+o(1))$$

where $0 < c_1 < c_2$ are constants. Chebyshev proved this in 1851 with constants $c_1 = \log(2^{1/2}3^{1/3}5^{1/5}30^{-1/30}) \approx 0.92129$ and $c_2 = \frac{6}{5}c_1 \approx 1.10555$. We present proofs giving constants $c_1 = \log 2 \approx 0.69314$ and $c_2 = \log 4 = 2c_1 \approx 1.38629$.

The upper bound $\pi(x) \le x(1+o(1))\log 4/\log x$. The upper bound for $\pi(x)$ is derived from that for $\prod_{p\le x} p$ or, in the logarithmic form, for $\sum_{p\le x} \log p$. Suppose that we have proved that

$$\prod_{p \le x} p \le (c+o(1))^x$$

for some constant $c > 1$. Then for $1 \le y \le x$,

$$y^{\pi(x)-\pi(y)} \le \prod_{y < p \le x} p \le (c+o(1))^x$$

and, taking logarithms, we get

$$\pi(x) \le \frac{x(1+o(1))\log c}{\log y} + y.$$

Setting $y = x/(\log x)^2$ we get the desired upper bound with constant $c_2 = \log c$:

$$\pi(x) \le \frac{x(1+o(1))\log c}{\log x}.$$

We show two approaches to obtain the bound $\prod_{p\le x} p \le (c+o(1))^x$. The first one gives $c = 4^2 = 16$. Improving upon it, we show that in fact $\prod_{p\le x} p \le 4^x$ for every $x \ge 1$. Clearly,

$$\prod_{n<p\le 2n} p \le \binom{2n}{n}$$

because the product even divides the number $\binom{2n}{n} = (2n)!/(n!)^2$. Since

$$\binom{2n}{n} \le 4^n$$

(by the binomial expansion of $(1+1)^{2n}$), taking logarithms we get

$$\sum_{n<p\le 2n} \log p \le n \log 4.$$

We cover $(1, x]$ by the intervals $(1, 2], (2, 4], (4, 8], \dots, (2^m, 2^{m+1}]$ where $m \in \mathbf{N}_0$ is such that $2^m \le x < 2^{m+1}$. The sum of $\log p$ over primes in $(2^k, 2^{k+1}]$ is, by the last inequality, $\le 2^k \log 4$. Summing these bounds over all intervals, we get

$$\sum_{p\le x} \log p \le (1+2+4+\cdots+2^m)\log 4 = (2^{m+1}-1)\log 4 < 2x \log 4.$$

This gives $c = 4^2 = 16$ and $c_2 = \log 16 = 2\log 4$.

A beautiful argument of Erdős and Kalmár gets rid of the factor 2 (in the logarithmic form). We prove by induction on n that

$$\prod_{p\le n} p \le 4^n.$$

For $n = 1, 2$ this holds. If $n > 2$ is even then this holds as well because $\prod_{p\le n} p = \prod_{p\le n-1} p \le 4^{n-1} < 4^n$. If $n = 2m+1 > 1$ is odd, we split the product as

$$\prod_{p\le n} p = \prod_{p\le m+1} p \prod_{m+1<p\le 2m+1} p.$$

The first product is $\le 4^{m+1}$ by induction. The second product divides $\binom{2m+1}{m}$ and thus it is at most $\binom{2m+1}{m} \le \frac{1}{2}2^{2m+1} = 4^m$ (use the binomial expansion of $(1+1)^{2m+1}$ and the fact that $\binom{2m+1}{m} = \binom{2m+1}{m+1}$). Alltogether,

$$\prod_{p\le n} p \le 4^{m+1+m} = 4^n.$$

This gives the upper Chebyshev bound with $c_2 = \log 4$.

The lower bound $\pi(x) \ge x(1+o(1))\log 2/\log x$. We derive the lower bound for $\pi(x)$ from that for the number d_n where d_n is the least common multiple of $1, 2, \ldots, n$. How is d_n related to $\pi(n)$? A moment of reflexion reveals that any prime p appears in the decomposition of d_n in the power equal to the highest power of p not exceeding n, that is, as p^a where $p^a \le n$ but $p^{a+1} > n$. Since there are $\pi(n)$ primes in the decomposition, we have

$$d_n \le n^{\pi(n)}.$$

This gives

$$\pi(n) \ge \frac{\log d_n}{\log n}$$

and we clearly need lower bounds of the form $d_n \ge (c+o(1))^n$. Such a bound follows from the fact that $\binom{2n}{n}$ divides d_{2n}, which can be proved by means of the identity $a = \lfloor n/p \rfloor + \lfloor n/p^2 \rfloor + \cdots$ where $a \in \mathbf{N}_0$ is the maximum number for which p^a divides $n!$.

We present an alternative and truly miraculous argument due to Nair [2]. For every $n \in \mathbf{N}$,

$$I = \int_0^1 x^n(1-x)^n\,dx = \int_0^1 \sum_{r=0}^{n}(-1)^r\binom{n}{r}x^{r+n} = \sum_{r=0}^{n}(-1)^r\binom{n}{r}\frac{1}{r+n+1}.$$

The integral I satisfies $0 < I < (1/4)^n$ because the integrand is positive on $(0,1)$ and $x(1-x) \le 1/4$ on $[0,1]$. Also, $d_{2n+1}I \in \mathbf{Z}$ because $r+n+1 \le 2n+1$ and so every $r+n+1$ divides d_{2n+1}. Thus $1 \le d_{2n+1}I < d_{2n+1}/4^n$ and

$$d_{2n+1} > 4^n = 2^{2n+1}/2.$$

Hence $d_n > 2^n/4$ for every $n \geq 1$ and we get the lower Chebyshev bound with $c_1 = \log 2$.

Using the more general integral $\int_0^1 x^{m-1}(1-x)^n \, dx$, Nair [2] proved by similar elementary arguments that $d_n \geq 2^n$ for $n \geq 7$.

1.2 A proof of the PNT

Step 1. An equivalent formulation. Consider the *Chebyshev function*

$$\vartheta(x) = \sum_{p \leq x} \log p.$$

We prove the equivalence

$$\pi(x) = \frac{x + o(x)}{\log x} \iff \vartheta(x) = x + o(x), \ x \to \infty.$$

This follows from the estimate

$$\frac{\vartheta(x)}{\log x} \leq \pi(x) \leq \frac{\vartheta(x)}{\log x}(1 + O(\log\log x/\log x)) + \frac{x}{(\log x)^2}.$$

The lower bound is immediate from $\sum_{p \leq x} \log p \leq \pi(x) \log x$. As for the upper bound, from $\vartheta(x) \geq \sum_{y < p \leq x} \log p \geq (\pi(x) - \pi(y)) \log y$ we get

$$\pi(x) \leq \frac{\vartheta(x)}{\log y} + \pi(y) \leq \frac{\vartheta(x)}{\log y} + y.$$

Setting $y = x/(\log x)^2$, we get the upper bound.

Step 2. Convergence of an integral implies the PNT. We prove the implication

$$\int_1^\infty \frac{\vartheta(x) - x}{x^2} \, dx \text{ converges } \implies \vartheta(x) = x + o(x).$$

Suppose that $\vartheta(x) \neq x + o(x)$. This means that $\limsup \vartheta(x)/x > 1$ or $\liminf \vartheta(x)/x < 1$. We suppose that the first case holds, the second case being handled similarly. There exists an $\lambda > 1$ such that $\forall y > 0 \ \exists x, x > y$,

with $\vartheta(x) > \lambda x$. The integral of our function over the interval $[x, \lambda x]$ is then (we use that $\vartheta(\cdot)$ is nondecreasing)

$$\int_x^{\lambda x} \frac{\vartheta(t) - t}{t^2}\, dt > \int_x^{\lambda x} \frac{\lambda x - t}{t^2}\, dt = \int_1^{\lambda} \frac{\lambda - u}{u^2}\, du = c > 0$$

where the constant c depends only on λ. So

$$\int_1^r \frac{\vartheta(t) - t}{t^2}\, dt$$

does not have finite limit as $r \to \infty$ (Cauchy condition is violated) and the integral does not converge.

Step 3. Laplace transform of $\vartheta(\mathrm{e}^t)\mathrm{e}^{-t} - 1$. We shall work with the complex functions

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} \quad \text{and} \quad F(s) = \sum_p \frac{\log p}{p^s}.$$

They are defined and holomorphic on $\mathrm{Re}(s) > 1$. We prove that for $\mathrm{Re}(z) > 0$

$$\int_0^{\infty} \Big(\frac{\vartheta(\mathrm{e}^t)}{\mathrm{e}^t} - 1 \Big) \mathrm{e}^{-zt}\, dt = \frac{F(z+1)}{z+1} - \frac{1}{z}.$$

It suffices to prove

$$s \int_0^{\infty} \vartheta(\mathrm{e}^t)\mathrm{e}^{-st}\, dt = F(s)$$

for $\mathrm{Re}(s) > 1$; then we set $s = z + 1$ and subtract $\int_0^{\infty} \mathrm{e}^{-tz}\, dt = 1/z$. So

$$\begin{aligned} s \int_0^{\infty} \vartheta(\mathrm{e}^t)\mathrm{e}^{-st}\, dt &= s \int_1^{\infty} \vartheta(x) x^{-s-1}\, dx = \sum_{n=1}^{\infty} \vartheta(n) \cdot s \int_n^{n+1} x^{-s-1}\, dx \\ &= \sum_{n=1}^{\infty} \vartheta(n)(n^{-s} - (n+1)^{-s}) = \sum_{n=1}^{\infty} n^{-s}(\vartheta(n) - \vartheta(n-1)) \\ &= \sum_p \frac{\log p}{p^s} = F(s). \end{aligned}$$

Step 4. Theorem of Wiener and Ikehara. Suppose $f : [0, \infty) \to \mathbf{R}$ is a bounded function that has integral on every bounded interval $[a, b] \subset [0, \infty)$. The Laplace transform

$$g(z) = \int_0^{\infty} f(t)\mathrm{e}^{-tz}\, dt$$

of $f(t)$ is then defined and holomorphic on $\mathrm{Re}(z) > 0$. We prove convergence of the integral in step 2 by means of the following theorem.

Theorem (Wiener and Ikehara, 1932). *If the functions $f(t)$ and $g(z)$ are as before and $g(z)$ has a holomorphic extension to $\mathrm{Re}(z) \geq 0$ (i.e., to an open set containing the right closed halfplane $\mathrm{Re}(z) \geq 0$), then the integral*

$$\int_0^\infty f(t)\, dt$$

converges and equals $g(0)$. (Informally, we may plug in 0 for z in the Laplace transform of $f(t)$.)

We postpone the proof of the theorem until step 9 and first finish the proof of the PNT.

Step 5. Proof of the PNT. We set

$$f(t) = \vartheta(\mathrm{e}^t)\mathrm{e}^{-t} - 1 \quad \text{and} \quad g(z) = F(z+1)(z+1)^{-1} - z^{-1}.$$

Function $f(t)$ is obviously integrable on every bounded interval and it is bounded because $\vartheta(\mathrm{e}^t) = O(\mathrm{e}^t)$ by the upper Chebyshev bound. Function $g(z)$ is a Laplace transform of $f(t)$ by step 3. If we show that $g(z)$ has the required holomorphic extension, we can apply the Wiener–Ikehara theorem and conclude that

$$\int_0^\infty f(t)\, dt = \int_0^\infty \left(\frac{\vartheta(\mathrm{e}^t)}{\mathrm{e}^t} - 1\right) dt = \int_1^\infty \frac{\vartheta(x) - x}{x^2}\, dx$$

converges. By steps 1 and 2 this proves the PNT.

Two things remain to be proved. First, that $g(z) = F(z+1)(z+1)^{-1} - z^{-1}$ has a holomorphic extension to $\mathrm{Re}(z) \geq 0$. Second, we have to prove the Wiener–Ikehara theorem itself. We derive the extension of $g(z)$ from two properties of $\zeta(s) = \sum_{n\geq 1} n^{-s}$.

Step 6. Meromorphic extension of $\zeta(s)$. We show that $\zeta(s) - (s-1)^{-1}$ has a holomorphic extension to $\mathrm{Re}(s) > 0$.

To this end we express the difference as

$$\begin{aligned} \zeta(s) - \frac{1}{s-1} &= \sum_{n=1}^\infty n^{-s} - \int_1^\infty x^{-s} dx = \sum_{n=1}^\infty \int_n^{n+1} (n^{-s} - x^{-s})\, dx \\ &= \sum_{n=1}^\infty F_n(s). \end{aligned}$$

By invoking results on integral representations of functions or simply by calculating the integral defining $F_n(s)$, we obtain that every $F_n(s)$ is an entire function. Further,

$$|F_n(s)| = \left| s \int_n^{n+1} \int_n^x u^{-s-1}\, du\, dx \right| \le |s| \max_{n\le u\le n+1} |u^{-s-1}| = \frac{|s|}{n^{\mathrm{Re}(s)+1}}.$$

Hence the sum $\sum_n F_n(s)$ converges uniformly on $\mathrm{Re}(s) > \varepsilon > 0$ and defines on $\mathrm{Re}(s) > 0$ a holomorphic function extending $\zeta(s) - (s-1)^{-1}$.

In fact, $\zeta(s) - (s-1)^{-1}$ has an extension to entire function.

Step 7. Nonvanishing of $\zeta(s)$. We show that $\zeta(s) \neq 0$ for $\mathrm{Re}(s) \ge 1$.

If $\mathrm{Re}(s) > 1$, we use the Euler identity

$$\zeta(s) = \prod_p \left(1 - \frac{1}{p^s}\right)^{-1}$$

and get the estimate

$$\begin{aligned} |\log(\zeta(s))| &= \left| \sum_p \log\left(1 - \frac{1}{p^s}\right) \right| \le \sum_p \sum_{n\ge 1} |p^{-ns}| \\ &= \sum_p \frac{1}{|p^s| - 1} = \sum_p \frac{1}{p^{\mathrm{Re}(s)} - 1} \\ &\le 2\zeta(\mathrm{Re}(s)). \end{aligned}$$

If $\zeta(s_0) = 0$ for some s_0 with $\mathrm{Re}(s_0) > 1$, then $\mathrm{Re}(\log(\zeta(s))) \to -\infty$ and $|\log(\zeta(s))| \to \infty$ as $s \to s_0$, which contradicts this estimate.

It remains to deal with the line $\mathrm{Re}(s) = 1$. The following argument proving nonvanishing of $\zeta(s)$ on it is due to Mertens in 1898. For $\mathrm{Re}(s) > 1$ Euler's identity gives

$$\begin{aligned} \log|\zeta(s)| &= \mathrm{Re}\Big(\sum_p \log(1 - p^{-s})^{-1} \Big) \\ &= \mathrm{Re}\Big(\sum_p \Big(p^{-s} + \frac{p^{-2s}}{2} + \frac{p^{-3s}}{3} + \cdots \Big) \Big) \\ &= \mathrm{Re}\Big(\sum_n a_n n^{-s} \Big) \end{aligned}$$

where a_n are real nonnegative numbers ($a_n = 1/r$ if $n = p^r$ and $a_n = 0$ else). Thus for the function

$$G(s) = G(u + it) = \zeta(u)^3 \zeta(u + it)^4 \zeta(u + 2it)$$

where $u > 1$ and $t \in \mathbf{R}$ we get, using the identity $\cos(2x) = 2\cos^2(x) - 1$, the estimate

$$\begin{aligned}
\log|G(s)| &= \mathrm{Re}\Big(\sum_{n=1}^{\infty} a_n n^{-u}(3 + 4n^{-it} + n^{-2it})\Big) \\
&= \sum_{n=1}^{\infty} a_n n^{-u}(3 + 4\cos(t\log n) + \cos(2t\log n)) \\
&= \sum_{n=1}^{\infty} 2a_n n^{-u}(1 + \cos(t\log n))^2 \\
&\geq 0.
\end{aligned}$$

Now suppose that $\zeta(1 + it_0) = 0$ and that $1 + it_0$ has as a zero multiplicity $k \in \mathbf{N}$. For $u \to 1^+$ then, for a nonzero $c \in \mathbf{C}$,

$$G(u + it_0) \sim (u-1)^{-3} \cdot c(u-1)^{4k} \cdot \zeta(u + 2it_0) = c\zeta(u + 2it_0)(u-1)^{4k-3} \to 0$$

implying $\log|G(u + it_0)| \to -\infty$, which contradicts the estimate.

Step 8. Holomorphic extension of $F(z+1)/(z+1) - 1/z$. We show that $F(s) - 1/(s-1)$ (recall that $F(s) = \sum_p p^{-s}\log p$) has a holomorphic extension to $\mathrm{Re}(s) \geq 1$. This clearly gives holomorphic extension of

$$\frac{F(z+1)}{z+1} - \frac{1}{z} = \frac{1}{z+1}\Big(F(z+1) - \frac{1}{z} - 1\Big)$$

to $\mathrm{Re}(z) \geq 0$.

Logarithmic derivative of Euler's identity gives ($\mathrm{Re}(s) > 1$)

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

Thus

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

The expression on the right side is holomorphic on $\mathrm{Re}(s) \geq 1$. Indeed, the sum defines a function holomorphic for $\mathrm{Re}(s) > 1/2$ and, by steps 6 and 7, it is clear that $\zeta'(s)/\zeta(s) + 1/(s-1)$ is holomorphic in a neighborhood of every point s with $\mathrm{Re}(s) \geq 1$ with the possible exception $s = 1$. But

in a neighborhood of $s = 1$ we have $\zeta(s) = (s-1)^{-1} + z(s)$ where $z(s)$ is holomorphic on $\mathrm{Re}(s) > 0$ (step 6). So

$$\begin{aligned}\frac{\zeta'(s)}{\zeta(s)} + \frac{1}{s-1} &= \frac{-(s-1)^{-2} + z'(s)}{(s-1)^{-1} + z(s)} + \frac{1}{s-1} \\ &= \frac{1}{s-1}\Big(\frac{-1 + (s-1)^2 z'(s)}{1 + (s-1)z(s)} + 1\Big) \\ &= \frac{z(s) + (s-1)z'(s)}{1 + (s-1)z(s)}\end{aligned}$$

is holomorphic in a neighborhood of $s = 1$ as well.

Step 9. A proof of the Wiener–Ikehara theorem. We assume that, for $\mathrm{Re}(z) > 0$,

$$g(z) = \int_0^\infty f(t)\mathrm{e}^{-tz}\, dt,$$

where the real function $f(t)$ is defined and bounded for $t \geq 0$ (and integrable on bounded intevals), and that $g(z)$ has a holomorphic extension to $\mathrm{Re}(z) \geq 0$. We prove that then

$$\int_0^\infty f(t)\, dt$$

must converge and be equal to $g(0)$. The proof presented here is due to Newman in 1980 ([3]).

For real $T > 0$ we set

$$g_T(z) = \int_0^T f(t)\mathrm{e}^{-zt}\, dt.$$

This is an entire function of z. (Why? Morera's theorem says that if $h : \Omega \to \mathbf{C}$ is continuous on an open set Ω and has zero integral over the boundary of every rectangle lying in Ω, then $h(z)$ is holomorphic on Ω. It is easy to check this condition for $g_T(z)$. Or see [1, chapter 17.2].) We want to prove that $\lim_{T\to\infty} g_T(0) = g(0)$. Let $R > 0$ be real and C be the domain

$$C = C(R) = \{z \in \mathbf{C} : \ |z| < R \ \&\ \mathrm{Re}(z) > -\delta\}$$

where $\delta = \delta(R) > 0$ is so small that $g(z)$ has a holomorphic extension to C (such $\delta > 0$ exists because of the compactness of the segment $[-iR, iR]$). By the Cauchy theorem,

$$g(0) - g_T(0) = \frac{1}{2\pi i}\int_{\partial C} \frac{g(z) - g_T(z)}{z}\, dz$$

where ∂C is the D-shaped boundary curve of C, oriented counterclockwise.

Newman's ingenious trick was to simplify estimation of the last integral by introducing in it an appropriate integration kernel $G(z)$: by the Cauchy theorem, we have also

$$g(0) - g_T(0) = \frac{1}{2\pi i} \int_{\partial C} \frac{g(z) - g_T(z)}{z} G(z)\, dz$$

provided that $G(z)$ is holomorphic on $\overline{C}$ and $G(0) = 1$. We take

$$G(z) = G(z, R, T) = \left(1 + \frac{z^2}{R^2}\right) \mathrm{e}^{zT} ,$$

where $R, T > 0$ are real parameters having the aforementioned meaning. This $G(z)$ is clearly entire and $G(0) = 1$. Its task is to tame the integrand on the circle K given by $|z| = R$; on K we have

$$\left|\frac{G(z)}{z}\right| = \left|\frac{\mathrm{e}^{zT}(z + \overline{z})}{R^2}\right| = 2\mathrm{e}^{\mathrm{Re}(z)T} \cdot \frac{|\mathrm{Re}(z)|}{R^2}.$$

We show that

$$I = \int_{\partial C} \frac{g(z) - g_T(z)}{z} G(z)\, dz \to 0$$

as $T \to \infty$. To this end we split the integral I in three summands. Denoting ∂C^- the arc of the curve ∂C lying in $\mathrm{Re}(z) \le 0$ and similarly for K^- and ∂C^+, we write

$$\begin{aligned} I &= \int_{\partial C^-} \frac{g(z)}{z} G(z)\, dz - \int_{\partial C^-} \frac{g_T(z)}{z} G(z)\, dz + \int_{\partial C^+} \frac{g(z) - g_T(z)}{z} G(z)\, dz \\ &= \int_{\partial C^-} \frac{g(z)}{z} G(z)\, dz - \int_{K^-} \frac{g_T(z)}{z} G(z)\, dz + \int_{\partial C^+} \frac{g(z) - g_T(z)}{z} G(z)\, dz \\ &= I_1 - I_2 + I_3. \end{aligned}$$

In I_2 we could replace ∂C^- with K^- without changing the integral because the integrand is holomorphic in $\mathrm{Re}(z) < 0$. We show that every $I_i \to 0$ as $T \to \infty$.

Step 9.1. Bounding I_2 and I_3.

Step 9.2. Bounding I_1.

1.3 Remarks

Chapter 2

Shnirel'man's theorem

One of the oldest unsolved problems in number theory and in all of mathematics is *Goldbach's conjecture* from 1742. It states that every even number bigger than 2 is a sum of two prime numbers. Assuming that the conjecture holds, it follows (via subtraction of 3) that every odd number bigger than 5 is a sum of three prime numbers. Since 2 and 3 are primes and $5 = 2 + 3$, Goldbach's conjecture has the corollary that every natural number bigger than 1 is a sum of at most three primes. In 1930, L. G. Shnirel'man (1905–1938) proved a weaker form of this corollary with the number of summands bounded by a bigger constant h (his proof gave $h = 8 \cdot 10^5$).

Theorem (Shnirel'man, 1930). *There exists an $h \in \mathbf{N}$ such that every natural number bigger than 1 is a sum of at most h prime numbers.*

We devote this chapter to the proof of Shnirel'man's remarkable result. In section 2.1 we introduce Shnirel'man's density $\sigma(A)$ of subsets $A \subset \mathbf{N}$ and prove the main result that every set A with $\sigma(A) > 0$ is an additive basis, that is, every $n \in \mathbf{N}$ is a sum of a bounded number of summands from A. This by itself does not imply that prime numbers (with 1 added) form an additive basis because they have zero Shnirel'man's density. However, the set $\{1\} \cup 2P$ consisting of 1 and sums of two primes has *positive* Shnirel'man's density. Thus, since $\sigma(\{1\} \cup 2P) > 0$, $\{1\} \cup 2P$ is an additive basis and Shnirel'man's theorem follows (it is an easy matter to get rid of the 1 summands). But how to prove that $\sigma(\{1\} \cup 2P) > 0$? Interestingly, this lower bound follows rather

easily by means of the Cauchy-Schwarz inequality from the upper bound

$$r(n) \ll \frac{n}{(\log n)^2} \prod_{p|n} \left(1 + \frac{1}{p}\right)$$

where $r(n)$ is the number of prime solutions p, q of the equation $n = p + q$. Hence we need to prove this upper bound on $r(n)$. With all due respect to the previous considerations, only now the proof starts at earnest. In section 2.2 we introduce a tool for obtaining such estimates, Selberg's sieve, and apply it to $r(n)$ to get the desired upper bound. In section 2.3 we finish the proof of Shnirel'man's theorem by deducing $\sigma(\{1\} \cup 2P) > 0$ from the upper bound on $r(n)$ In Section 4 we prove a technical lemma needed to complete the proof of the bound in Selberg's sieve.

2.1 Shnirel'man's density

For a set $A \subset \mathbf{N}_0$ and $n \in \mathbf{N}_0$ we let $A(n)$ denote the cardinality $|A \cap \{1, 2, \ldots, n\}|$, so $A(0) = 0$. *Shnirel'man's density* $\sigma(A)$ of A is

$$\sigma(A) = \inf_{n \in \mathbf{N}} \frac{A(n)}{n}.$$

The following properties of $\sigma(A)$ follow easily from the definition.

- $0 \leq \sigma(A) \leq 1$.
- $A(n) \geq \sigma(A)n$ for every $n \geq 0$.
- $\sigma(A) > 0$ if and only if $1 \in A$ and there are constants $c > 0$ and $n_0 \in \mathbf{N}$ such that $A(n)/n > c$ for every $n > n_0$. In particular, $\sigma(A) = 0$ if $1 \notin A$.
- For $A \subset \mathbf{N}$ we have $\sigma(A) = 1$ if and only if $A = \mathbf{N}$.

For $A, B \subset \mathbf{N}_0$ we define

$$A + B = \{a + b : \ a \in A, b \in B\}$$

and similarly for more summands. For $A \subset \mathbf{N}_0$ and $h \in \mathbf{N}$ we write briefly

$$hA = A + A + \cdots + A \ \ (h \text{ summands}).$$

We say that a set $A \subset \mathbf{N}$ is an *additive basis* if there is an $h \in \mathbf{N}$ such that every number $n \in \mathbf{N}$ can be expressed as a sum of at most h elements of A. The smallest such h is then called the *order* of the basis. Equivalently, $A \subset \mathbf{N}$ is an additive basis if there is an $h \in \mathbf{N}$ such that $h(\{0\} \cup A) = \mathbf{N}_0$.

Another simple but important property of $\sigma(A)$ is the following.

- If $A, B \subset \mathbf{N}_0$ satisfy $\sigma(A)+\sigma(B) \geq 1$ and $0 \in A \cap B$ then $A+B = \mathbf{N}_0$. In particular, if $\sigma(A) \geq 1/2$ then $2(\{0\} \cup A) = \mathbf{N}_0$ and A is an additive basis of order at most 2.

Indeed, for $n \in \mathbf{N}_0$ the sum of cardinalities of the sets $A \cap \{0, 1, \ldots, n\}$ and $\{n - b : \ b \in B, 0 \leq b \leq n\}$ is at least $\sigma(A)n + 1 + \sigma(B)n + 1 \geq n + 2$. But these sets are subsets of the $(n + 1)$-element set $\{0, 1, \ldots, n\}$ and must therefore intersect, which gives representation $n = a + b$ with $a \in A$, $b \in B$.

Shnirel'man discovered that, more generally, A is an additive basis whenever $\sigma(A) > 0$. This follows from an inequality (also due to him) relating $\sigma(A + B)$ to $\sigma(A)$ and $\sigma(B)$.

Lemma. *If $A, B \subset \mathbf{N}_0$ satisfy $0 \in A \cap B$, then*

$$\sigma(A + B) \geq \sigma(A) + \sigma(B) - \sigma(A)\sigma(B).$$

Proof. Let $n \in \mathbf{N}_0$ be arbitrary and $0 = a_0 < a_1 < \ldots < a_k \leq n$ be the elements of the set $A \cap \{0, 1, \ldots, n\}$. The elements

$$\{a_i : \ 1 \leq i \leq k\}, \ \{a_i + b : \ 0 \leq i \leq k - 1, b \in B, 1 \leq b \leq a_{i+1} - a_i - 1\}$$

and

$$\{a_k + b : \ b \in B, 1 \leq b \leq n - a_k\}$$

lie in $(A+B) \cap \{1, 2, \ldots, n\}$ and are all mutually distinct. Denoting $\sigma(A) = \alpha$ and $\sigma(B) = \beta$, this implies

$$\begin{aligned}
(A + B)(n) &\geq A(n) + \sum_{i=0}^{k-1} B(a_{i+1} - a_i - 1) + B(n - a_k) \\
&\geq A(n) + \beta \sum_{i=0}^{k-1} (a_{i+1} - a_i - 1) + \beta(n - a_k) \\
&= A(n) + \beta(n - k) = A(n) - \beta A(n) + \beta n \geq (1 - \beta)\alpha n + \beta n \\
&= (\alpha + \beta - \alpha\beta)n.
\end{aligned}$$

So $\sigma(A+B) \geq \alpha + \beta - \alpha\beta$. □

Theorem (Shnirel'man, 1930). *Every set $A \subset \mathbf{N}$ with $\sigma(A) > 0$ is an additive basis.*

Proof. We rewrite the inequality in the lemma as

$$1 - \sigma(A+B) \leq (1-\sigma(A))(1-\sigma(B)).$$

Iterating, we get $1-\sigma(hA) \leq (1-\sigma(A))^h$. Since $\sigma(A) > 0$, we have $\sigma(hA) \geq 1/2$ for sufficiently big h. Then, by the above observation, $2(\{0\} \cup hA) = \mathbf{N}_0$ and A is an additive basis of order at most $2h$. □

2.2 Bounding $r(n)$ via Selberg's sieve

We prove the promised upper bound on the number $r(n)$ counting expressions of n as a sum of two primes. First we state and prove, modulo one technical lemma, an upper bound due to A. Selberg (1917) that applies in rather general situations. Using this *Selberg's sieve*, we obtain the upper bound for $r(n)$.

We have a finite set (or a finite sequence with possible repetitions) $A \subset \mathbf{N}$ and a real number $z > 0$ and we want to estimate from above the quantity

$$S(A,z) = |\{a \in A \mid p|a \Rightarrow p \geq z\}|$$

that counts numbers in A having no prime factor smaller than z. For $d \in \mathbf{N}$ we denote

$$A_d = |\{a \in A \mid d|a\}|,$$

the number of multiples of d in A. Suppose that $g : \mathbf{N} \to (0,1]$ is a completely multiplicative function, which means that $g(1) = 1$ and $g(ab) = g(a)g(b)$ for every $a, b \in \mathbf{N}$, that moreover satisfies $0 < g(n) < 1$ for $n > 1$. We define $r_d \in \mathbf{Z}$ by

$$A_d = g(d)|A| + r_d.$$

If the quantities $g(d)|A|$ approximate A_d well, that is, the errors r_d are small, at least in average, one has a good upper bound on $S(A,z)$:

Theorem (Selberg, 1947). *In the above notation,*

$$S(A,z) < \frac{|A|}{\sum_{k<z} g(k)} + \sum_{\mu(f)\neq 0, f<z^2} 3^{\omega(f)}|r_f.$$

Proof. We rewrite the definition of $S(A,z)$ so that it does not mention prime numbers. Let $D = D(z)$ be the set of squarefree natural numbers smaller than z. Then

$$S(A,z) = |\{a \in A \mid \forall d \in D:\ (a,d) = 1\}.$$

Hence, if λ_d for $d \in D$ are any real numbers satisfying only the condition $\lambda_1 = 1$, we have the bound

$$S(A,z) \le \sum_{a\in A} \Big(\sum_{d\in D, d|a} \lambda_d \Big)^2 = \sum_{a\in A} \Big(\sum_{d\in D, d|a} \lambda_d \Big) \Big(\sum_{e\in D, d|a} \lambda_d \Big)$$

because each a in A counted by $S(A,z)$ contributes 1 and each remaining a contributes nonnegative amount. Changing the summation order, we get (recall that $[e,d]$ is the smallest common multiple of e and d)

$$\begin{aligned} S(A,z) &\le \sum_{e,d\in D} \lambda_e \lambda_d \sum_{a\in A, [e,d]|a} 1 \\ &= \sum_{e,d\in D} \lambda_e \lambda_d (g([e,d])|A| + r_{[e,d]}) \\ &= |A| \sum_{e,d\in D} g([e,d]) \lambda_e \lambda_d + \sum_{e,d\in D} r_{[e,d]} \lambda_e \lambda_d. \end{aligned}$$

We see that values of the quadratic form

$$G(x_d:\ d \in D) = \sum_{e,d\in D} g([e,d]) x_e x_d = \sum_{e,d\in D} \frac{g(e)g(d)x_e x_d}{g((e,d))}$$

(we have used $[e,d](e,d) = ed$ and the multiplicativity of g) play an important role.

Lemma. *For the given completely multiplicative function g and the set of squarefree numbers $D = D(z)$, there exist real numbers λ_d^* such that $\lambda_1^* = 1$, $|\lambda_d^*| \le 1$ for all $d \in D$, and*

$$G(\lambda_d^*:\ d \in D) < \frac{1}{\sum_{k<z} g(k)}.$$

We defer proof of the lemma in Section 2.4. With the selection $\lambda_d = \lambda_d^*$, the upper bound on $S(A,z)$ turns in

$$S(A,z) < \frac{|A|}{\sum_{k<z} g(k)} + \sum_{e,d\in D} |r_{[e,d]}.$$

In order to bound the last sum, we consider the equality $f = [e,d]$ where $e,d \in D$. Since $e,d < z$ are squarefree, so is f and $f < z^2$. If f is fixed and $f = p_1p_2\dots p_k$ for k distinct primes, the number of pairs $e,d \in \mathbf{N}$ (we may drop the condition $e,d \in D$ since we are proving an upper bound) for which $f = [e,d]$ equals to the number of ways to write $\{1,2,\dots,k\}$ as $A \cup B$ for a pair of sets A,B. These pairs are in 1-1 correspondence with colorings of $\{1,2,\dots,k\}$ by three colors—one color for the elements in $A\backslash B$, another color for the elements in $B\backslash A$, and the third color for the elements in $A\cap B$—and are therefore counted by $3^k = 3^{\omega(f)}$. Summing over squarefree $f < z^2$, we get

$$\sum_{e,d\in D} |r_{[e,d]}| \le \sum_{\mu(f)\neq 0, f<z^2} 3^{\omega(f)}|r_f.$$

□

Now we bound $r(n)$. Let $n \in \mathbf{N}$ be even. We set

$$A = (m(n-m) :\ 1 \le m \le n-1)$$

and $z = n^{1/8}$. It follows that $r(n)$ is the number of terms a of the sequence A with $\Omega(n) = 2$ and that

$$r(n) \le 2z + S(A,z) = 2n^{1/8} + S(A, n^{1/8}).$$

For a squarefree $d \in \mathbf{N}$, A_d is the number of solutions of the congruence

$$m(n-m) \equiv 0 \pmod{d}$$

in the set $1 \le m \le n-1$. For $d = p_1\dots p_a q_1 \dots q_b$, where the primes p_i divide n and the primes q_i do not divide n, the congruence has 2^b solutions in congruence classes modulo d. Indeed, by the Chinese remainder theorem, the congruence is equivalent to the system

$$m \equiv 0 \vee m \equiv n \pmod{p} \quad p \in \{p_1,\dots,p_a,q_1,\dots,q_b\}$$

where for $p = p_i$ the two corresponding conguences coincide while for $p = q_i$ they are distinct (and have no common solution) and the 2^b possible selections of the right sides in the system 1-1 correspond to 2^b solutions of the original congruence. Fixing a congruence class c modulo d, we have at least $\lfloor (n-1)/d \rfloor$ and at most $\lceil (n-1)/d \rceil$ solutions to $m \equiv c \pmod{d}$ in the set $1 \le m \le n-1$. We see that

$$2^b \left\lfloor \frac{n-1}{d} \right\rfloor \le A_d \le 2^b \left\lceil \frac{n-1}{d} \right\rceil .$$

We set, for any $d \in \mathbf{N}$,

$$g(d) = \frac{2^b}{d}$$

where b is the sum of exponents in the prime decomposition of d of those prime factors that do not divide n. The above estimate of A_d shows that for squarefree d the error r_d in

$$A_d = g(d)|A| + r_d = g(d)(n-1) + r_d$$

satisfies

$$|r_d| \le 2^b \le 2^{\omega(d)}.$$

It is easy to see that $g(d)$ is completely multiplicative and that $0 < g(d) < 1$ for $d > 1$ because n is even.

Theorem. *There is a constant $c > 0$ such that for every $n \ge 2$ we have*

$$r(n) = \#(\textit{solutions to } n = p + q) < \frac{cn}{(\log n)^2} \prod_{p \mid n} \Big(1 + \frac{1}{p}\Big).$$

Proof. By Selberg's sieve, we have ($z = n^{1/8}$)

$$r(n) \le 2n^{1/8} + S(A, n^{1/8}) \le 2n^{1/8} + \frac{n-1}{\sum_{k<z} g(k)} + \sum_{\mu(k)\neq 0, k<z^2} 3^{\omega(k)} |r_k|.$$

As for the last sum, the above estimate on $|r_d|$ shows that

$$\begin{aligned}
\sum_{\mu(k)\neq 0, k<z^2} 3^{\omega(k)} |r_k| &\le \sum_{\mu(k)\neq 0, k<z^2} 6^{\omega(k)} = \sum_{\mu(k)\neq 0, k<z^2} \left(2^{\omega(k)}\right)^{\log_2 6} \\
&\le \sum_{\mu(k)\neq 0, k<z^2} k^{\log_2 6} \quad (2^{\omega(k)} \le k \text{ for squarefree } k) \\
&\le z^2 \cdot z^{2\log_2 6} = n^{(2+2\log_2 6)/8} \\
&< n^{9/10}.
\end{aligned}$$

Thus the last sum of errors $|r_k|$ is negligible.

We need a lower bound on $\sum_{k<z} g(k)$. Let $d(k)$ be the number of divisors of k, $d_n(k)$ be the number of divisors of k that are coprime with n, and P_n be the set of numbers composed only of primes dividing n:

$$\prod_{p|n}\left(1-\frac{1}{p}\right)^{-1} = \sum_{k\in P_n}\frac{1}{k}.$$

Let s_i be the exponents in the prime decomposition of k of the prime factors not dividing n. Then

$$g(k) = \frac{2^{s_1+s_2+\cdots}}{k} \geq \frac{\prod(s_i+1)}{k} = \frac{d_n(k)}{k}.$$

We have

$$\begin{aligned}
\frac{\sum_{k<z} g(k)}{\prod_{p|n}(1-1/p)} &\geq \sum_{k<z}\frac{d_n(k)}{k}\sum_{l\in P_n}\frac{1}{l} \\
&= \sum_{k<z} d_n(k) \sum_{k|t,\ t/k\in P_n}\frac{1}{t} \\
&= \sum_{t=1}^{\infty}\frac{1}{t}\sum_{k<z,\ k|t,\ t/k\in P_n} d_n(k) \\
&\geq \sum_{t<z}\frac{1}{t}\sum_{k|t,\ t/k\in P_n} d_n(k).
\end{aligned}$$

To evaluate the inner sum, we split $t = t_1t_2$ where $t_1 \in P_n$ and $(t_2, n) = 1$. Then k runs through the numbers it_2 with $i|t_1$ and the sum contains $d(t_1)$ summands, each of which equals $d(t_2)$. Thus the inner sum equals $d(t_1)d(t_2) = d(t)$ (since $(t_1, t_2) = 1$ and $d(\cdot)$ is multiplicative). So

$$\frac{|A|}{\sum_{k<z} g(k)} < n\cdot\frac{\prod_{p|n}(1-1/p)^{-1}}{\sum_{t<z} d(t)/t}.$$

Replacing $(1-1/p)^{-1}$ with $1+1/p$ in the product costs only a constant factor:

$$\prod_{p|n}\left(1-\frac{1}{p}\right)^{-1} = \prod_{p|n}\left(1+\frac{1}{p}\right)\prod_{p|n}\left(1-\frac{1}{p^2-1}\right) \ll \prod_{p|n}\left(1+\frac{1}{p}\right)$$

because the product $\prod_p(1+(p^2-1)^{-1})$ converges. Dirichlet's asymptotics in the divisor problem implies that

$$D(x)=\sum_{t<x} d(t) = x\log x + O(x).$$

This by Abel's partial summation gives

$$\begin{aligned}\sum_{t<z}\frac{d(t)}{t} &= \frac{D(z)}{z} - \int_1^z D(t)(1/t)'\,dt \\ &= \frac{(\log z)^2}{2} + O(\log z) \\ &= \frac{(\log n)^2}{128} + O(\log n).\end{aligned}$$

We conclude that

$$r(n) \le 2n^{1/8} + S(A, n^{1/8}) \le \frac{|A|}{\sum_{k<z} g(k)} + O(n^{9/10}) \ll \frac{n}{(\log n)^2}\prod_{p|n}\Big(1+\frac{1}{p}\Big).$$

□

2.3 Conclusion of the proof

Let $X = \{1\}\cup 2P$ be the set consisting of 1 and sums of two primes. Using the upper bound on $r(n) = \#$ (solutions of $p+q=n$) obtained in the previous section, we prove that $X(n) = |X\cap\{1,2,\dots,n\}| > cn$ for all $n\in\mathbf{N}$ for an absolute constant $c>0$. This gives $\sigma(X)>0$. Assuming this result, the theorem in section 2.1 tells us that X is an additive basis of order at most h. Thus every $n\in\mathbf{N}$ is a sum of at most $2h$ summands, each of which is 1 or a prime number. To get rid of the 1 summands for $n>1$, we take $n>2$ (for $n=2$ Shnirel'man's theorem holds) and write

$$n = 2 + (n-2) = 2 + a_1 + \cdots + a_k$$

where $a_i\in\{1\}\cup P$ and $k\le 2h$. If no $a_i = 1$, n is expressed as a sum of $k+1$ primes. If $a_i=1$ for exactly one i, we delete this a_i and the first 2 and replace them by 3. If $a_i=1$ for more than one i, we delete all 1's and replace them by smaller number of 2's and 3's with the same sum (convince

yourself that this is possible). In all cases we obtain an expression of n as a sum of at most $k+1 \le 2h+1$ primes.

Now we obtain the desired lower bound on $X(n)$. We prove that $(2P)(n) > cn$ for every $n \ge 4$ for a constant $c > 0$. Let $n \ge 4$ be arbitrary. By the Cauchy–Schwarz inequality,

$$\begin{aligned}\Big(\sum_{m=4}^{n} r(m)\Big)^2 &\le \Big(\sum_{m=4, r(m)>0}^{n} 1^2\Big)\Big(\sum_{m=4}^{n} r(m)^2\Big) \\ &= (2P)(n)\cdot\Big(\sum_{m=4}^{n} r(m)^2\Big)\end{aligned}$$

and we have

$$X(n) > (2P)(n) \ge \frac{(\sum_{m=4}^{n} r(m))^2}{\sum_{m=4}^{n} r(m)^2}.$$

The sum in the numerator counts all pairs of primes p, q such that $p+q \le n$ and therefore, for $n \ge 4$,

$$\sum_{m=4}^{n} r(m) \ge \pi(n/2)^2 \gg n^2/(\log n)^2.$$

By the upper bound on $r(m)$ proved in the theorem in section 2.3, the sum in the denominator satisfies

$$\sum_{m=4}^{n} r(m)^2 \ll \sum_{m=4}^{n} \frac{m^2}{(\log m)^4} \prod_{p|m}\Big(1+\frac{1}{p}\Big)^2 \le \frac{n^2}{(\log n)^4} \sum_{m=4}^{n} \prod_{p|m}\Big(1+\frac{1}{p}\Big)^2.$$

As for the last sum,

$$\begin{aligned}\sum_{m=4}^{n} \prod_{p|m}\Big(1+\frac{1}{p}\Big)^2 &\le \sum_{m=4}^{n}\Big(\sum_{d|m}\frac{1}{d}\Big)^2 = \sum_{m=4}^{n}\Big(\sum_{d|m}\frac{1}{d}\Big)\Big(\sum_{e|m}\frac{1}{e}\Big) \\ &= \sum_{d,e}\frac{1}{de}\sum_{m=4, d|m, e|m}^{n} 1 \le n\sum_{d,e}\frac{1}{de[d,e]} \\ &\le n\sum_{d,e}\frac{1}{(de)^{3/2}} \quad (\text{since } [d,e] \ge \max(d,e) \ge \sqrt{de}) \\ &= n\Big(\sum_{d=1}^{\infty}\frac{1}{d^{3/2}}\Big)^2 \ll n.\end{aligned}$$

Thus, for $n \ge 4$,

$$(2P)(n) \gg \frac{(n^2/(\log n)^2)^2}{n^3/(\log n)^4} \gg n.$$

This finishes the proof of Shnirel'man's theorem.

2.4 Lemma on numbers λ_d^*

It remains to prove the lemma on numbers λ_d^* used in derivation of Selberg's sieve. Recall that $z > 0$,

$$D = D(z) = \{n \in \mathbf{N} \mid \mu(n) \neq 0, n < z\}$$

is the set of squarefree numbers smaler than z,

$$g : \mathbf{N} \to (0, 1]$$

is a completely multiplicative function satisfying $g(1) = 1$ and $0 < g(n) < 1$ for $n > 1$, and G is a quadratic form in $|D|$ variables λ_d corresponding to the elements $d \in D$, defined by

$$G(\lambda_d : \ d \in D) = \sum_{d_1, d_2 \in D} \frac{g(d_1)\lambda_{d_1} g(d_2)\lambda_{d_2}}{g((d_1, d_2))}.$$

For $l \in \mathbf{N}$ we define

$$f(l) = \sum_{d|l} \frac{\mu(d)}{g(l/d)} = \frac{1}{g(l)} \sum_{d|l} \mu(d)g(d) = \frac{1}{g(l)} \prod_{p|l} (1 - g(p)) > 0.$$

Note that $f(l)$ is multiplicative ($f(ab) = f(a)f(b)$ whenever $(a, b) = 1$) and that by Möbius inversion formula

$$\frac{1}{g(k)} = \sum_{d|k} f(d).$$

For $d \in D$ we set

$$\alpha_d = \sum_{l, \, dl \in D} \frac{1}{f(l)}.$$

In the next lemma we define the numbers λ_d^* and in the parts 1, 3, and 4 we prove all their required properties. In the proof we use the notation, for a condition C,

$$\langle C \rangle = \begin{cases} 1 & C \text{ holds} \\ 0 & C \text{ does not hold.} \end{cases}$$

Lemma. *For $d \in D$, let*

$$\lambda_d^* = \frac{\mu(d)\alpha_d}{f(d)g(d)\alpha_1}.$$

Then

1. $\lambda_1^* = 1$ *and* $|\lambda_d^*| \le 1$ *for every* $d \in D$.

2. *For every* $k \in D$ *we have* $\sum_{d\in D}\langle k|d\rangle \cdot g(d)\lambda_d^* = \mu(k)/(\alpha_1 f(k))$.

3. $G(\lambda_d^* :\ d \in D) = (\alpha_1)^{-1}$.

4. $\alpha_1 = \sum_{k\in D} f(k)^{-1} \ge \sum_{k<z} g(k)$.

Proof. **1.** Since $\mu(1) = g(1) = f(1) = 1$, $\lambda_1^* = 1$. For any $d \in D$ we have

$$\begin{aligned}
\alpha_1 &= \sum_{k\in D}\frac{1}{f(k)} = \sum_l \langle l|d\rangle \sum_{k\in D}\frac{\langle (k,d) = l\rangle}{f(k)} \\
&= \sum_l \frac{\langle l|d\rangle}{f(l)} \sum_m \frac{\langle ml \in D \ \&\ (m, d/l) = 1\rangle}{f(m)} \\
&\ge \sum_l \frac{\langle l|d\rangle}{f(l)} \sum_m \frac{\langle md \in D\rangle}{f(m)} = \sum_m \frac{\langle md \in D\rangle}{f(m)} \sum_l \frac{\langle l|d\rangle}{f(l)} \\
&= \frac{\alpha_d}{f(d)} \sum_l \langle l|d\rangle f(d/l) \\
&= \frac{\alpha_d}{f(d)g(d)}
\end{aligned}$$

where in the last transformation we used the above expression for $1/g(k)$. Thus $|\lambda_d^*| = \alpha_d/(f(d)g(d)\alpha_1) \le 1$.
2. Let $k \in D$. Then

$$\begin{aligned}
\sum_{d\in D}\langle k|d\rangle g(d)\lambda_d^* &= \sum_{d\in D}\langle k|d\rangle g(d)\frac{\mu(d)\alpha_d}{f(d)g(d)\alpha_1} \\
&= \frac{1}{\alpha_1}\sum_l \langle kl \in D\rangle \frac{\mu(kl)\alpha_{kl}}{f(kl)} \\
&= \frac{\mu(k)}{\alpha_1 f(k)} \sum_l \langle kl \in D\rangle \frac{\mu(l)}{f(l)} \sum_m \frac{\langle klm \in D\rangle}{f(m)} \\
&= \frac{\mu(k)}{\alpha_1 f(k)} \sum_l \langle kl \in D\rangle \mu(l) \sum_m \frac{\langle klm \in D\rangle}{f(lm)} \\
&= \frac{\mu(k)}{\alpha_1 f(k)} \sum_n \frac{\langle kn \in D\rangle}{f(n)} \sum_l \langle l|n\rangle \mu(l) \\
&= \frac{\mu(k)}{\alpha_1 f(k)} \quad \text{(by the property } \textstyle\sum_{l|n}\mu(l) = \delta_{1,n}\text{)}.
\end{aligned}$$

3.

$$\begin{aligned} G(\lambda_d :\ d \in D) &= \sum_{d_i \in D} \frac{g(d_1)\lambda_{d_1} g(d_2)\lambda_{d_2}}{g((d_1,d_2))} \\ &= \sum_{d_i \in D} \sum_{k|d_i} f(k) g(d_1)\lambda_{d_1} g(d_2)\lambda_{d_2} \text{ (above expression for } 1/g(k)) \\ &= \sum_{k \in D} f(k) \sum_{d_i \in D} \langle k|d_1 \ \&\ k|d_2\rangle g(d_1)\lambda_{d_1} g(d_2)\lambda_{d_2} \\ &= \sum_{k \in D} f(k) \Big(\sum_{d \in D} \langle k|d\rangle g(d)\lambda_d \Big)^2. \end{aligned}$$

For $\lambda_d = \lambda_d^*$ we get, using part 2,

$$G(\lambda_d^* :\ d \in D) = \sum_{k \in D} f(k) \left(\frac{\mu(k)}{\alpha_1 f(k)} \right)^2 = \frac{1}{\alpha_1^2} \sum_{k \in D} \frac{1}{f(k)} = \frac{1}{\alpha_1}.$$

4.

$$\begin{aligned} \alpha_1 = \sum_{k \in D} \frac{1}{f(k)} &= \sum_{k \in D} g(k) \prod_{p|k} (1 - g(p))^{-1} \\ &= \sum_{k \in D} g(k) \prod_{p|k} (1 + g(p) + g(p^2) + \cdots) \\ &= \sum_{k \in D} g(k) \sum_{l} \langle p|l \Rightarrow p|k\rangle g(l) \\ &= \sum_{k,l} \langle k \in D \ \&\ (p|l \Rightarrow p|k)\rangle g(kl) \\ &= \sum_{m} g(m) \sum_{k} \langle k \in D \ \&\ k|m \ \&\ (p|(m/k) \Rightarrow p|k)\rangle \\ &\geq \sum_{m<z} g(m) \end{aligned}$$

because for $m < z$ the last inner sum is always ≥ 1 (set k to be the product of all prime factors of m). □

2.5 Remarks

Chapter 3

Roth's theorem on arithmetic progressions

Two famous theorems were proved by Klaus Roth (1925). The first one ([7]), which weighted most in awarding him the Fields medail in 1958, asserts that for every real irrational algebraic number α and every $\varepsilon > 0$ the inequality

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{q^{2+\varepsilon}}$$

has only finitely many solutions in rational numbers $p/q \in \mathbf{Q}$. Here we will be interested in other Roth's theorem that concerns arithmetic progressions of length 3. Let $r_3(n)$ be the maximum size of a subset $A \subset [n] = \{1, 2, \ldots, n\}$ which contains no proper arithmetic progression of length 3, that is, there does not exist a triple $a, a+d, a+2d$ of elements in A with $d > 0$. In 1952 Roth proved in [5] (see also [6]) that $r_3(n)/n \to 0$ as $n \to \infty$. An equivalent formulation of this result is as follows.

Theorem 3.0.1 (Roth 1952) *For every $\delta > 0$ there is an $n_0 \in \mathbf{N}$ such that if $n > n_0$ and $A \subset [n]$ is any set with $|A| > \delta n$ elements, then A must contain a proper arithmetic progression of length* 3.

In Sections 3.2 and 3.3 we give an analytic proof of Roth's theorem using so called *circle method*. In Section 3.4 and 3.5 we give a completely different combinatorial proof by means of extremal graph theory. More comments and references on Roth's theorem follow in Section 3.5.

3.1 An analytic proof

For real t we consider the function ($i = \sqrt{-1}$ is the imaginary unit)

$$e(t) = \exp(2\pi i t)$$

that maps $\mathbf{R}$ to the unit circle in $\mathbf{C}$. Properties of the exponential function tell us that

$$e(t)e(u) = e(t+u),\ e(t)^m = e(mt) \text{ for } m \in \mathbf{Z}, \text{ and } \overline{e(t)} = e(-t) = 1/e(t).$$

For integers $m \in \mathbf{Z}$ we have the fundamental identity

$$\int_0^1 e(mt)\ dt = \langle m = 0 \rangle = \begin{cases} 1 & \text{if } m = 0 \\ 0 & \text{if } m \neq 0 \end{cases}$$

because $m = 0$ gives $\int_0^1 e(0)\ dt = \int_0^1 1\ dt = 1$ and $m \neq 0$ gives

$$\int_0^1 e(mt)\ dt = \left[\frac{1}{2\pi i m} \exp(2\pi i m t)\right]_0^1 = \frac{1-1}{2\pi i m} = 0.$$

Let $f, g \in \mathbf{C}[z, z^{-1}]$ be two Laurent polynomials, that is to say

$$f(z) = \sum_{k \in \mathbf{Z}} a_k z^k \quad \text{and} \quad g(z) = \sum_{k \in \mathbf{Z}} b_k z^k$$

with $a_k, b_k \in \mathbf{C}$ and only finitely many nonzero a_ks and b_ks. The properties of $e(t)$ imply that

$$\int_0^1 f(e(t))\overline{g(e(t))}\ dt = \sum_{j,k} a_j \overline{b_k} \int_0^1 e((j-k)t)\ dt = \sum_j a_j \overline{b_j}.$$

For $f(z) = g(z)$ we get the *Parseval identity*

$$\int_0^1 |f(e(t))|^2\ dt = \sum_j |a_j|^2.$$

Finally, we have the Cauchy–Schwarz inequality

$$\int_0^1 |f(e(t))g(e(t))|\ dt \leq \left(\int_0^1 |f(e(t))|^2\ dt\right)^{1/2} \left(\int_0^1 |g(e(t))|^2\ dt\right)^{1/2},$$

which of course holds in far more general situations than just for Laurent polynomials.

For a set $A \subset [n]$ we denote

$$f_A(z) = \sum_{a \in A} z^a = \sum_{k=1}^{n} \langle k \in A \rangle z^k$$

and

$$p_3(A) = \#\{(a, a+d, a+2d) \in A^3 \mid d \geq 0\}.$$

The function $p_3(A)$ counts arithmetical progressions of length 3 in A, including the degenerate ones (a, a, a) with $d = 0$. Thus

$$p_3(A) = |A| + \#(\text{proper APs of length 3 in } A)$$

and $p_3(A) = |A| \leq n$ if A contains no proper AP of length 3. Notice that the parametrization $x = a, y = a + 2d, z = a + d$ gives equivalent formula

$$p_3(A) = \#\{(x, y, z) \in A^3 \mid x + y = 2z\}.$$

This and the fundamental identity give the integral reprezentation

$$p_3(A) = \int_0^1 f_A(e(t))^2 \cdot f_A(e(-2t))\, dt.$$

To handle this integral, we return to combinatorics. The following result (we leave its proof as an exercise) is sometimes called the *Fekete lemma*.

Lemma. *Let $(a_n)_{n \geq 1}$ be a sequence of nonnegative real numbers satisfying $a_{m+n} \leq a_m + a_n$ for all $m, n \geq 1$ or $a_{m+n} \geq a_m + a_n$ for all $m, n \geq 1$. Then the limit*

$$L = \lim_{n \to \infty} \frac{a_n}{n}$$

exists (in the former case it may be $L = \infty$). Moreover, $a_n/n \geq L$ for all $n \geq 1$ in the former case and $a_n/n \leq L$ for all $n \geq 1$ in the latter case.

It is easy to see that

$$r_3(m+n) \leq r_3(m) + r_3(n)$$

for all $m, n \geq 1$: If $A \subset [m+n]$ contains no AP of length 3 and $|A| = r_3(m+n)$, then the sets $A \cap [n]$ and $A \cap [m+1, m+n]$ contain no AP

of length 3 and their sizes, which sum up to $r_3(m+n)$, are thus bounded by $r_3(m)$ and $r_3(n)$, respectively (if the latter set is shifted to lie in $[n]$, no arithmetical progression in it is created). By Fekete lemma, there exists a constant

$$d_3 = \lim_{n\to\infty} \frac{r_3(n)}{n} \in [0,1]$$

and $r_3(n) \geq d_3 n$ for all $n \geq 1$. Roth's theorem amounts to proving that $d_3 = 0$.

We have used certain combinatorial properties of sets free of APs of length 3 and it will be useful to state them explicitly. If A contains no AP of length 3, then

- any subset $B \subset A$ contains no AP of length 3 as well and
- any affine image $\alpha A + \beta = \{\alpha a + \beta \mid a \in A\}$, where $\alpha \in \mathbf{Q}$ and $\beta \in \mathbf{Z}$, contains no AP of length 3 as well.

These two properties are crutial for dealing with $r_3(n)$. They are valid for the class of sets free of APs of length $k \geq$ too.

Let $A_n \subset [n]$ be a set witnessing $r_3(n)$ (A contains no AP of length 3 and has size $r_3(n)$) and let

$$g_n(z) = d_3 z + d_3 z^2 + \cdots + d_3 z^n.$$

Then

$$f_{A_n}(1) - g_n(1) = |A_n| - d_3 n = r_3(n) - d_3 n = o(n) \text{ as } n \to \infty.$$

The key step is to prove that this remains true when 1 is replaced with any z on the unit circle $|z| = 1$.

Proposition 3.1.1 *In the above notation, for $n \to \infty$ we have, uniformly on the unit circle $|z| = 1$, that*

$$f_{A_n}(z) - g_n(z) = o(n).$$

This means that for every $\varepsilon > 0$ there is an $n_0 = n_0(\varepsilon)$ such that for every $n > n_0$ we have

$$\max_{|z|=1} |f_{A_n}(z) - g_n(z)| < \varepsilon n.$$

We defer the proof in the next section. Note that one has the weaker trivial bound

$$\max_{|z|=1} f_{A_n}(z) - g_n(z)| \leq \max(d_3, 1-d_3)n.$$

Using the Proposition, we conclude the proof of Roth's theorem. We take a sequence of sets $A_n \subset [n]$ as before (i.e., witnessing the values $r_3(n)$), set

$$f(z) = f_{A_n}(z) = \sum_{a \in A_n} z^a \quad \text{and} \quad g(z) = g_n(z) = d_3 \sum_{k=1}^{n} z^k,$$

and write

$$f(z) = g(z) + (f(z) - g(z)) = g(z) + h(z).$$

By the integral formula for $p_3(A)$,

$$\begin{aligned} n \geq |A_n| = p_3(A_n) &= \int_0^1 f(e(t))^2 \cdot f(e(-2t))\, dt \\ &= \int_0^1 (g(e(t)) + h(e(t)))^2 \cdot (g(e(-2t)) + h(e(-2t)))\, dt \\ &= \int_0^1 g(e(t))^2 \cdot g(e(-2t))\, dt + \text{ seven integrals} \\ &= d_3^3 \int_0^1 g_0(e(t))^2 \cdot g_0(e(-2t))\, dt + \text{ seven integrals} \end{aligned}$$

where $g_0(z) = z + z^2 + \cdots + z^n$ and each of the seven integrals has the form

$$\int_0^1 a(e(t)) \cdot b(e(t)) \cdot c(e(-2t))\, dt$$

with $a(z), b(z), c(z) \in \{g(z), h(z)\}$ and at least one of $a(z), b(z), c(z)$ equal to $h(z)$.

Since $g_0(z) = z + z^2 + \cdots + z^n = f_{[n]}(z)$, the first integral is equal to

$$\begin{aligned} p_3([n]) &= n + (n-2) + (n-4) + \cdots + (n - 2\lfloor n/2 \rfloor) \\ &= \lfloor n/2 \rfloor (\lceil n/2 \rceil - 1) + n \\ &= n^2/4 + O(n). \end{aligned}$$

We show that each of the remaining seven integrals is $o(n^2)$. We demostrate it on the case $b(z) = h(z)$, the other cases are virtually identical. By the

Proposition, $h(e(t)) = o(n)$ uniformly in $t \in [0,1]$ and so

$$\begin{aligned}\left|\int_0^1 a(e(t)) \cdot h(e(t)) \cdot c(e(-2t))\, dt\right| &\leq \int_0^1 |h(e(t))| \cdot |a(e(t)) \cdot c(e(-2t))|\, dt \\ &= o(n) \int_0^1 |a(e(t)) \cdot c(e(-2t))|\, dt.\end{aligned}$$

By the Cauchy-Schwarz inequality this is at most

$$o(n) \left(\int_0^1 |a(e(t))|^2\, dt\right)^{1/2} \left(\int_0^1 |c(e(-2t))|^2\, dt\right)^{1/2}.$$

Since $a(z)$ and $c(z^{-2})$ are Laurent polynomials with at most n nonzero coefficients which all lie in $[0,1]$, the Parseval identity tells us that each of the two integrals is $\leq n$. Thus

$$\left|\int_0^1 a(e(t)) \cdot h(e(t)) \cdot c(e(-2t))\, dt\right| \leq o(n)\sqrt{n}\sqrt{n} = o(n^2).$$

Alltogether, we conclude that

$$n \geq |A_n| = p_3(A_n) = d_3^3 n^2/4 + o(n^2).$$

This forces $d_3 = \lim_{n\to\infty} r_3(n)/n = 0$, which proves Roth's theorem.

3.2 Proof of the uniform bound on the unit circle

In this section we prove Proposition 3.1.1 on the uniform behaviour of $\sum_{a\in A_n} z^a - d_3 \sum_{k=1}^n z^k$ on the unit circle. We begin with four lemmas.

For a polynomial $p(z) = a_0 + a_1 z + \cdots + a_n z^n$ and $0 \leq m \leq n$ we define $p_m(z) = a_0 + a_1 z + \cdots + a_m z^m$; thus $p_n(z) = p(z)$.

Lemma 3.2.1 *Suppose that $p(z) = a_0 + a_1 z + \cdots + a_n z^n$ is a polynomial, numbers $u, \zeta \in \mathbf{C}$ lie on the unit circle, and $|p_m(\zeta)| \leq M$ holds for all $0 \leq m \leq n$ with a constant $M > 0$. Then*

$$|p(u)| \leq M(n|u - \zeta| + 1).$$

Proof. If z is a variable and $\zeta \in \mathbf{C}$ is nonzero, we have the identity

$$\frac{p(z)}{1-z/\zeta} = \sum_{m=0}^{n-1} p_m(\zeta)(z/\zeta)^m + \frac{p(\zeta)(z/\zeta)^n}{1-z/\zeta},$$

which follows by expanding the left side in geometric series:

$$p(z)\sum_{n\geq 0}(z/\zeta)^n = a_0 + (a_0+a_1\zeta)(z/\zeta) + (a_0+a_1\zeta+a_2z^2)(z/\zeta)^2 + \cdots.$$

Hence, because $|u| = |\zeta| = 1$,

$$\begin{aligned} |p(u)| &\leq |1-u/\zeta| \sum_{m=0}^{n-1} |p_m(\zeta)| \cdot |(u/\zeta)^m| + |p(\zeta)| \cdot |(u/\zeta)^n| \\ &= |\zeta - u| \sum_{m=0}^{n-1} |p_m(\zeta)| + |p(\zeta)| \\ &\leq |\zeta - u| \cdot nM + M. \end{aligned}$$

□

Lemma 3.2.2 *For every $u \in \mathbf{C}$ lying on the unit circle and every $N \in \mathbf{N}$ there is an ω lying on the unit circle and such that $\omega^a = 1$ for some $a \leq N$ (ω is the a-th root of unity) and*

$$|u - \omega| < \frac{2\pi}{a(N+1)}.$$

Proof. Consider the $N+1$ numbers $1, u, u^2, \ldots, u^N$ on the unit circle. Its length is 2π and therefore two of the numbers lie within arc distance at most $2\pi/(N+1)$. Thus $|u^j - u^i| < 2\pi/(N+1)$ for some $0 \leq i < j \leq N$ and $|u^a - 1| < 2\pi/(N+1)$ where $0 < a = j - i \leq N$. Consider the a-th roots of the number u^a. These are vertices of a regular a-gon R inscribed in the unit circle; one vertex w coincides with u and another vertex v is close to 1, $|v-1| < 2\pi/a(N+1)$. We rotate R around the origin so that v is moved to 1 and obtain a regular a-gon R' whose vertices are a-th roots of unity. Vertex w is rotated to a number ω which is an a-th root of unity and satisfies $|u - \omega| < 2\pi/a(N+1)$. □

We extend the function $r_3(n)$ to positive real numbers by setting $r_3(x) = r_3(\lceil x \rceil)$. We know that $r_3(x) - d_3 x = o(x)$ but it is not clear whether $r_3(x) -$

d_3x is monotonous. It would be convenient to have a monotonous quantity and therefore we define

$$R(x) = \max_{1 \le t \le x} r_3(t) - d_3 t.$$

Lemma 3.2.3 *$R(x) \ge 0$, is nondecreasing and $R(x) = o(x)$ as $x \to \infty$*

Proof. The first two properties are clear from the definition. We show that $R(x) = o(x)$. Given $\varepsilon > 0$, we take x_0 such that $t > x_0$ implies $r_3(t) - d_3 t < \varepsilon t$ and then an $x_1 > x_0$ such that $x_0/x_1 < \varepsilon$. Then for $x > x_1$ we have $R(x) = r_3(t_0) - d_3 t_0$ for some $t_0 \in [1, x]$. If $t_0 > x_0$, $R(x) = r_3(t_0) - d_3 t_0 < \varepsilon t_0 \le \varepsilon x$. If $t_0 \le x_0$, $R(x) = r_3(t_0) - d_3 t_0 \le r_3(t_0) \le t_0 \le x_0 < \varepsilon x_1 < \varepsilon x$. □

Let $A_n \subset [n]$ be a set of size $r_3(n)$ not containing any AP of length 3 and

$$q(z) = f_{A_n}(z) - g_n(z) = \sum_{k=1}^{n} (\langle k \in A_n \rangle - d_3) z^k.$$

Recall that for $0 \le m \le n$, $q_m(z)$ is the initial sum of $q(z)$ obtained by replacing the upper summation index n by m.

Lemma 3.2.4 *If $n \in \mathbf{N}$ and $\omega \in \mathbf{C}$ is an a-th root of unity, i.e. $\omega^a = 1$, then for every $0 \le m \le n$*

$$|q_m(\omega)| < 2aR(n/a) + R(n).$$

Proof. For $a, b, m \in \mathbf{N}$ we denote $\alpha(b, a, m)$ and $\beta(b, a, m)$ the number of elements in $A_n \cap [m]$, respectively in $[m]$, which are congruent to b modulo a. Note that

$$\sum_{b=1}^{a} \alpha(b, a, m) = |A_n \cap [m]| \quad \text{and} \quad \sum_{b=1}^{a} \beta(b, a, m) = m.$$

Also,

$$|A_n \cap [m]| = r_3(n) - |A_n \cap [m+1, n]| \ge r_3(n) - r_3(n-m) \ge d_3 n - r_3(n-m).$$

Now, because $\omega^c = \omega^b$ whenever $c \equiv b$ modulo a, we can write

$$\begin{aligned} |q_m(z)| &= \left|\sum_{b=1}^{a} \omega^b(\alpha(b,a,m) - d_3\beta(b,a,m))\right| \\ &\le \sum_{b=1}^{a} |\alpha(b,a,m) - r_3(m/a) + r_3(m/a) - d_3\beta(b,a,m)| \\ &\le \sum_{b=1}^{a}(r_3(m/a) - \alpha(b,a,m)) + \sum_{b=1}^{a}(r_3(m/a) - d_3\beta(b,a,m)), \end{aligned}$$

where we have used that $r_3(m/a) \ge \alpha(b,a,m)$ because the set $\{c \in A_n \mid c \le m \ \&\ c \equiv b \pmod a\}$, counted by $\alpha(b,a,m)$, can be affinely mapped to $[\lfloor m/a \rfloor]$, and that $r_3(m/a) = r_3(\lceil m/a \rceil) \ge d_3\lceil m/a \rceil \ge d_3\beta(b,a,m)$. Thus

$$\begin{aligned} |q_m(z)| &\le 2ar_3(m/a) - |A_n \cap [m]| - d_3 m \\ &\le 2ar_3(m/a) - d_3 n + r_3(n-m) - d_3 m \\ &= 2a(r_3(m/a) - d_3 m/a) + (r_3(n-m) - d_3(n-m)) \\ &\le 2aR(n/a) + R(n) \end{aligned}$$

due to the monotonicity of $R(x)$. □

Proof of Proposition 3.1.1. Let $\varepsilon > 0$ be given. We want to estimate

$$|q(z)| = \left|\sum_{k=1}^{n}(\langle k \in A_n \rangle - d_3)z^k\right|$$

when n is big and z, $|z| = 1$, is arbitrary. We take an $n_0 \in \mathbf{N}$ such that $x \ge n_0$ implies $R(x) < \varepsilon x$ and then an $n_1 \in \mathbf{N}$ such that $x \ge n_1$ implies $R(x) < (\varepsilon/n_0)x$ (Lemma 3.2.3). Let $n > n_1$ and $z \in \mathbf{C}$ be an arbitrary number on the unit circle. We set $N = \lfloor n/n_0 \rfloor$ and use Lemma 3.2.2:

$$|z - \omega| < \frac{2\pi}{a(N+1)}$$

for some a-th root of unity ω, where $1 \le a \le N$. By Lemma 3.2.1, applied with $M = 2aR(n/a) + R(n)$ (Lemma 3.2.4),

$$\begin{aligned} |q(z)| &\le (2aR(n/a) + R(n)) \cdot (1 + n|z - \omega|) \\ &< (2aR(n/a) + R(n)) \cdot (1 + 2\pi n_0/a). \end{aligned}$$

We distinguish two cases. If $a \leq n_0$, then using $R(n/a) \leq R(n) < (\varepsilon/n_0)n$ we get

$$\begin{aligned} |q(z)| &\leq R(n) \cdot (2a+1)(1+2\pi n_0/a) \\ &\leq R(n) \cdot (3a + 6\pi n_0) \\ &< (\varepsilon/n_0)n \cdot 22n_0 \\ &= 22\varepsilon n. \end{aligned}$$

If the remaining case $n_0 \leq a \leq N$ we have $n/a \geq n/N \geq n_0$ and $R(n/a) < \varepsilon n/a$. Thus

$$\begin{aligned} |q(z)| &\leq (2aR(n/a) + R(n)) \cdot (1+2\pi) \\ &\leq (2a\varepsilon n/a + \varepsilon n) \cdot (1+2\pi) \\ &\leq 3\varepsilon n(1+2\pi) \\ &< 22\varepsilon n. \end{aligned}$$

□

3.3 A graph-theoretical proof

A *graph* G is a pair

$$G = (V, E)$$

where V is a finite set of *vertices* and $E \subset \binom{V}{2}$ is a set of two-element subsets of V, called *edges*. A *triangle* T in G is a triple of vertices $\{a, b, c\}$ such that every pair $\{a, b\}$, $\{a, c\}$, and $\{b, c\}$ is an edge of G. A set of triangles in G is *edge-disjoint* if every two triangles from the set are either disjoint or intersect only in one vertex, that is, they do not share an edge.

Theorem 3.3.1 (triangle removal lemma) *For every $\delta > 0$ there is an $n_0 \in \mathbf{N}$ such that the following holds. If $n > n_0$ and G is a graph on n vertices that contains m edge-disjoint triangles $T_1, T_2, \ldots, T_m$ where $m > \delta n^2$, then G must contain a triangle distinct from all triangles T_i.*

Note that always $m \leq \binom{n}{2}/3 < n^2/6$ because the edge sets of $T_1, \ldots, T_m$ are disjoint. The theorem says that if G has, in the order of magnitude, so many edge-disjoint triangles, then for big enough n there must be three of them

whose vertex sets pairwise intersect so that the three intersections do not coincide in one vertex. In the next Section we establish Theorem 3.3.1 in a stronger form by proving that in fact G must contain $\gg n^3$ triangles. Let us see now how Theorem 3.3.1 implies Roth's theorem.

Corollary 3.3.2 *For every $\delta > 0$ there is an $n_0 \in \mathbf{N}$ such that the following holds. If $n > n_0$ and $X \subset [n] \times [n]$ is a set satisfying $|X| > \delta n^2$, then X must contain a proper equilateral right-angle triangle, that is, a triple of elements (a, b), $(a + d, b)$, and $(a, b + d)$ where d is not zero.*

Proof. Let $n > n_0$ and $X \subset [n] \times [n]$ with $|X| > \delta n^2$ be given. We define an appropriate graph $G = (V, E)$ on the set V of horizontal, vertical and skew lines in $[n] \times [n]$. A horizontal (vertical) line is an n-element set $\{(m, n) \mid m \in [n]\}$ where $n \in [n]$ (a set $\{(m, n) \mid n \in [n]\}$ where $m \in [n]$) and a skew line is a set $\{(m, n) \mid m, n \in [N], m + n = p\}$ where $p \in [2, 2n]$. Thus $|V| = 4n - 1$. Two lines k, l form an edge in E if and only if $k \cap l \in X$. Note that a triangle in G is formed by three lines, one horizontal, one vertical and one skew, which pairwise intersect in points lying in X. It may happen that the three intersections coincide in one point $v \in X$ (so the three lines go through the common point v). We denote such triangle as T_v. Triangles T_v are edge-disjoint because two lines intersect in at most one point, and we have exactly $|X| > \delta n^2$ of them. By Theorem 3.3.1, G contains a triangle T that is distinct from all T_v, $v \in X$. The lines in T must intersect in three distinct points, which form the desired proper equilateral right-angle triangle in X. □

Corollary 3.3.3 *For every $\delta > 0$ there is an $n_0 \in \mathbf{N}$ such that if $n > n_0$ and $A \subset [n]$ satisfies $|A| > \delta n$, then A contains a proper arithmetic progression of length* 3.

Proof. Let $n > n_0$ and $A \subset [n]$ with $|A| > \delta n$ be given. Consider the set $X \subset [n] \times [n]$ defined by

$$(a, b) \in X \iff a + 2b \in A.$$

Since for fixed $c \in A$ the number of solutions of $a + 2b = c$ for $a, b \in [n]$ is at least $\lfloor n/2 \rfloor - 1$, we have

$$|X| \geq \sum_{c \in A} \lfloor n/2 \rfloor - 1 = |A|(\lfloor n/2 \rfloor - 1) > \delta n^2/2 - 2\delta n.$$

By the previous corollary, for big enough n the set X contain a triple (a, b), $(a+d, b)$, and $(a, b+d)$ with $d \neq 0$. Hence the proper arithmetic progression $a + 2b$, $a + d + 2b$, $a + 2b + 2d$ with difference d lies in A. □

3.4 Szemerédi's regularity lemma

We first introduce ε-regular pairs and prove that a tripartite graph in which each two parts form an ε-regular pair contains many triangles (counting lemma). Then we state the regularity lemma and prove by means of it and by means of the counting lemma Theorem 3.3.1. In conclusion we prove the regularity lemma.

Let $G = (V, E)$ be a graph and $X, Y \subset V$ be two disjoint sets. We denote by $e(X, Y)$ the number of edges in G joining X and Y, and by $d(X, Y)$ the density of these edges:

$$d(X, Y) = \frac{e(X, Y)}{|X| \cdot |Y|}.$$

We say that (X, Y) is an *ε-regular pair* if for every subset $X_1 \subset X$ and $Y_1 \subset Y$ satisfying $|X_1| \geq \varepsilon|X|$ and $|Y_1| \geq \varepsilon|Y|$ one has

$$|d(X_1, Y_1) - d(X, Y)| < \varepsilon.$$

For $x \in X$ we denote

$$\Gamma_Y(x) = \{y \in Y \mid \{x, y\} \in E\},$$

the set of neighbors of x in Y. Clearly, $|\Gamma_Y(x)| = e(\{x\}, Y)$.

Lemma 3.4.1 *Let (X, Y) be an ε-regular pair with edge density $d = d(X, Y)$ in a graph $G = (V, E)$ and let $X_1 \subset X$ and $Y_1 \subset Y$ be subsets satisfying $|X_1| \geq \varepsilon|X|$ and $|Y_1| \geq \varepsilon|Y|$. Then there exists a vertex $x \in X_1$ such that*

$$|\Gamma_{Y_1}(x)| > (d - \varepsilon)|Y_1|.$$

The same is true with the inequality $< (d + \varepsilon)|Y_1|$.

Proof. If this were not true, then $|\Gamma_{Y_1}(x)| \leq (d - \varepsilon)|Y_1|$ would hold for every $x \in X_1$ and we would get

$$e(X_1, Y_1) = \sum_{x \in X_1} |\Gamma_{Y_1}(x)| \leq (d - \varepsilon)|X_1| \cdot |Y_1|$$

and $d(X_1, Y_1) \leq d - \varepsilon$, contradicting the regularity of the pair (X, Y). The proof of the other inequality is similar. □

Proposition 3.4.2 (counting lemma) *If $G = (U \cup V \cup W, E)$ is a tripartite graph (edges go only between the sets U and V, U and W, and V and W) with all three pairs (U, V), (U, W), and (V, W) ε-regular, then, denoting the edge densities by $\kappa = d(U, V)$, $\lambda = d(U, W)$, and $\mu = d(V, W)$,*

$$\#\textit{triangles in } G > (\kappa\lambda\mu - 5\varepsilon - \varepsilon^3)|U| \cdot |V| \cdot |W|.$$

Proof. The proposition holds trivially if one of the densities is smaller than 2ε (then $\kappa\lambda\mu - 5\varepsilon - \varepsilon^3 < 2\varepsilon - 5\varepsilon - \varepsilon^3 < 0$). We will therefore assume that $\kappa, \lambda, \mu \geq 2\varepsilon$. Let

$$U_1 = \{u \in U \mid |\Gamma_V(u)| \leq (\kappa - \varepsilon)|V|\}.$$

By the previous lemma, we must have $|U_1| < \varepsilon|U|$. The same is true, by Lemma 3.4.1, for the subset U_2 defined by the condition

$$|\Gamma_W(u)| \leq (\lambda - \varepsilon)|W|.$$

Thus the set $U_0 = U \backslash (U_1 \cup U_2)$ satisfies $|U_0| \geq (1 - 2\varepsilon)|U|$ and if $u \in U_0$ then

$$|\Gamma_V(u)| > (\kappa - \varepsilon)|V| \quad \text{and} \quad |\Gamma_W(u)| > (\lambda - \varepsilon)|W|.$$

Because $\kappa - \varepsilon, \lambda - \varepsilon \geq \varepsilon$ and the pair (V, W) is ε-regular, for every $u \in U_0$ we have

$$e(\Gamma_V(u), \Gamma_W(u)) > (\mu - \varepsilon)|\Gamma_V(u)| \cdot |\Gamma_W(u)|.$$

This is a lower bound on the number of triangles with one vertex in u because every edge joining $\Gamma_V(u)$ and $\Gamma_W(u)$ forms together with u a triangle. Summing over U_0, we get the lower bound

$$\begin{aligned}
\#\text{triangles in } G &> \sum_{u \in U_0} e(\Gamma_V(u), \Gamma_W(u)) > (\mu - \varepsilon) \sum_{u \in U_0} |\Gamma_V(u)| \cdot |\Gamma_W(u)| \\
&> (\mu - \varepsilon)|U_0| \cdot (\kappa - \varepsilon)|V| \cdot (\lambda - \varepsilon)|W| \\
&> (1 - 2\varepsilon)(\kappa - \varepsilon)(\lambda - \varepsilon)(\mu - \varepsilon)|U| \cdot |V| \cdot |W| \\
&> (\kappa\lambda\mu - 3\varepsilon - \varepsilon^3 - 2\varepsilon)|U| \cdot |V| \cdot |W|,
\end{aligned}$$

which proves the stated bound. □

The following decomposition of sufficiently large graphs in ε-regular pairs is one of the most important results in graph theory.

Theorem 3.4.3 (Szemerédi's regularity lemma) *For every $\varepsilon > 0$ and every $m \in \mathbf{N}$, there exists an $M \in \mathbf{N}$ with the property that in every graph $G = (V, E)$ on more than M vertices the vertex set V can be partitioned as*

$$V = V_1 \cup V_2 \cup \ldots \cup V_r$$

so that (i) $m \leq r \leq M$, (ii) the cardinalities $|V_i|$ differ among themselves at most by 1, and (iii) with the exception of at most $\varepsilon\binom{r}{2}$ pairs, the pairs (V_i, V_j), $1 \leq i < j \leq r$, are ε-regular.

We postpone the proof. Given a parameter $h > 0$ and a partition of V described in the regularity lemma, we say that an edge e in G is *h-good* with respect to the partition, if e joins two parts V_i and V_j such that the pair (V_i, V_j) is ε-regular and $d(V_i, V_j) \geq h$. Remaining edges of G are called *h-bad*; we bound their number.

Lemma 3.4.4 *The number of h-bad edges is at most*

$$2(1/m + \varepsilon + h)n^2.$$

Proof. An edge e is h-bad iff it lies inside one part V_i or joins two distinct parts V_i and V_j such that the pair (V_i, V_j) is not ε-regular or $d(V_i, V_j) < h$. Thus the number of h-bad edges is at most

$$r\binom{\lceil n/r \rceil}{2} + \varepsilon\binom{r}{2}\lceil n/r \rceil^2 + h\binom{r}{2}\lceil n/r \rceil^2.$$

From $n/r > M/M = 1$ we have $\lceil n/r \rceil \leq 2n/r$. This and $r \geq m$ imply the stated bound. □

Proof of Theorem 3.3.1. We prove that if $G = (V, E)$ is a graph on n vertices containing $> \delta n^2$ edge-disjoint triangles T_i and n is big (depending on $\delta > 0$), then

$$\#\text{triangles in } G > \kappa n^3$$

for some constant $\kappa > 0$ depending only on δ. Thus for big n there are many more triangles in G than the at most $n^2/6$ edge-disjoint triangles T_i.

Let $\delta > 0$ be given. We fix sufficiently small $\varepsilon > 0$ and sufficiently large $m \in \mathbf{N}$ such that

$$2(1/m + \varepsilon + (6\varepsilon + \varepsilon^3)^{1/3}) < \delta.$$

Let $M \in \mathbf{N}$ be the constant corresponding to these ε and m in the regularity lemma and let $G = (V, E)$ be any graph that has $n > M$ vertices and contains $> \delta n^2$ edge-disjoint triangles T_i. We consider the partition $V = V_1 \cup V_2 \cup \ldots \cup V_r$, $m \leq r \leq M$, ensured by the regularity lemma and delete from G all h-bad edges, where $h = (6\varepsilon + \varepsilon^3)^{1/3}$. By Lemma 3.4.4 and by the selection of ε and m, the resulting graph G' must still contain at least one triangle T_i (their edge sets are disjoint and to get rid of all of them, we have to delete more than δn^2 edges). But G' consists only of h-good edges and it follows that there must be three parts V_i, V_j, and V_k, $1 \leq i < j < k \leq r$, in the partition of V such that all three pairs (V_i, V_j), (V_i, V_k), and (V_j, V_k) are ε-regular and their edge densities are $\geq h$. By Proposition 3.4.2, the tripartite graph H induced by G on $V_i \cup V_j \cup V_k$ satisfies that

$$\begin{aligned} \#\text{triangles in } H &> (h^3 - 5\varepsilon - \varepsilon^3)|V_i| \cdot |V_j| \cdot |V_k| \\ &> \varepsilon \lfloor n/r \rfloor^3 \\ &> (\varepsilon/8M^3)n^3. \end{aligned}$$

Thus the theorem holds with $\kappa = \varepsilon/8M^3$. □

3.5 Remarks

Bibliography

[1] J. Bak and D. J. Newman, *Complex Analysis*, Springer, Berlin, 1997.

[2] M. Nair, On Chebyshev-type inequalities for primes, *Amer. Math. Monthly* **89** (1982) 126–129.

[3] D. J. Newman, Simple analytic proof of the prime number theorem, *Amer. Math. Monthly* **87** (1980) 693–696.

[4] D. J. Newman, *Analytic Number Theory*, Springer, Berlin, 1998.

[5] K. Roth, Sur quelques ensembles d'entiers, *C. R. Acad. Sci. Paris* **234** (1952) 388–390.

[6] K. F. Roth, On certain sets of integers, *J. London Math. Soc.* **28** (1953) 104–109.

[7] K. F. Roth, Rational approximations to algebraic numbers, *Mathematika* **2** (1955) 1–20; corrigendum, 168.

[8] G. Tenenbaum, *Introduction to analytic and probabilistic number theory*, Cambridge University Press, Cambridge, U.K., 1997.

[9] J. Veselý, *Komplexní analýza pro učitele*, Karolinum, Praha, 2000

[10] D. Zagier, Newman's short proof of the prime number theorem, *Amer. Math. Monthly* **104** (1997) 705–708.