1. Basic Setup

Recall that we are estimating $A(x) = \sum_{n \leq x} a_n$ by looking at $X_d = \sum_{n \in dP} a_n$, the latter of which is generally estimated by considering $A_d = g(d)X + r(d)$, where $g$ is some set of weights coming close to being multiplicative and $X$ well approximates $A(x)$. Recall that we also have some distinguished primes $P$ to avoid, up to some height $z$ (denoted $P(z)$). The M"obius $\mu$-function lets us express this coprimality as a convolution, but it has the downside of being hard to estimate. We will choose some replacements $\lambda_n$ for $\mu(n)$, supported up to height $D$, the level of distribution. If $D(x) = x^\alpha$, then call $\alpha$ the exponent of distribution. Recall further from the Legendre form that we expect $P(d, P(z)) = 1$ $X_d$ to be something like $X \cdot \prod (1 - g(p)) \approx G = \sum_{d \in P(z)} \lambda_d g(d)$.

In what follows, we will be heavily using convolutions of arithmetic functions. If this is confusing, just know that every convolution takes the form $C \ast (\sum_{cd=n} f(c)g(d))$ in some appropriate ambient group. In our case, we have M"obius inversion: convoluting by 1 is inverse to convoluting by $\mu$.

2. Correction to First Week

There was a problem in the first week about how I explained Zhang’s result on “raising the level” or “breaking through the square root barrier”. Indeed, I meant that $\alpha$ goes from $\frac{1}{2}$ to $\frac{1001}{2000}$. The level of distribution $D$ must satisfy that $R = \sum_{d \in P(z)} \lambda_d g(d) \ll A(x)(\log x)^{-A}$. Of course, raising the exponent of distribution is non-trivial (and even that it has one!).

3. Dirichlet Hyperbola Trick

3.1. Basic Example. The function $\tau(n)$ returns the number of positive divisors of $n$. Then an over count gives

$$\sum_{n \leq x} \tau(n) = \sum_{d \leq x} \lfloor x/d \rfloor = x \log(x) + O(x).$$

We know that divisors come in pairs, as solutions to $cd = n$, so

$$\sum_{n \leq x} \tau(n) = 2 \sum_{d \leq \sqrt{x}} \lfloor x/d \rfloor - \lfloor \sqrt{x} \rfloor^2 = x(\log x + 2\gamma - 1) + O(\sqrt{x}).$$

The equation $cd = n$ can be viewed as the level set of a convolution, and apparently the technique above applies more generally to convolutions.
3.2. Another Example. Let \( r(n) = 4 \sum_{d \mid n} \chi_4(d) \). We use the trick above, but this time there’s not a direct pairing, but rather we get

\[
\sum_{n \leq x} r(n) = 4 \sum_{d \leq \sqrt{x}} \chi_4(d) \lfloor x/d \rfloor 4 \sum_{c \leq \sqrt{x}} \sum_{d \leq x/c} \chi_4(d) + O(\sqrt{x}).
\]

Our target is the first summation on the RHS. Splitting into integral and fractional parts we get that

\[
\sum_{d \leq \sqrt{x}} \chi_4(d) \frac{x}{d} = L(1, \chi_4) + O(x^{1/2})
\]

from which, after applying \( L(1, \chi_4) = \pi/4 \), we get the estimate

\[
\sum_{n \leq x} r(n) = \pi x + O(\sqrt{x}).
\]

3.3. A Harder Example. I have not studied this well. But the start is

\[
\sum_{n \leq x} \tau(n^2 + 1) = 2 \sum_{d \leq x} \sum_{n \leq d, d \neq n^2 + 1} 1 - \sum_{c \mid d, d^2 = n^2 + 1} 1 + O(x) = 2x \sum_{d \leq x} \frac{\rho(d)}{d} + O(x).
\]

From the generating series for \( \rho(d) \) as \( \zeta(2s)^{-1} \zeta(s) L(s, \chi_4) \), one has \( \rho(d) = \sum_{a^2bc = d} \mu(a) \chi_4(c) \) and we can get \( \sum_{d \leq x} \frac{\rho(d)}{d} = \frac{L(1, \chi_4)}{\zeta(2)} \log x + O(1) \). Hence,

\[
\sum_{n \leq x} \tau(n^2 + 1) = \frac{3}{\pi} x \log x + O(x).
\]

4. The Von Mangoldt Function

It is the convolution \( \Lambda \) of \( \mu \) and the logarithm. Also, convoluting the constant function with \( \Lambda \) gives the logarithm. We can use this function to detect large prime divisors. Let \( y \) be the cutoff and set

\[
S(x, y) = \sum_{n \leq x} a_n \left( \sum_{d \mid n, d > y} \Lambda(d) \right).
\]

Note that ignoring \( d > y \) means we get a regular convolution \( A'(x) = \sum a_n \log n = A(x)(\log x + O(1)) \), where the bound holds for many sequences. So, \( S(x, y) \) is the difference between \( A'(x) \) and the contribution of low primes. Using the forms in §1, and noting a few natural approximations, we achieve

**Proposition 4.1.** Suppose the remainder terms satisfy \( \sum_{d \leq x} \Lambda(d) r_d(x) \ll A(x) \) for a power of \( x \) between 0 and 1. Then,

\[
S(x, y) \ll_x A(x)(\log(x) + O(1)).
\]

If \( x \) is large enough to outpace the \( O(1) \), we get that some \( n \) has a primary factor larger than \( x^\theta \).
5. Further Generalities About Sieves

We want some regularity on the weights \( g(p) \). Typically one has

\[
\sum_{p \leq z} g(p) \log(p) = \kappa \log z + O(1).
\]

Partial summation gives

\[
\sum_{p \leq z} g(p) = \kappa \log \log z + \alpha + O((\log z)^{-1})
\]

(\( \kappa \) is the sieve dimension) whence a formula of Mertens gives

\[
\prod_{p \leq z} (1 - g(p)) = e^{-\gamma \kappa} H(\log z)^{-\kappa}(1 + O((\log z)^{-1})).
\]

We will want \( V \) to grow at most like a logarithmic power:

\[
\frac{V(w)}{V(z)} \ll (\log w)^{\kappa}
\]

and hence

\[
g(p) \leq 1 - \frac{1}{K} - 1.
\]

If this \( K \) is too large, cut out some small primes from \( \mathcal{P} \) so that you can replace 2 with something larger, and then handle these small primes some other way. You can also instead enlarge \( \kappa \). Increasing it by 1 adds a \( (\log z) \) to the denominator.

If you have this logarithmic growth of \( V \), you get nice bounds.

**Lemma 5.1.** Let \( h \) be continuous, non-negative, and non-decreasing. Let \( \Delta = \sum_{y \leq p < z} g(p)h(p)V(p) \).

\[
\Delta \leq -KV(z) = \int_y^z h(w)d(\log w)\kappa + (K-1)h(z)V(z),
\]

\[
\Delta \leq -V(z) \int_y^z h(w)d(\log w)\kappa + \left(1 - \frac{1}{K}\right)h(z)V(y),
\]

\[
\Delta_{\text{without } V} \leq \kappa \int_y^z \frac{h(w)}{w \log w} dw + h(z) \log K.
\]

5.1. **Monotonicity.** As a precursor, we will be taking upper and lower bound sieves with weights \( \lambda_d^+ \) and \( \lambda_d^- \) respectively. It turns out that in the multiplicative weight case, the signs refer to the signs of the corresponding \( \theta = 1 * \lambda \) when things are not coprime to \( \mathcal{P} \). It says nothing about the signs of the \( \lambda \) themselves.

By inversion,

\[
\lambda_d = \sum_{m|d} \mu(m)\theta_n
\]

whence

\[
G = \sum_{m,n} \mu(m)g(mn)\theta_n.
\]

If \( g \) is multiplicative, we can eventually derive that \( G \geq V \) (resp. \( G \leq V \)) when \( \theta_n^+ \geq 0 \) (resp. \( \theta_n^- \leq 0 \)) for \( n \) not coprime to the sifting set. Admitting all of this, we have the following

**Proposition 5.2.** With assumptions as above, we have

\[
\sum_{(d,q)=1} \lambda_d^+ g(d) \leq (1 - g(q))^{-1} \sum_d \lambda_d^+ g(d)
\]

and the reverse for the minus case.

We can also fiddle with the sieve weights.
Proposition 5.3. If $g'(d) \geq g(d)$, one has

$$G' \geq G \prod \left( \frac{1 - g'(p)}{1 - g(p)} \right).$$

6. Quick Comments on Composition

Composition of sieves with weights $\lambda'$ and $\lambda''$ just multiplies the $\theta$'s involved. Also,

$$G' \ast G'' = \sum_{(d_1,d_2)=1} \lambda_{d_1}' \lambda_{d_2}'' g'(d_1)g''(d_2)$$

which is useful when the two sequences are linearly dependent. See §14.7.