§4. Vinogradov's Three-Primes Theorem.

Vinogradov's famous theorem asserts that every sufficiently large odd number is the sum of three primes. Together with Chen's theorem (every sufficiently large even number is the sum of p and q, where p is prime and q is the product of at most two primes) this is one of the strongest results in the direction of Goldbach's conjecture. In this section we shall see how to use exponential-sum estimates to prove Vinogradov's theorem, and we shall also gain some insight into why Goldbach's conjecture itself is out of reach.

We begin with some definitions and simple lemmas. Given $n \in \mathbb{N}$, let $\Lambda(n)$ be $\log p$ if $n = p^k$ with p prime, $k \geqslant 1$ and zero otherwise. Let $\mu(n) = (-1)^k$ if n is a product of k distinct primes (interpreting this as 1 when $n = 1$) and zero otherwise. These functions are called von Mangoldt's function and the Möbius function respectively.

Lemma 1. *Let $x \in \mathbb{N}$. Then $\sum_{d|x} \Lambda(d) = \log x$.*

Proof. Write x as a product of prime powers and it becomes obvious. □

Lemma 2. *Let $x \in \mathbb{N}$. Then $\sum_{d|x} \mu(d)$ equals zero unless $x = 1$ in which case it equals one.*

Proof. Let $x \geqslant 2$ and write $x = p_1^{a_1} \dots p_k^{a_k}$. Then every subset $A \subset [k]$ contributes $(-1)^{|A|}$ to the sum $\sum_{d|x} \mu(d)$. But

$$\sum_{A \subset [k]} (-1)^{|A|} = \sum_{j=0}^{k} (-1)^j \binom{k}{j} = (1-1)^k = 0 \ .$$

(Another way of looking at the last calculation is that a randomly chosen subset of $[k]$ has the same chance of being of even as of odd size.) □

Recall that $d(x)$ is defined to be the number of divisors of x. We know from the previous section that $d(x)$ is sometimes quite large. The next lemma shows that this does not happen all that often.

Lemma 3. *Let $n \in \mathbb{N}$. Then $\sum_{x \leqslant n} d(x)^2 \leqslant 2n(\log n)^3$.*

Proof. This is surprisingly easy to prove. Indeed,

$$\begin{aligned}
\sum_{x\leqslant n} d(x)^2 &= \sum_{x\leqslant n}\sum b|x \sum_{c|x} 1 \\
&= \sum_{b\leqslant n}\sum_{c\leqslant n}\sum_{y.\mathrm{lcm}(a,b)\leqslant n} 1 \\
&\leqslant \sum_{a\leqslant n}\sum_{d\leqslant n/a}\sum_{e\leqslant n/ad}\sum_{y\leqslant n/ade} 1 \\
&\leqslant \sum_{a\leqslant n}\sum_{d\leqslant n/a}\sum_{e\leqslant n/ad} n/ade \\
&\leqslant \sum_{a\leqslant n}\sum_{d\leqslant n/a} (n/ad)(\log n+1) \\
&\leqslant \sum_{a\leqslant n}(n/a)(\log n+1)^2 \\
&\leqslant n(\log n+1)^3
\end{aligned}$$

which proves the lemma. □

It is easy to check that the number of ways of writing n as the sum of three primes is $\int F(\alpha)^3 e(-\alpha n)d\alpha$, where $F(\alpha)$ is the function $\sum_{p\leqslant n} e(\alpha p)$. Roughly speaking, our aim will be to estimate $F(\alpha)$ for every α, and use this estimate to prove that the integral is non-zero. As in the previous section, $F(\alpha)$ turns out to be small when α is not too close to a rational with small denominator. When it is close to such a rational, we shall use results about the distribution of primes in an arithmetic progression to estimate $F(\alpha)$ directly.

There are, however, certain advantages in weighting the primes so that their density is approximately constant through the interval. Since the density near m is $(\log m)^{-1}$, the appropriate weight to give p is $\log p$. Accordingly, we shall estimate the function $f(\alpha) = \sum_{p\leqslant n}\log p e(\alpha p)$. The integral $\int f(\alpha)^3 e(-\alpha n)d\alpha$ gives us the sum of $(\log p_1)(\log p_2)(\log p_3)$ over all triples (p_1, p_2, p_3) such that $p_1 + p_2 + p_3 = n$, so for the purposes of Vinogradov's theorem it is enough to prove that this integral is non-zero for large enough odd n.

Finally, even this function is not always the most convenient to estimate. The next lemma shows that we may replace it by $g(\alpha) = \sum_{x\leqslant n}\Lambda(x)e(\alpha x)$, with only a small error.

Lemma 4. *$|f(\alpha) - g(\alpha)| \leqslant C\sqrt{n}$ for every α and some absolute constant C.*

Proof. $g(\alpha) - f(\alpha) = \sum_{p^k\leqslant n, k\geqslant 2}\log p e(\alpha p^k)$ which in modulus is at most $(\log_2 n)\sum_{p\leqslant\sqrt{n}} 1$. By Chebyshev's theorem the result follows. □

The next lemma is similar to the lemma we kept using during the proof of Weyl's inequality, and follows from it. Since we are about to prove several results with the same hypotheses, let us state them once and for all before starting. Thus, a and q will be positive integers with $(a, q) = 1)$ and α is a real number with $|\alpha - a/q| \leqslant q^{-2}$.

Lemma 5. *Let Q, R be positive integers with $q \leqslant Q$. Then*

$$\sum_{x=1}^{R} \min\{\|\alpha x\|^{-1}, Qx^{-1}\} \leqslant 200 \log Q \log R(q + R + Qq^{-1}) \ .$$

Proof. We know from §3 that the numbers $0, \alpha, 2\alpha, \ldots, \lfloor (q/2) \rfloor \alpha$ are $(2q)^{-1}$-separated. Therefore,

$$\begin{aligned} \sum_{x \leqslant q/2} \min\{\|\alpha x\|^{-1}, Qx^{-1}\} &\leqslant 2 \sum_{x \leqslant \lceil q/4 \rceil} 2q/x \\ &\leqslant 4q \log q \ . \end{aligned}$$

Given an integer i, let S_i be the sum

$$\sum_{x=2^{i-1}}^{2^i - 1} \min\{\|\alpha x\|^{-1}, Qx^{-1}\} \ .$$

Then

$$S_i \leqslant \sum_{x=2^{i-1}}^{2^i - 1} \min\{\|\alpha x\|^{-1}, Q/2^{i-1}\}$$

which, by Lemma 2 of §3, is at most $48 \log Q(2^{-(i-1)}Q + 2^{i-1} + q + Qq^{-1})$. Summing over all i such that $2^i > q/2$ and $2^{i-1} \leqslant R$, we obtain the desired result. □

We now prove an identity due to Vaughan, which will allow us to show that $g(\alpha)$ is small when α is not close to a rational with small denominator. This identity seems mysterious when it is just drawn out of a hat, but the mystery can be reduced with a few remarks.

We wish to show that $g(\alpha) = \sum_{x \leqslant n} \Lambda(x) e(\alpha x)$ is appreciably smaller than n when q is not too small (or too large). The function which is hard to understand is of course Λ, but we know that Λ has the nice property that $\sum_{d|x} \Lambda(d) = \log x$, which is much more familiar. Therefore, we try to express $g(\alpha)$ as a sum of pieces of this form. As a first observation, we notice (or rather, it has been noticed) that

$$\sum_{x \leqslant n} \sum_{y \leqslant n/x} \Lambda(x) e(\alpha x y) = \sum_{u \leqslant n} \sum_{x|u} \Lambda(x) e(\alpha u) \ .$$

This is very promising, because

$$\begin{aligned}\sum_{x\leqslant n}\Lambda(x)e(\alpha x)&=\sum_{x\leqslant n}\sum_{y\leqslant n/x}\sum_{d|y}\mu(d)\Lambda(x)e(\alpha xy)\\&=\sum_{d\leqslant n}\mu(d)\sum_{z\leqslant n/d}\sum_{x\leqslant n/zd}\Lambda(x)e(\alpha dxz)\ ,\end{aligned}$$

which is a ±1-combination of sums of the required form, and therefore seems to have a chance of being small.

Now it is clearly not easy to obtain a good estimate for the last quantity directly, because d takes n possible values and for each one we are not going to do better than a modulus of 1. (In principle one might show that there was considerable cancellation, but then one would be back to trying to understand mysterious functions rather than writing $g(\alpha)$ as a clever combination of terms that can be estimated by elementary means.) It is therefore essential to restrict d. However, this introduces a new error term which must be shown to be small. Moreover, showing that this error term is small turns out not to be possible unless we also restrict x to be not too small. (This is the point that I did not appreciate until shortly after the lecture where I had problems with $x=y$. It results in the extra error term T, and the extra restriction $X<x$ which helps to estimate U.) So it is likely that Vaughan arrived at the identity which we now prove by a process of trial and error, starting with the observations above.

Lemma 6. *Let $X=n^{2/5}$. Then $g(\alpha)=\sum_{x\leqslant n}\Lambda(x)e(\alpha x)=S-T-U+O(n^{2/5})$, where*

$$S=\sum_{d\leqslant X}\mu(d)\sum_{z\leqslant n/d}\sum_{x\leqslant n/zd}\Lambda(x)e(\alpha dxz)\ ,$$

$$T=\sum_{d\leqslant X}\mu(d)\sum_{z\leqslant n/d}\sum_{x\leqslant X,x\leqslant n/zd}\Lambda(x)e(\alpha dxz)$$

and

$$U=\sum_{X<u\leqslant n}\sum_{d|u,d\leqslant X}\mu(d)\sum_{X<x\leqslant n/u}\Lambda(x)e(\alpha xu)\ .$$

Proof. Let us write τ_u for $\sum_{d|u,d\leqslant X}\mu(d)$. Then, by Lemma 2, we know that τ_u is 1 when $u=1$ and 0 when $1<u\leqslant X$. Therefore,

$$\sum_{u\leqslant n}\tau_u\sum_{X<x\leqslant n/u}\Lambda(x)e(\alpha xu)=U+\sum_{X<x\leqslant n}\Lambda(x)e(\alpha x)\ .$$

But, by Chebyshev's theorem (as in the proof of Lemma 4),

$$\sum_{X<x\leqslant n} \Lambda(x)e(\alpha x) = g(\alpha) + O(n^{2/5}) \ .$$

We also know that

$$\begin{aligned}\sum_{u\leqslant n} \tau_u \sum_{X<x\leqslant n/u} \Lambda(x)e(\alpha xu) &= \sum_{u\leqslant n} \sum_{d|u, d\leqslant X} \mu(d) \sum_{X<x\leqslant n/u} \Lambda(x)e(\alpha xu) \\ &= \sum_{d\leqslant X} \mu(d) \sum_{z\leqslant n/d} \sum_{X<x\leqslant n/dz} \Lambda(x)e(\alpha xzd) \\ &= S - T \ .\end{aligned}$$

The identity follows. □

In the next three lemmas, we show that each of S, T and U is small. Notice that S is the sum we originally expected to be able to bound, and is therefore in a sense the important one, while T and U are error terms that we were unable to avoid introducing.

Lemma 7. $|S| \leqslant 80(\log n)^3(q + X + n/q)$.

Proof. Writing u for xz, we have

$$|S| = \Big|\sum_{d\leqslant X} \mu(d) \sum_{u\leqslant n/d} \sum_{x|u} \Lambda(x)e(\alpha du)\Big| \leqslant \sum_{d\leqslant X} \Big|\sum_{u\leqslant n/d} \log u e(\alpha du)\Big|$$

by Lemma 1. But

$$\begin{aligned}\Big|\sum_{u\leqslant n/d} \log u e(\alpha du)\Big| &= \Big|\sum_{u\leqslant n/d} \int_1^u e(\alpha du)\, dt/t\Big| \\ &\leqslant \int_1^{n/d} \Big|\sum_{t\leqslant u\leqslant n/d} e(\alpha du)\Big|\, dt/t \\ &\leqslant \int_1^{n/d} \min\{\|\alpha d\|^{-1}, n/d\}\, dt/t \\ &\leqslant \log n \min\{\|\alpha d\|^{-1}, n/d\} \ .\end{aligned}$$

Summing over $d \leqslant X$ and applying Lemma 5 (taking into account that $\log X = (2/5)\log n$) we obtain the bound claimed. □

Because I did not prove the next lemma in lectures, it is starred. However, it is no harder than the other ones.

Lemma 8. $|T| \leqslant 160(\log n)^3(q + X^2 + n/q)$.

Proof. Interchanging the order of summation of z and x in the definition of T, and using the fact that $|\mu(d)| \leqslant 1$, we have

$$|T| \leqslant \sum_{d\leqslant X}\sum_{x\leqslant X}\Lambda(x)\Big|\sum_{z\leqslant n/dx} e(\alpha dxz)\Big| .$$

Now let $y = dx$, and this becomes

$$\sum_{y\leqslant X^2}\sum_{x\leqslant X, x|y}\Lambda(x)\Big|\sum_{z\leqslant n/y} e(\alpha yz)\Big| .$$

By Lemma 1, $\sum_{x\leqslant X, x|y}\Lambda(x) \leqslant \log y \leqslant \log n$, so we can bound this above by

$$\log n \sum_{k\leqslant X^2}\min\{\|\alpha y\|^{-1}, n/y\}$$

which is at most the bound stated, by Lemma 5. □

The next lemma is the correct version of the one I got stuck on in lectures. The extra ingredient needed is the bounding of x away from zero, which stops u from getting too large.

Lemma 9. $|U| \leqslant 40(\log n)^4(n^{1/2}q^{1/2} + n/X^{1/2} + nq^{-1/2})$.

Proof. Given a positive integer i, let U_i be the sum

$$\sum_{u=2^{i-1}}^{2^i-1}|\tau_u|\Big|\sum_{X<x\leqslant n/u}\Lambda(x)e(\alpha xu)\Big| .$$

Notice that $U_i = 0$ when $2^{i-1} \geqslant n/X$ (because it is then impossible to satisfy the inequality $X < x \leqslant n/u$), and that $|U|$ is therefore at most the sum of all U_i over all i such that $2^i > X$ and $2^{i-1} < n/X$. It is easy to check that there are at most $\log n$ such values of i. (The fact that 2^i is between roughly $n^{2/5}$ and roughly $n^{3/5}$ more than compensates for the replacement of $\log_2 n$ by $\log n$.) We shall estimate the U_i separately.

By the Cauchy-Schwarz inequality,

$$U_i^2 \leqslant \Big(\sum_{u=2^{i-1}}^{2^i-1}|\tau_u|^2\Big)\Big(\sum_{u=2^{i-1}}^{2^i-1}\Big|\sum_{X<x\leqslant n/u}\Lambda(x)e(\alpha xu)\Big|^2\Big) .$$

Now $|\tau_u|$ is obviously at most $d(u)$, so

$$\begin{aligned}\sum_{u=2^{i-1}}^{2^i-1} |\tau_u|^2 &\leqslant \sum_{u=2^{i-1}}^{2^i-1} d(u)^2 \\ &\leqslant \sum_{u=1}^{2^i} d(u)^2 \ ,\end{aligned}$$

which is at most $2^i(\log n)^3$, by Lemma 3. (The factor 2 disappeared, because $2^i < 2n^{3/5}$, so $\log(2^i)$ is actually a bit smaller than $\log n$, but this is of course a desperately unimportant point.)

As for the other bracket, if we expand out the modulus squared, we find that it equals

$$\sum_{u=2^{i-1}}^{2^i-1} \sum_{X<x\leqslant n/u} \sum_{X<y\leqslant n/u} \Lambda(x)\Lambda(y)e(\alpha(x-y)u) \ .$$

Interchanging the sum over u with those over x and y, we find that this is at most

$$\sum_{X<x\leqslant n/2^{i-1}} \sum_{X<y\leqslant n/2^{i-1}} \Lambda(x)\Lambda(y)\Big|\sum_{2^{i-1}\leqslant u<2^i, u\leqslant \min\{n/x,n/y\}} e(\alpha(x-y)u)\Big|$$

which is at most

$$\sum_{X<x\leqslant n/2^{i-1}} \sum_{X<y\leqslant n/2^{i-1}} \Lambda(x)\Lambda(y)\min\{\|\alpha(x-y)\|^{-1}, 2^{i-1}\} \ .$$

Writing z for $x-y$ and observing that each z occurs at most $n/2^{i-1}$ times, we can bound this sum above by

$$(\log n)^2(n/2^{i-1}) \sum_{n/2^{i-1}<z\leqslant n/2^{i-1}} \min\{\|\alpha z\|^{-1}, 2^{i-1}\} \ ,$$

which, by Lemma 2 of §3, is at most

$$(\log n)^2 .48 \log n(q + n/2^{i-2} + 2^{i-1} + 2n/q) \ .$$

Multiplying the two estimates together, we have shown that

$$U_i^2 \leqslant 96n(\log n)^6(q + 4n/2^i + 2^{i-1} + 2n/q) \ ,$$

which implies, since $n/2^i$ and 2^{i-1} are at most n/X, that

$$U_i \leqslant 40(\log n)^3(n^{1/2}q^{1/2} + n/X^{1/2} + nq^{-1/2}) \ .$$

Since there are at most $\log n$ values of i such that U_i contributes to U, the result follows. □

Remarks. In the lectures I used Lemma 5 of this section where I have just used Lemma 2 of §3. However, since z is roughly constant, it is clear that that could not have achieved anything. (I was imitating Nathanson, which was a mistake, and he appears not to notice that he has to take account of the case $z = 0$.) It may look complicated to split the sum into $\log n$ (or so) further pieces, but this was a good (and standard) thing to do because we were estimating something of the form $\sum_u f(u)g(u)$, where $f(u)$ appeared to be roughly proportional to u and $g(u)$ roughly proportional to u^{-1}. So applying the Cauchy-Schwarz inequality straight away would have been disastrous. Note that the choice of $X = n^{2/5}$ was made in order to minimize $\max\{X^2, nX^{-1/2}\}$.

If we put together Lemmas 4 and 6 to 9 we obtain the following result.

Theorem 10. *Let a, q be positive integers with $(a, q) = 1$ and let α be a real number such that $|\alpha - a/q| \leqslant 1/q^2$. Then $\sum_{x \leqslant n} \Lambda(x)e(\alpha x)$ and $\sum_{p \leqslant n} \log p \, e(\alpha p)$ are both at most $50(\log n)^4(n^{1/2}q^{1/2} + n^{4/5} + nq^{-1/2})$, when n is sufficiently large.* □

We have now managed to show that $f(\alpha)$ is small, provided that q is not too small. The usual approach to the rest of the proof is to estimate $f(\alpha)$ when α *is* close to a rational with small denominator, using the Siegel-Walfisz theorem, and then combine these results to obtain a fairly accurate estimate for $\int f(\alpha)^3 e(-\alpha n)\, d\alpha$ (in particular, accurate enough to show that it is non-zero). In these notes, I use a different argument, which I believe explains in a more intuitive way why the integral comes out to be positive. It has the added advantage that we do not actually need to estimate the integral at all accurately, although it is possible to work harder in order to do so.

The main idea is to work out exactly what is meant by the familiar idea that the primes are somehow randomly distributed. A minor problem to worry about first is that there are more small primes than large ones, but we have already dealt with that by weighting a prime p by $\log p$. Now, in §2, we thought of a subset A of $\{1, 2, \ldots, n\}$ as being random if the Fourier coefficients $\hat{A}(r)$ were all much smaller than n, for non-zero r. However, it is clear that the primes are not random in this sense, because, for example, only one prime is a multiple of five.

Which constraints of this kind have an effect on Fourier coefficients? It is an easy exercise to show that congruence conditions mod q have an effect if and only if q is small. Motivated by this observation, we let $p_1, \dots, p_k$ be the primes less than or equal to $(\log n)^A$, in ascending order, and define Q to be the set of integers less than or equal to n that are not multiples of any p_i. Here, A is an absolute constant (in fact we shall choose $A = 16$), but there is some freedom in the argument, and we could have made p_k quite a bit larger. What we shall do in the rest of the section is show that the weighted primes behave like a random subset of Q.

It is not hard to work out how to interpret this statement. It means that the Fourier transforms $f(\alpha) = \sum_{p \leqslant n} \log p \, e(\alpha p)$ and $h(\alpha) = \sum_{x \in Q} e(\alpha x)$ are roughly proportional. This implies that integrals involving these functions are also roughly proportional, so that, roughly speaking, whatever is true for Q is true for the weighted primes as well. (That "roughly speaking" is important: a good exercise is to see why Lemma 20 below does not translate into a solution of the Goldbach conjecture.)

We begin by obtaining an estimate similar to Theorem 10 for the function $h(\alpha)$. The proof is much simpler, however.

Lemma 11. *Suppose that $(a, q) = 1$ and $|\alpha - a/q| \leqslant q^{-2}$. Then*

$$|h(\alpha)| \leqslant 100(\log n)^2 (n^{1/2} + q + nq^{-1} + n^{1-1/4A}) \; .$$

Proof. Notice first that

$$h(\alpha) = \sum_{s=0}^{k} (-1)^s \sum_{1 \leqslant i_1 < \dots < i_s \leqslant k} \; \sum_{y \leqslant n/p_{i_1} \dots p_{i_s}} e(\alpha p_{i_1} \dots p_{i_s} y) \; .$$

The justification of this is similar to the proof of Lemma 2. If $z \in Q$ then $e(\alpha z)$ is added when $s = 0$, and otherwise does not appear. If $z \notin Q$ then $z = p_{j_1}^{a_1} \dots p_{j_r}^{a_r} w$ for some $w \in Q$, and $a_i \geqslant 1$, and $e(\alpha z)$ is added $(-1)^{|B|}$ times for every subset B of $\{j_1, \dots, j_r\}$, giving a total contribution of zero.

The inner sum is at most $\min\{\|\alpha p_{i_1} \dots p_{i_s}\|^{-1}, n/p_{i_1} \dots p_{i_s}\}$. Let $t = \log n / 2A \log\log n$ and note that $p_k^t \leqslant \sqrt{n}$. These estimates and the fundamental theorem of arithmetic imply that

$$\Big|\sum_{s=0}^{t}(-1)^s \sum_{1\leqslant i_1<\ldots<i_s\leqslant k} \ \sum_{y\leqslant n/p_{i_1}\ldots p_{i_s}} e(\alpha p_{i_1}\ldots p_{i_s}y)\Big|$$

is at most $\sum_{x\leqslant\sqrt{n}}\min\{\|\alpha x\|^{-1},n/x\}$, which, by Lemma 5, is at most $100(\log n)^2(n^{1/2}+q+nq^{-1})$.

The rest of the sum is, in modulus, at most

$$\sum_{s=t+1}^{k}\ \sum_{1\leqslant i_1<\ldots<i_s\leqslant k} n\prod_{j=1}^{s}p_{i_j}^{-1}\ ,$$

which is at most

$$n\sum_{s=t+1}^{k}(s!)^{-1}(p_1^{-1}+\ldots+p_k^{-1})^s\ .$$

It is well known (and follows from the prime number theorem) that $p_1^{-1}+\ldots+p_k^{-1}$ is about $\log\log k$, and so at most $2\log\log\log n$, when n is sufficiently large. Approximating $s!$ by $(s/e)^s$, we obtain an upper bound of $2n(2e\log\log\log n/t)^t$, since $t\geqslant 4e\log\log\log n$. It is not hard to check that this is at most $n^{-1/4A}$ when n is sufficiently large. This, together with the first estimate, proves the lemma. □

We now turn to the "major-arcs" estimates, that is, estimates for $f(\alpha)$ and $h(\alpha)$ when α is close to a rational with small denominator. It turns out that such estimates are more or less equivalent to estimating $\sum_{p\in X}\log p$ and $|X\cap Q|$ for certain long arithmetic progressions X. In the case of the primes themselves, we shall appeal to known estimates of this type, as given in the next result, the Siegel-Walfisz theorem.

Theorem 12. *Let A be a positive real number, let x be an integer, let $q\leqslant(\log x)^A$ be another integer and let $(a,q)=1$. Then*

$$\sum_{p\leqslant x,p\equiv a\ (q)}\log p=\frac{x}{\phi(q)}+O(\exp(-C\sqrt{\log x}))\ ,$$

where C is a constant depending on A only. □

Notice that from Theorem 12 it follows that, if $q\leqslant(\log n)^A$, and X is the arithmetic progression $\{a,a+q,\ldots,a+(m-1)q\}$, where $(a,q)=1$ and $1\leqslant a\leqslant n-(m-1)q$, then for any constant B, we have

$$\sum_{p\in X}\log p=\frac{mq}{\phi(q)}+O(n/(\log n)^B)\ ,$$

with the implied constant in the error term depending on A and B only.

We shall now obtain an estimate for $|X \cap Q|$, when X is an arithmetic progression of the kind above.

Lemma 13. *Let $q \leqslant (\log n)^A$, let $X = \{a, a+q, \ldots, a+(m-1)q\}$ be a subset of $[N]$ with $m \geqslant N^{1/2}$ and suppose that $(q,a) = 1$. Then*

$$|X \cap Q| = \frac{mq}{\phi(q)} \prod_{i=1}^{k} (1 - p_i^{-1}) + O\big(mn^{-1/4A}\big) \ .$$

Proof. Let $x \in X$ be chosen uniformly at random, and for each i let X_i be the event $p_i | x$. Then the probability of X_i is $p_i^{-1} + O(m^{-1})$ if $p_i \not| q$ and $O(m^{-1})$ if $p_i | q$. More generally, for any choice $1 \leqslant i_1 < \ldots < i_s \leqslant k$ we have

$$\mathrm{Prob}(X_{i_1} \cap \ldots \cap X_{i_s}) = \prod_{j=1}^{s} \epsilon_{i_j}/p_{i_j} + O(m^{-1}) \ ,$$

where $\epsilon_i = 1$ if $p_i \not| q$ and 0 if $p_i | q$. It follows from this and the inclusion-exclusion formula that, for any t,

$$1 - \mathrm{Prob}\Big(\bigcup_{i=1}^{k} X_i\Big) = \sum_{s=0}^{t} (-1)^s \sum_{1 \leqslant i_1 < \ldots < i_s \leqslant k} \prod_{j=1}^{s} \epsilon_{i_j}/p_{i_j} + O(m^{-1}) \sum_{s=1}^{t} \binom{k}{s} \ .$$

Now

$$\prod_{i=1}^{k} (1 - \epsilon_i/p_i) = \sum_{s=0}^{k} (-1)^s \sum_{1 \leqslant i_1 < \ldots < i_s \leqslant k} \prod_{j=1}^{s} \epsilon_{i_j}/p_{i_j}$$

and

$$\begin{aligned} \sum_{1 \leqslant i_1 < \ldots < i_s \leqslant k} \prod_{j=1}^{s} \epsilon_{i_j}/p_{i_j} &\leqslant (s!)^{-1}(p_1^{-1} + \ldots + p_k^{-1})^s \\ &\leqslant (4e \log\log\log n/s)^s \end{aligned}$$

when n is sufficiently large. If $t \geqslant 8e \log\log\log n$, then this quantity summed from $t+1$ to k is at most $(4e \log\log\log n/t)^t$. Furthermore, $\sum_{s=1}^{t} \binom{k}{s}$ is easily seen to be at most k^t. It follows that

$$1 - \mathrm{Prob}\Big(\bigcup_{i=1}^{k} X_i\Big) = \prod_{i=1}^{k} (1 - \epsilon_i/p_i) + O\big((\log n)^{At} + (4e \log\log\log n/t)^t\big) \ .$$

Choosing t to be $\log n/2A\log\log n$ gives an error of at most $O(n^{-1/4A})$, as in the proof of Lemma 11. Note finally that

$$\begin{aligned}\prod_{i=1}^k(1-\epsilon_i/p_i) &= \prod_{i=1}^k(1-1/p_i)\prod_{p_i|q}(1-1/p_i)^{-1}\\ &= \prod_{i=1}^k(1-1/p_i)\prod_{p|q}(1-1/p)^{-1}\\ &= \frac{q}{\phi(q)}\prod_{i=1}^k(1-p_i^{-1})\ .\end{aligned}$$

Multiplying everything by m proves the lemma. □

Corollary 14. *Let a,q,X be as in Lemma 13, let $K=\prod_{i=1}^k(1-p_i^{-1})^{-1}$ and let B be any positive constant. Then*

$$K|X\cap Q|-\sum_{p\in X}\log p=O(n(\log n)^{-B})\ .$$

Proof. This follows immediately from Lemma 13 and the remark following Lemma 12. (Strictly speaking one must consider what happens if $(a,q)\neq 1$ but then it is easy to see that both $K|X\cap Q|$ and $\sum_{p\in X}\log p$ are very small.) □

I have a more streamlined (and rigorous) presentation of the next part of the proof than I gave in lectures. But once again I stress that something very simple is going on, and the best way to understand this is to prove it for yourself.

Lemma 15. *Let $q\leqslant(\log n)^A$, let $(b,q)=1$ and let α be a real number such that $|\alpha-b/q|\leqslant(\log n)^A/qn$. Let G be a function from $\{1,2,\ldots,n\}$ to $\mathbb{R}$ such that $|G(x)|\leqslant\log n$ for every x and such that*

$$\Big|\sum_{x\in X}G(x)\Big|=O(n(\log n)^{-B})$$

for every arithmetic progression $X=\{a,a+q,\ldots,a+(m-1)q\}$, where $B\geqslant 4A+2$. Then

$$\Big|\sum_{x\leqslant n}G(x)e(\alpha x)\Big|=O(n(\log n)^{-A})\ .$$

Proof. Let $\beta=\alpha-b/q$ and let X be one of the arithmetic progressions of the above type. Notice that, if $x,y\in X$, then

$$|e(\beta x)-e(\beta y)|=|1-e(\beta(x-y))|\leqslant 2\pi|x-y||\beta|\leqslant 2\pi m(\log n)^A/n\ .$$

Therefore, letting x_0 be an arbitrary element of X, we have

$$\begin{aligned}\Big|\sum_{x\in X} G(x)e(\alpha x)\Big| &= \Big|\sum_{x\in X} G(x)e(bx/q)e(\beta x)\Big| \\ &\leqslant \Big|e(ab/q)\sum_{x\in X} G(x)(e(\beta x)-e(\beta_0 x))\Big| + \Big|e(ab/q)e(\beta x_0)\sum_{x\in X} G(x)\Big| \\ &= \Big|\sum_{x\in X} G(x)(e(\beta x)-e(\beta x_0))\Big| + \Big|\sum_{x\in X} G(x)\Big| \\ &\leqslant (2\pi m(\log n)^A/n)m\log n + O(n(\log n)^{-B}) \\ &= O\big((\log n)^{A+1}m^2n^{-1} + n(\log n)^{-B}\big)\ .\end{aligned}$$

But we can partition $[n]$ into $2n/m_0$ arithmetic progressions of the form of X, with $m \leqslant m_0$ in each case. Therefore, choosing $m_0 = n(\log n)^{-B/2}$ and summing over all these, we find that

$$\Big|\sum_{x\leqslant n} G(x)e(\alpha x)\Big| = O\big(n(\log n)^{A+1-B/2}\big)$$

which proves the result. □

Recall that $f(\alpha) = \sum_{p\leqslant n}\log p\, e(\alpha p)$. Let us define $h_1(\alpha)$ to be $K\sum_{x\in Q} e(\alpha x) = Kh(\alpha)$.

Corollary 16. *Let $A = 16$. Then, for every real number α, $f(\alpha) - h_1(\alpha) = O(n(\log n)^{-A/4})$.*

Proof. Let α be a real number. Then we can find $q \leqslant n(\log n)^{-A}$ and b with $(b,q) = 1$ such that $|\alpha - b/q| \leqslant (\log n)^A/nq$. If $q \geqslant (\log n)^A$, then Theorem 10 implies that $f(\alpha) = O(n(\log n)^{4-A/2})$, while Lemma 11 (with an easy estimate for K) implies that $h_1(\alpha) = O(n(\log n)^{3-A})$, so the result holds.

If on the other hand $q \leqslant (\log n)^A$, then set $G(x) = \log x - KQ(x)$ if x is prime, and $-KQ(x)$ otherwise. Corollary 14 tells us that G satisfies the conditions for Lemma 15. But $\sum_{x\leqslant n} G(x)e(\alpha x) = f(\alpha) - h_1(\alpha)$, so Lemma 15 gives us the result in this case. □

This is all we need for the three-primes theorem. However, it is perhaps of some interest to obtain an actual estimate for $f(\alpha)$ and $h_1(\alpha)$ when q is small, rather than merely showing that they are close. So the next two lemmas are here for interest only.

For notational convenience, when we write $(a,q) = 1$ in the next lemma we shall mean that a and q are coprime and that $1 \leqslant a \leqslant q$.

Lemma 17. *For every q, $\sum_{(a,q)=1} e(a/q) = \mu(q)$.*

Proof. If $q = 1$ then the result holds. If q is a prime, then

$$\sum_{(a,q)=1)} e(a/q) = \sum_{1 \leqslant a < q} e(a/q) = 0 - 1 = -1 \ .$$

If $q = p^k$ with p prime and $k \geqslant 2$, then

$$\sum_{(a,q)=1} e(a/q) = \sum_{1 \leqslant a \leqslant q} e(a/q) - \sum_{1 \leqslant b \leqslant p^{k-1}} e(b/p^{k-1}) = 0 - 0 = 0 \ .$$

Finally, if q and r are coprime, then

$$\sum_{(a,q)=1} e(a/q) \sum_{(b,r)=1} e(b/r) = \sum_{(a,q)=1,(b,r)=1} e(ar + bq/qr) \ .$$

But $ar + bq$ runs through all residues mod qr, and $(ar + bq, qr) = 1$ if and only if $(a,q) = 1$ and $(b,r) = 1$. So the sum is $\sum_{(a,qr)=1} e(a/qr)$.

These properties of the left hand side force it to equal μ. □

Now, given $q \leqslant (\log n)^A$, let us define a function $H_q : [n] \to \mathbb{R}$ by letting $H_q(x)$ equal $q/\phi(q)$ if $(x,q) = 1$ and zero otherwise.

Lemma 18. *Let $q \leqslant (\log n)^A$, let $(b,q) = 1$ and let α be a real number such that $|\alpha - b/q| \leqslant (\log n)^A/nq$. Let $\beta = \alpha - b/q$. Then*

$$\sum_{x \leqslant n} H_q(x) e(\alpha x) = \frac{\mu(q)}{\phi(q)} \sum_{x \leqslant n} e(\beta x) + O((\log n)^{2A}) \ .$$

Proof. Let us write X_a for the set of integers less than or equal to n and congruent to a mod q. If $(a,q) \neq 1$, then clearly $\sum_{x \in X_a} H_q(x) e(\alpha x) = 0$. On the other hand, if $(a,q) = 1$, then

$$\begin{aligned} \sum_{x \in X_a} H_q(x) e(\alpha x) &= \frac{q}{\phi(q)} \sum_{x \in X_a} e(bx/q) e(\beta x) \\ &= \frac{q}{\phi(q)} e(ab/q) \sum_{x \in X_a} e(\beta x) \ . \end{aligned}$$

Now, if $a_1, a_2 \leqslant q$, then

$$\Big| \sum_{x \in X_{a_1}} e(\beta x) - \sum_{x \in X_{a_2}} e(\beta x) \Big| \leqslant 1 + \Big| \sum_{x \in X_{a_1}} e(\beta x) \Big| |1 - e(\beta(a_1 - a_2))| \ .$$

Since $|a_1 - a_2| \leqslant q$, we know that $1 - e(\beta(a_1 - a_2)) = O((\log n)^A/n)$, so this shows that, for every a,

$$\sum_{x \in X_a} e(\beta x) = q^{-1} \sum_{x \leqslant n} e(\beta x) + O((\log n)^A) \ .$$

(In words, the numbers $\sum_{x \in X_a} e(\beta x)$ are all approximately equal, and therefore all approximately equal to their average.) It follows that

$$\sum_{0 \leqslant a < q} \sum_{x \in X_a} H_q(x) e(\alpha x) = \frac{q}{\phi(q)} \sum_{(a,q)=1} e(ab/q) \Big(q^{-1} \sum_{x \leqslant n} e(\beta x) + O(\log n)^A \Big) \ .$$

Since $(b, q) = 1$, the result follows from Lemma 17. □

Corollary 19. *Let α, b, q and β be as in Lemma 18. Then $f(\alpha)$ and $h_1(\alpha)$ are both equal to $(\mu(q)/\phi(q)) \sum_{x \leqslant n} e(\beta x) + O(n(\log n)^{-A})$.*

Proof. This follows easily from Theorem 12 and Lemmas 13, 15 and 18. Let $P(x)$ be the function $\log x$ if x is prime and zero otherwise. Setting $G(x) = P(x) - H_q(x)$, Theorem 12 tells us that the conditions for Lemma 15 are satisfied. But this implies that $f(\alpha) = \sum_{x \leqslant n} H_q(x) e(\alpha x) + O(n(\log n)^{-A})$. Then Lemma 18 gives us our estimate for $f(\alpha)$. The same argument works for $h_1(\alpha)$ if we use Lemma 13 instead of Theorem 12. □

After that diversion, let us now finish the proof of the three-primes theorem. There are two steps to the proof. First, we show that every sufficiently large odd integer is the sum of three elements of Q (or fake primes) in many ways, using the Brun sieve once again. Then we deduce, from the fact that f and h_1 are uniformly close, that the same is true of the genuine primes.

Lemma 20. *Let m be an integer. Then the number of ways of writing $m = x + y$ with x and y both in Q is at least $m \prod_{i=1}^k (1 - r_i/p_i) + O(m^{-1} n^{1/2} + m n^{-1/4A})$, where $r_i = 1$ if $p_i | m$ and 2 otherwise.*

Proof. Choose x randomly and uniformly from the set $[m]$. For each i let X_i be the event that $p_i | x$ or $p_i | m - x$. As in the proof of Lemma 13, it is easy to show that $\mathrm{Prob}(X_i) = r_i/p_i + O(m^{-1})$. (The point about the r_i is that the events $p_i | x$ and $p_i | m - x$ are the same if $p_i | m$ and mutually exclusive otherwise.) More generally, it is not hard to show that

$$\mathrm{Prob}(X_{i_1} \cap \ldots \cap X_{i_s}) = \prod_{j=1}^s \frac{r_{i_j}}{P_{i_j}} + O(m^{-1}) \ .$$

Therefore, by the inclusion-exclusion formula,

$$1-\mathrm{Prob}\Big(\bigcup_{i=1}^k X_i\Big)=\sum_{s=0}^t(-1)^s\sum_{1\leqslant i_1<\ldots<i_s\leqslant k}\prod_{j=1}^s r_{i_j}/p_{i_j}+O(m^{-1})\sum_{s=1}^t\binom{k}{s}\ .$$

But

$$\prod_{i=1}^k(1-r_i/p_i)=\sum_{s=0}^k(-1)^s\sum_{1\leqslant i_1<\ldots<i_s\leqslant k}\prod_{j=1}^s r_{i_j}/p_{i_j}$$

and

$$\begin{aligned}\sum_{1\leqslant i_1<\ldots<i_s\leqslant k}\prod_{j=1}^s r_{i_j}/p_{i_j}&\leqslant(s!)^{-1}(2p_1^{-1}+\ldots+2p_k^{-1})^s\\&\leqslant(8e\log\log\log n/s)^s\ .\end{aligned}$$

As in the proof of Lemma 13, it follows that

$$1-\mathrm{Prob}\Big(\bigcup_{i=1}^k X_i\Big)=\prod_{i=1}^k(1-r_i/p_i)+O\big(m^{-1}(\log n)^{At}+(8e\log\log\log n/t)^t\big)$$

for any $t\geqslant 16e\log\log\log n$. Choosing t to be $\log n/2A\log\log n$ implies the lemma. □

On writing out the next corollary it occurs to me that I can't have done it correctly in lectures, where I think I absent-mindedly estimated the number of ways of writing n as $x+y+z$ with only x and y in Q.

Corollary 21. *If n is sufficiently large and odd, then the number of ways of writing n as the sum of three elements of Q is at least $(n^2/16)K^{-1}\prod_{i=2}^k(1-2p_i^{-1})$.*

Proof. Note first that Lemma 13 implies that the number of elements of Q less than or equal to $n/2$ is at least $K^{-1}n/4$ (when n is sufficiently large). For every odd $z\leqslant n/2$, the number of ways of writing $n-z$ as the sum of two elements of Q is, by Lemma 20, at least $(n/4)\prod_{i=2}^k(1-2/p_i)$. The result follows. □

It is possible to be much more careful and work out the number of ways of writing n as the sum of three elements of Q to within a factor $1+o(1)$, but we do not need this.

Theorem 22. *(Vinogradov) Every sufficiently large odd integer is the sum of three primes.*

Proof. Note first that $(16K)^{-1}\prod_{i=2}^k(1-2p_i^{-1})$ is easily shown to be at least $(\log n)^{-1}$ when n is sufficiently large, so the number of ways of writing n as the sum of three elements of Q is at least $n^2/\log n$. On the other hand, it is also $\int h(\alpha)^3 e(-\alpha n)\,d\alpha$, so we certainly have $\int h_1(\alpha)^3 e(-\alpha n)\,d\alpha \geqslant n^2/\log n$.

As we commented at the beginning, it is enough for our purposes to show that $\int f(\alpha)^3 e(-\alpha n)\,d\alpha \neq 0$. But, by Corollary 16,

$$\begin{aligned}\Big|\int f(\alpha)^3 e(-\alpha n)\,d\alpha - \int h_1(\alpha)^3 e(-\alpha n)\,d\alpha\Big| \\ &= O\big(n(\log n)^{-A/4}\big)\int |f(\alpha)^2 + f(\alpha)h_1(\alpha) + h_1(\alpha)^2|\,d\alpha \\ &= O\big(n(\log n)^{-A/4}\big)\int |f(\alpha)|^2 + |h_1(\alpha)|^2\,d\alpha \\ &= O\big(n(\log n)^{-A/4}\big)\Big(\sum_{p\leqslant n}(\log p)^2 + K^2|Q|\Big) \\ &= O\big(n^2\log n(\log n)^{-A/4}\big)\ .\end{aligned}$$

Since we chose A to be 16, this and our estimate for the integral with h_1 are enough to prove the theorem. □