Homework 3 solutions

Additive number theory seminar

Problem 1. Let $h \geq 3$ be an integer, and $N \geq 2$ an integer such that $N \equiv h \pmod 2$, i.e. N and h are either both odd or both even. Using similar methods and lemmas to the proof of Vinogradov's theorem in Nathanson, find an asymptotic formula for the number of ways $r_h(N)$ that N can be written as a sum of h primes.

Solution. There are three major components to Vinogradov's formula: the major arc decomposes into the product of the singular series (which we need to bound) and the singular integral (which we need to estimate), and the minor arc gives an error term we need to bound.

Following Nathanson, we first worry about the singular series. Let

$$\mathfrak{S}(N, Q) = \sum_{q \leq Q} \frac{\mu(q)^h c_q(N)}{\varphi(q)^h}$$

where φ is Euler's totient function and

$$c_q(N) = \sum_{\substack{a=1 \\ (a,q)=1}}^{q} e(aN/q),$$

and write $\mathfrak{S}(N) = \lim_{Q\to\infty} \mathfrak{S}(N, Q)$. (Notice that $\mu(q)^h$ is $\mu(q)$ if h is odd and $\mu(q)^2$ if h is even, i.e. just detecting squarefreeness.)

Proposition 1. *The singular series $\mathfrak{S}(N)$ converges absolutely and uniformly in N, and has the Euler product*

$$\mathfrak{S}(N) = \prod_{p \nmid N} \left(1 - \frac{1}{(1-p)^h}\right) \prod_{p \mid N} \left(1 - \frac{1}{(1-p)^{h-1}}\right),$$

and there exist positive constants c_1, c_2 such that $c_1 < \mathfrak{S}(N) < c_2$ for all positive N. For any $\epsilon > 0$,

$$\mathfrak{S}(N, Q) = \mathfrak{S}(N) + O(Q^{-h+2+\epsilon}).$$

The proof is very similar to that of Theorem 8.2 in Nathanson, but we sketch it for clarity:

Proof. Since $c_q(N) \ll \varphi(q)$ (since it is a sum with $\varphi(q)$ terms each of absolute value 1), the absolute value of each summand is bounded by $\frac{1}{\varphi(q)^{h-1}}$. We have $\varphi(q) \gg q^{1-\epsilon}$ for all $\epsilon > 0$ and q sufficiently large, so

$$\sum_{q=1}^{\infty} \left| \frac{\mu(q) c_q(N)}{\varphi(q)^h} \right| \ll \sum_{q=1}^{\infty} \frac{1}{q^{(h-1)(1-\epsilon)}}$$

which converges since $h \geq 3$, so $\mathfrak{G}(N)$ converges absolutely (and uniformly in N, since these bounds are uniform in N). By taking only the terms with $q > Q$ we get

$$\mathfrak{G}(N) - \mathfrak{G}(N,Q) \ll \sum_{q>Q} \frac{1}{q^{(h-1)(1-\epsilon)}} \ll \int_Q^\infty \frac{1}{q^{(h-1)(1-\epsilon)}}\, dq \ll Q^{-h+2+(h-1)\epsilon},$$

which recovers the claimed bound by replacing ϵ by $\epsilon/(h-1)$.

The Euler product factors follows from the description of $c_q(N)$ in Nathanson and the theory of Dirichlet series for multiplicative functions, exactly as there (replacing 3 by h and 2 by $h-1$); we won't use this description and so don't dwell on it. However, it does mean that if N and h are both odd, we can write

$$\mathfrak{G}(N) = \prod_p \left(1 + \frac{1}{(p-1)^h}\right) \cdot \prod_{p|N} \frac{1-(p-1)^{-h+1}}{1+(p-1)^{-h}},$$

so the first factor is independent of N and the second factor satisfies

$$\prod_{p\geq 3} \frac{1-(p-1)^{-h+1}}{1+(p-1)^{-h}} < \prod_{p|N} \frac{1-(p-1)^{-h+1}}{1+(p-1)^{-h}} < 1$$

(since $2 \nmid N$, so we can exclude it), and then it is straightforward to check that the left-hand side converges and so is positive so we get the desired constants c_1 and c_2 since neither bound depends on N. If h and N are both even, we work similarly except handling the prime 2 separately:

$$\mathfrak{G}(N) = \prod_{p\geq 3} \left(1 - \frac{1}{(p-1)^h}\right) \cdot 2 \cdot \prod_{\substack{p|N \\ p\geq 3}} \frac{1+(p-1)^{-h+1}}{1-(p-1)^{-h}}$$

and so again the first term is independent of N and the last is bounded by

$$1 < \prod_{\substack{p|n \\ p\geq 3}} \frac{1+(p-1)^{-h+1}}{1-(p-1)^{-h}} < \prod_{p\geq 3} \frac{1+(p-1)^{-h+1}}{1-(p-1)^{-h}}$$

which again converges and so we get the desired bounds. $\square$

We'll take the same major arc/minor arc decomposition as Nathanson, so for fixed Q, N as above the major arc $\mathfrak{M}$ is given by the union over $1 \leq q \leq Q$, $1 \leq a \leq q$ with $(a,q) = 1$ of $\alpha \in [0,1]$ such that $|\alpha - a/q| \leq \frac{Q}{N}$. Let

$$R_h(N) = \sum_{p_1+\cdots+p_h=N} \log p_1 \cdot \log p_2 \cdots \log p_h,$$

and

$$F(\alpha) = \sum_{p\leq N} e(p\alpha) \log p.$$

By the usual circle method argument

$$R_h(N) = \int_0^1 F(\alpha)^h e(-N\alpha)\,d\alpha = \int_{\mathfrak{M}} F(\alpha)^h e(-N\alpha)\,d\alpha + \int_{\mathfrak{m}} F(\alpha)^h e(-N\alpha)\,d\alpha.$$

We now turn towards estimating the major arc term.

Let

$$u(\beta) = \sum_{m=1}^{N} e(m\beta)$$

and

$$J(N) = \int_{-1/2}^{1/2} u(\beta)^h e(-N\beta)\,d\beta.$$

Exactly as in the proof of Lemma 8.1 in Nathanson, we see (essentially this is a formulation of the circle method for sums of integers) that this is the number of ways of writing N as the sum of h integers, which we know from Theorem 5.1 is

$$\binom{N-1}{h-1} = \frac{N^{h-1}}{(h-1)!} + O(N^{h-2}).$$

We can now prove our next result:

Proposition 2. *For any $B, C, \epsilon > 0$ with $C > 2B$, $Q = (\log N)^B$, the major arc contribution is*

$$\int_{\mathfrak{M}} F(\alpha)^h e(-N\alpha)\,d\alpha = \mathfrak{G}(N)\frac{N^{h-1}}{(h-1)!} + O\left(\frac{N^{h-1}}{(\log N)^{(1-\epsilon)B}}\right) + O\left(\frac{N^{h-1}}{(\log N)^{C-5B}}\right).$$

The proof is almost identical to that of Theorem 8.4 in Nathanson. The key point is that we have unchanged the result of Lemma 8.3: if α is in the (a,q) component of the major arc and $\beta = \alpha - a/q$, then

$$F(\alpha) = \frac{\mu(q)}{\varphi(q)}u(\beta) + O\left(\frac{Q^2 N}{(\log N)^C}\right)$$

so

$$F(\alpha)^h = \frac{\mu(q)^h}{\varphi(q)^h}u(\beta)^h + O\left(\frac{Q^2 N^h}{(\log N)^C}\right).$$

Integrating over the major arc, one gets a bound on

$$\sum_{\substack{q\le Q\\ a\le q\\ (a,q)=1}}^{q} \int_{\mathfrak{M}(a,q)} F(\alpha)^h e(-N\alpha) - \frac{\mu(q)^h}{\varphi(q)^h}u\left(\alpha - \frac{a}{q}\right)^h d\alpha = \int_{\mathfrak{M}} F(\alpha)^h e(-N\alpha)\,d\alpha - \mathfrak{G}(N,Q)J(N).$$

Since we can estimate $J(N)$ well, the result follows.

The last thing we need to do is bound the minor arc contribution. Since this mostly comes down to bounding $F(\alpha)$, which is already done for us (Theorem 8.5 in Nathanson), this is not too bad:

Proposition 3. *For any $B > 0$, we have*

$$\int_{\mathfrak{m}} F(\alpha)^h e(-N\alpha)\, d\alpha \ll \frac{N^{h-1}}{(\log N)^{(B/2-4)(h-2)-1}}.$$

Again, the proof is very similar to Nathanson's Theorem 8.6. Exactly as there, we have

$$\int_{\mathfrak{m}} |F(\alpha)|^2\, d\alpha \ll N \log N,$$

and so the minor arc integral is bounded by

$$\sup_{\alpha\in\mathfrak{m}} |F(\alpha)|^{h-2} \cdot N \log N.$$

By Nathanson's Theorem 8.5, $F(\alpha) \ll \frac{N}{(\log N)^{B/2-4}}$ and so $|F(\alpha)|^{h-2} \ll \frac{N^{h-2}}{(\log N)^{(B/2-4)(h-2)}}$, so in all we get the claimed bound.

In particular, taking B sufficiently large in Proposition 3 and using the fact that $\mathfrak{G}(N)$ is bounded between two positive constants, we see that the major arc term is much larger than the minor arc term as $N \to \infty$, and so (relabeling our constants) combining all three propositions we find that

$$R_h(N) = \mathfrak{G}(N)\frac{N^{h-1}}{(h-1)!} + O\left(\frac{N^{h-1}}{(\log N)^A}\right)$$

for every $A > 0$.

To conclude, we want to get a formula for $r_h(N)$ rather than $R_h(N)$. We relate them via one final proposition:

Proposition 4. *We have*

$$R_h(N) = r_h(N)(\log N)^h + O\left(\frac{N^{h-1}\log\log N}{\log N}\right).$$

Proof. Recall

$$R_h(N) = \sum_{p_1+\cdots+p_h=N} \log p_1 \cdots \log p_h, \qquad r_h(N) = \sum_{p_1+\cdots+p_h=N} 1,$$

so since each $p_i \le N$ we have $R_h(N) \le (\log N)^h r_h(N)$. For $0 \le \delta < 1$, let $r_{h,\delta}(n)$ denote the number of ways we can write N as a sum of h primes $p_1 + \cdots + p_h = N$ such that at least one p_i is bounded by $N^{1-\delta}$. There are h options for which p_i will be bounded; let's say it's p_1. Then this is really the number of ways of choosing $p_1 \le N^{1-\delta}$, $p_2, \ldots, p_{h-1} \le N$ prime such that $N - p_1 - \cdots - p_{h-1}$ is also prime. This is bounded by the number of such primes, dropping the last condition, i.e. $\pi(N^{1-\delta}) \cdot \pi(N)^{h-2}$; allowing any p_i to be the bounded one instead of just p_1, we get

$$r_{h,\delta}(N) \le h\pi(N^{1-\delta})\pi(N)^{h-2} \ll \frac{N^{h-1-\delta}}{(\log N)^{h-1}}.$$

Imposing any conditions on the primes p_i reduces the total, so we have

$$\begin{aligned}
R_h(N) &\geq \sum_{\substack{p_1+\cdots+p_h=N \\ p_i>N^{1-\delta}\ \forall i}} \log p_1 \cdots \log p_i \\
&> \sum_{\substack{p_1+\cdots+p_h=N \\ p_i>N^{1-\delta}\ \forall i}} (\log N^{1-\delta})^h \\
&= (1-\delta)^h (\log N)^h (r_h(N) - r_{h,\delta}(N))
\end{aligned}$$

and so

$$\begin{aligned}
r_h(N)(\log N)^h - R_h(N) &< ((1-\delta)^{-1} - 1)R_h(N) + r_{h,\delta}(N)(\log N)^h \\
&\ll \delta N^{h-1} + N^{h-1-\delta} \log N \\
&= N^{h-1}\left(\delta + \frac{\log N}{N^\delta}\right)
\end{aligned}$$

since as we've seen $R_h(N) \ll N^{h-1}$. Choosing $\delta = \frac{2\log\log N}{\log N}$, we have (identical to Nathanson)

$$\delta + \frac{\log N}{N^\delta} \ll \frac{\log\log N}{\log N}$$

and so, recalling from the start that $r_h(N)(\log N)^h \geq R_h(N)$, we have

$$0 \leq r_h(N)(\log N)^h - R_h(N) \ll N^{h-1}\frac{\log\log N}{\log N}$$

and so the claim follows. □

Combining Proposition 4 with our estimate of $R_h(N)$ yields our final formula:

Theorem 5. *For any $h \geq 3$ and $N \equiv h \pmod 2$,*

$$r_h(N) = \frac{\mathfrak{S}(N)}{(h-1)!} \cdot \frac{N^{h-1}}{(\log N)^h}\left(1 + O\left(\frac{\log\log N}{\log N}\right)\right).$$

The case $h = 3$ recovers Vinogradov's formula.

Problem 2. Explain where your solution to Problem 1 fails for $h = 2$, if you did it, or equivalently where Vinogradov's method fails to prove the strong Goldbach conjecture.

Solution. The method used to show the convergence of the singular series for Proposition 1 fails, but a slightly more careful analysis (using the Euler product expansion) shows that it does converge. The real issue is in the proof of Proposition 3: the $h-2$ factors of $F(\alpha)$ become trivial and so the bound on the minor arc is just $N \log N$, which is larger than the major arc estimate $\mathfrak{S}(N)N$.

Nevertheless, the final statement of the result still makes sense, and one can computationally verify that it seems to hold. In this case, we can rewrite $\mathfrak{G}(N)$ slightly as

$$\mathfrak{G}(N) = \prod_{p\geq 3}\left(1 - \frac{1}{(p-1)^2}\right)\cdot 2\cdot \prod_{\substack{p\mid N\\ p\geq 3}}\frac{p-1}{p-2};$$

the first product, which is independent of N, is the twin prime constant Π_2, which one can obtain via similar methods to problem 2 from homework 2: the heuristic for the number of twin primes up to x is $\pi_2(x) \simeq 2\Pi_2 \cdot \frac{x}{(\log x)^2}$. It is a sign of how closely related Goldbach and twin primes are that it shows up here as well. If we call the final product $\mathfrak{G}_2(N)$, which is the only factor that depends on N, then we expect that

$$r_2(N)\cdot\frac{(\log N)^2}{\mathfrak{G}_2(N)N}$$

approaches a constant, with slowly decreasing error bounded by $\frac{\log\log N}{\log N}$. Below is a graph of the values of this ratio for even N up to 200000, so that you can convince yourself this claim is plausible: