On Square-Primitive Sets with Small Gaps

Xiaoyu He

September 11, 2015

1 Introduction

Call a sequence $S$ of positive integers square-primitive if there do not exist $a, b \in S$ for which $\frac{b}{a}$ is a perfect square greater than 1. We say a sequence has gaps bounded by $g(N)$ if for every $N \geq 1$, the subsequence $S \cap [1, N]$ has no gaps $> g(N)$. We will prove the following.

**Theorem 1.** For any $\varepsilon > 0$, there exists a square-primitive sequence $S$ with gaps bounded by $C(\log N)^2(\log \log N)^\varepsilon$ for some absolute constant $C$ depending only on $\varepsilon$.

In fact, we can think of constructing an arbitrary square-primitive sequence $S$ as follows. For each squarefree integer $m$, consider the set of all its square multiples $T_m = \{ma^2 : a \in \mathbb{N}\}$. The sets $T_m$ are all disjoint as we vary $m$ over all squarefree numbers, and cover all of $\mathbb{N}$. Now each $T_m$ is naturally a poset under divisibility, and to pick $S$ square-primitive is equivalent to independently picking an antichain from each $T_m$. Obviously since every antichain extends to a maximal one this amounts to giving a probability distribution over all of the maximal antichains in $T_m$.

2 The Construction

There are certain distinguished antichains in each $T_m$. Write $A_{m,k}$ to be the antichain of all $n \in T_m$ such that $n/m$ is the square of a product of $k$ not necessarily distinct primes. Thus for example $A_{1,0} = \{1\}$ and $A_{1,1}$ is the set of squares of primes. From each $T_m$ we independently pick one $A_{m,k}$ with probability $C/f(k+1)$ where $f$ is any function for which $\sum_n 1/f(n)$ converges and

$$\sum_{n \geq 1} \frac{1}{f(n)} = \frac{1}{C},$$

and put $A_{m,k}$ in $S$. Now, the probability that a given $n$ is put in $S$ is just $C/f(\Omega_2(n))$, where $\Omega_2(n)$ is the number of primes $p$ such that $p^2|n$, counted by multiplicity, i.e.

$$\Omega_2(n) = \sum_{p|n} \left\lfloor \frac{\nu_p(n)}{2} \right\rfloor.$$
3 The Proof

It remains to show that $S$ has no small gaps with high probability. Let $h = h(x)$ be a slow-growing gap function, i.e. $h \ll x^{1/2-\varepsilon}$ and $h \rightarrow \infty$. It is easy to check that the $[x, x + h)$ intersects every $T_m$ at most once, for all $x$ sufficiently large, since the difference between $ma^2$ and $mb^2$ is at least $m(a + b) \geq \sqrt{\min(ma^2, mb^2)}$. Thus all of the events $n \in S$ are independent when $n \in [x, x + h)$ for $x$ sufficiently large.

We can now say that the probability $S$ misses the entire interval is

$$P[S \cap [x, x + h) = \emptyset] = \prod_{n \in [x, x + h)} \left(1 - \frac{C}{f(\Omega_2(n) + 1)}\right)$$

exactly. We wish to say this probability decreases rapidly as $x \rightarrow \infty$. Now there exists a positive proportion $Ah$ of $n \in [x, x + h)$ such that no $p^2|n$ for $p \leq h$, whenever $h$ is sufficiently large. For the $n$ in this set, we have

$$\Omega_2(n) \leq \frac{\log n}{2 \log h}.$$

Thus, if we also apply $1 - t \leq e^{-t}$ to the previous identity, we get

$$P[S \cap [x, x + h) = \emptyset] \leq \exp\left(-\frac{ACH}{f(\log x/2 \log h + 1)}\right).$$

Thus if we take $h$ to satisfy

$$h \gg \log x \cdot f(\log x/2 \log h + 1)$$

we are immediately done; the probability that $S$ has gaps $> h$ is negligible. In fact if we take

$$h = C' (\log x)^2 (\log \log x)^\varepsilon$$

the result follows from the choice

$$f(n) = n (\log n)^{1+\varepsilon}.$$