DIMACS Technical Report 2000-39
December 2000

Tight Bound for the Density of Sequence of Integers the Sum of No Two of which is a Perfect Square

by

Ayman Khalfalah[1]
Email: akhalfal@cs.rutgers.edu
Address: Dept. of Computer Science
Rutgers University
New Brunswick, New Jersey 08903

Sachin Lodha
Email: lodha@cs.rutgers.edu
Address: Dept. of Computer Science
Rutgers University
New Bruswick, New Jersey 08903

Endre Szemerédi
Email: szemered@cs.rutgers.edu
Address: Dept. of Computer Science
Rutgers University
New Brunswick, New Jersey 08903

[1]This work was supported by DIMACS which was funded by the NSF under grant STC 91-19999.

ABSTRACT

P. Erdős and D. Silverman [EG-80] proposed the problem of determining the maximal density attainable by a set S of positive integers having the property that no two distinct elements of S sum up to a perfect square. J. P. Massias exhibited such a set consisting of all $x \equiv 1 \pmod 4$ with $x \equiv 14, 26, 30 \pmod{32}$ in [M]. In [LOS-82], J. C. Lagarias, A. M. Odylzko and J. B. Shearer showed that for any positive integer n, one cannot find more than $\frac{11}{32}n$ residue classes (mod n) such that the sum of any two is never congruent to a square (mod n), thus essentially proving that the Massias' set has the best possible density. They [LOS-83] also proved that the density of such a set S is never more than 0.475 when we allow general sequences.

We improve on the lower bound for general sequences, essentially proving that it is not 0.475, but arbitrarily close to $\frac{11}{32}$, the same as that for sequences made up of only arithmetic progressions.

1 Introduction

P. Erdős and D. Silverman [EG-80] proposed the problem of determining the maximal density attainable by a set $S = \{s_i\}$ of positive integers having the following property :

$$\text{Property NS} : \forall i \neq j, s_i + s_j \text{ is not a perfect square.}$$

J. P. Massias exhibited such a set consisting of all $x \equiv 1 \pmod 4$ with $x \equiv 14, 26, 30 \pmod{32}$ in [M]. Its density is $\frac{11}{32}$.

In [LOS-82], J. C. Lagarias, A. M. Odylzko and J. B. Shearer proved following theorem

Theorem 1 *Let S be a union of arithmetic progressions (mod M) having the property NS. Then the density $d(S) \leq \frac{11}{32}$ with equality possible if and only if $32|M$. For all other M, $d(S) \leq \frac{1}{3}$.*

Therefore, they essentially showed that the Massias' set has the best possible density if S were to be a union of arithmetic progressions. The same authors [LOS-83] also proved

Theorem 2 *Let S denote a finite set with all elements $\leq N$ which has property NS, and let*

$$d(N) = \max_S \frac{|S|}{N}$$

Then, there exists an absolute constant N_0 such that for all $N > N_0$, $d(N) \leq 0.475$.

It is an improvement over trivial upper bound of 0.5 on $d(N)$, and yet there is a wide gap between the results of theorem 1 and theorem 2.

One must also note that the behavior of sets S having the property NS is quite different from those sets S having following property :

$$\text{Property DS} : \forall i \neq j, s_i - s_j \text{ is not a perfect square.}$$

In [S-78], A. Sárközy proved that any set S having property DS has density 0, to be precise, it has at most $[x(\log\log x)^{2/3}/(\log x)^{1/3}]$ elements $\leq x$.

In this paper, we strengthen theorem 2 and show that the bound in theorem 1 applies in general case too.

Theorem 3 *Let S denote a finite set with all elements $\leq N$ which has property NS, and let*

$$d(N) = \max_S \frac{|S|}{N}$$

Then, for any positive real number δ, there exists an absolute constant $N_0(\delta)$ such that for all $N > N_0(\delta)$, $d(N) < \frac{11}{32} + \delta$.

2 Outline of Proof

We sketch outline of our proof in this section. We are using notation from section 3.

Given a subset S of $[N]$ of density $\frac{11}{32}+\delta$, the number of solutions to the equation $x+y=z^2$ where $x,y \in S$ is equal to

$$\frac{1}{2N}\sum_{t=0}^{2N-1} f_{SQ}(\frac{-t}{2N}) \cdot f_S(\frac{t}{2N}) \cdot f_S(\frac{t}{2N}). \tag{1}$$

Assuming that there is no solution to this equation then the sum in equation (1) must be 0.

Let M be a "highly composite number". We shift S by jM, $0 \le j \le N^{1-\epsilon}$ and then count the number of solutions to the equation $x+y=z^2$ where $x \in S$ and $y \in S+jM$ for each j. We show that it's "almost" the same as in equation (1). The average error is $O(\frac{N\sqrt{N}}{\sqrt{P}})$, where P is the largest prime divisor of M.

Then we partition $[N]$ into M residue classes mod M and observe how well S gets partitioned into these different pieces of $[N]$. If S is "well-distributed" in these residue classes, then the average number of solutions to the equation $x+y=z^2$ where $x \in S$ and $y \in S+jM$ (averaging over $0 \le j \le N^{1-\epsilon}$) is $\Omega(\frac{N\sqrt{N}}{M^2})$. But this is not good enough to get contradiction !

In fact, we will show that we can combinatorially get a "highly composite" M such that

- there exist "quite a few" pairs of "well distributed dense" residue classes modulo M which add up to a quadratic residue modulo M,
- shifting gives at most $O(\frac{N\sqrt{N}}{\sqrt{P}})$ average error in analytical counting
- and combinatorial counting estimates the average number of solutions to be $\Omega(\frac{N\sqrt{N}}{\log^2 P})$.

The last two statements give us a contradiction.

3 Notation

Let N be a sufficiently large positive integer and let S be a subset of $[N]$ of density $\frac{11}{32}+\delta$, δ being a positive constant.

Let $l = \frac{56}{\delta^3}$. For $1 \le i \le l$, let us define primes p_i in following way

$$p_1 = 17, \quad p_{i+1} \ge 2^{16^{p_i}}.$$

Using these primes, we define l numbers, q_1 through q_l

$$q_0 = 1, \quad q_{i+1} = q_i \cdot \prod_{p \text{ prime}, p \le p_{i+1}, p^{\alpha_p} \le {p_{i+1}}^2} p^{\alpha_p}.$$

Let $\sigma = \frac{\frac{11}{32}+\delta-(\frac{11}{32}+\delta)^2}{l}$. Note that

$$\sigma \leq \frac{\frac{1}{4}}{l} = \frac{\delta^3}{224}.$$

Let $\epsilon_{i,j}(S)$ be the density of set S in residue class j modulo q_i. That is

$$\epsilon_{i,j}(S) = \frac{|\{s \in S | s \equiv j \ mod \ q_i\}|}{n/q_i}.$$

Since we will be working with some fixed subset S of $[N]$, we will drop S from above notation and just use $\epsilon_{i,j}$ throughout this paper.

We use the abbreviations

$$e(\alpha) = e^{2\pi i\alpha}.$$

We denote the trigonometrical sum over a set X of integers as

$$f_X(\alpha) = \sum_{x \in X} e(\alpha x).$$

Let SQ be the set of all perfect squares which are less than or equal to $2N$. Note that

$$f_{SQ}(\alpha) = \sum_{s \in SQ} e(\alpha s) = \sum_{x=0}^{\sqrt{2N}} e(\alpha x^2).$$

For sake of simplification of writing, and without loss of generality, we may assume that $\sqrt{2N}$ is an integer.

Notice that given any subset A and B of $[N]$, the number of triplets (x, y, z) such that $x \in A$, $y \in B$ and $x + y = z^2$ is

$$\frac{1}{2N} \sum_{t=0}^{2N-1} f_{SQ}(\frac{-t}{2N}) \cdot f_A(\frac{t}{2N}) \cdot f_B(\frac{t}{2N}).$$

4 Definitions

Definition 1

$$\alpha_i = \frac{\sum_{j=0}^{q_i-1} {\epsilon_{i,j}}^2}{q_i}.$$

Definition 2 *The residue class $j \equiv q_i$ is* full *if $\epsilon_{i,j} = 1 - o(1)$.*

Definition 3 *The residue class $j \equiv q_i$ is* bad *if*

$$\frac{|\{k | 0 \leq k < \frac{q_{i+1}}{q_i} \ and \ |\epsilon_{i,j} - \epsilon_{i+1,j+kq_i}| \geq \delta/4\}|}{q_{i+1}/q_i} \geq \frac{1}{8}.$$

Definition 4 *The residue class $j \equiv q_i$ is* good *if it is not* bad.

5 Lemmas

Lemma 1

$$q_{i+1} \ll 16^{p_{i+1}} \text{ and } 2^{q_i} \ll p_{i+1}.$$

Proof : It could be easily done using induction and left as an exercise to the reader. $\square$

Lemma 2 $\alpha_{i+1} \geq \alpha_i$.

Proof :

$$\alpha_{i+1} - \alpha_i = \frac{\sum_{j=0}^{q_i-1} \sum_{k,k'=0}^{q_{i+1}/q_i-1} (\epsilon_{i+1,j+kq_i} - \epsilon_{i+1,j+k'q_i})^2}{q_{i+1}^2/q_i} \geq 0.$$

$\square$

Lemma 3 $\forall i \;\; (\frac{11}{32}+\delta)^2 \leq \alpha_i \leq \frac{11}{32}+\delta.$

Proof : Using Cauchy-Schwartz inequality, we get

$$\alpha_i = \frac{\sum_{j=0}^{q_i-1} {\epsilon_{i,j}}^2}{q_i} \geq \frac{(\sum_{j=0}^{q_i-1} \epsilon_{i,j})^2}{{q_i}^2} = (\frac{11}{32}+\delta)^2.$$

Since $0 \leq \epsilon_{i,j} \leq 1$, we get

$$\alpha_i = \frac{\sum_{j=0}^{q_i-1} {\epsilon_{i,j}}^2}{q_i} \leq \frac{\sum_{j=0}^{q_i-1} \epsilon_{i,j}}{q_i} = \frac{11}{32}+\delta.$$

Notice that $\alpha_i = \frac{11}{32}+\delta$ will mean that $(\frac{11}{32}+\delta)q_i$ residue classes modulo q_i are *full.* $\square$

Lemma 4 (Improved Cauchy-Schwartz Inequality) *If for the integers* $0 < m < n$*,*

$$\frac{\sum_{k=1}^{m} x_k}{m} = \frac{\sum_{k=1}^{n} x_k}{n} + \gamma,$$

then

$$\sum_{k=1}^{n} {x_k}^2 \geq \frac{(\sum_{k=1}^{n} x_k)^2}{n} + \frac{\gamma^2 mn}{n-m}$$

Proof :

$$\begin{aligned}
\sum_{k=1}^{n} {x_k}^2 &= \sum_{k=1}^{m} {x_k}^2 + \sum_{k=m+1}^{n} {x_k}^2 \\
&\geq \frac{(\sum_{k=1}^{m} x_k)^2}{m} + \frac{(\sum_{k=m+1}^{n} x_k)^2}{n-m} \\
&= (\frac{m(\sum_{k=1}^{n} x_k)^2}{n^2} + \frac{2m\gamma(\sum_{k=1}^{n} x_k)}{n} + m\gamma^2) + \\
&\quad (\frac{(n-m)(\sum_{k=1}^{n} x_k)^2}{n^2} - \frac{2m\gamma(\sum_{k=1}^{n} x_k)}{n} + \frac{m^2\gamma^2}{n-m}) \\
&= \frac{(\sum_{k=1}^{n} x_k)^2}{n} + \frac{\gamma^2 mn}{n-m}.
\end{aligned}$$

□

Lemma 5 *If $\alpha_{i+1} - \alpha_i < \sigma$, then the number of* bad *residue classes mod q_i is less than $\frac{\delta q_i}{2}$.*

Proof :

$$\alpha_{i+1} - \alpha_i = \frac{1}{q_{i+1}} \sum_{j=0}^{q_i - 1} (\sum_{k'=0}^{q_{i+1}/q_i - 1} \epsilon^2_{i+1,k'+j*q_i} - \frac{(\sum_{k'=0}^{q_{i+1}/q_i - 1} \epsilon_{i+1,j+k'q_i})^2}{q_{i+1}/q_i})$$

Let number of bad residue classes mod q_i be B. Then using lemma 4 with $n = \frac{q_{i+1}}{q_i}$, $m = \frac{n}{8}$ and $\gamma = \frac{\delta}{4}$, we get that

$$\alpha_{i+1} - \alpha_i \geq \frac{1}{q_{i+1}} \cdot B \cdot \frac{q_{i+1}\delta^2}{112 q_i} = \frac{B\delta^2}{112 q_i}$$

Since this difference was less than σ, we get

$$B < \frac{112\sigma q_i}{\delta^2} < \frac{\delta q_i}{2}.$$

□

Lemma 6 *If $\alpha_{i+1} - \alpha_i < \sigma$, then there exists a pair of* good *residue classes, say $a \neq b$, modulo q_i such that*

1. $\epsilon_{i,a} \geq \frac{\delta}{2}$,

2. $\epsilon_{i,b} \geq \frac{\delta}{2}$,

3. *$a + b$ is a quadratic residue modulo q_i.*

Proof : Since $\alpha_{i+1} - \alpha_i < \sigma$, lemma 4 says that the number of *bad* residue classes modulo q_i is less than $\frac{\delta q_i}{2}$. So the number of *good* residue classes is at least $(1 - \frac{\delta}{2})q_i$. Also note that

$$\frac{1}{q_i} \sum_{k \text{ is } good} \epsilon_{i,k} \geq (\frac{11}{32} + \delta) - \frac{\delta}{2} = \frac{11}{32} + \frac{\delta}{2}.$$

$$G = \frac{|\{k | k \text{ is } good \text{ and } \epsilon_{i,k} \geq \frac{\delta}{2}\}|}{q_i}.$$

If $G > 1/2$, then we have the required pair a and b simply by pigeonhole principle. We will show now that $G > \frac{11}{32}$. Therefore, such a pair must exist in the light of theorem 1.

If $G \leq \frac{11}{32}$, then

$$\frac{1}{q_i} \sum_{k \text{ is } good} \epsilon_{i,k} \leq G \cdot 1 + (1 - G)\frac{\delta}{2} \leq \frac{11}{32} + \frac{21\delta}{64}.$$

contradicting lower bound we got above. □

Lemma 7 *Let z be a quadratic residue modulo q, n be a sufficiently large positive integer. Then the number of perfect squares in any interval of length H of $[n] \equiv z \bmod q$ is at least $\frac{H}{4\sqrt{n}}$.*

Proof :

Let $z \equiv k^2 \bmod q$. Suppose the interval is $[r, r+q, r+2q, \ldots, r+(H-1)q]$, where $0 \leq r \leq n - (H-1)q$.

So we want to count the number of integral l such that $\frac{\sqrt{r}-k}{q} \leq l \leq \frac{\sqrt{r+(H-1)q}-k}{q}$. It follows that the number of such l that satisfy these inequalities is at least

$$\frac{(\sqrt{r+(H-1)q} - k) - (\sqrt{r} - k)}{q} - 2 \geq \frac{(H-1)q}{(\sqrt{r+(H-1)q} + \sqrt{r})q} - 2 \geq \frac{H}{4\sqrt{n}}.$$

□

Lemma 8 (Shifting Lemma) *Let $S \subset [N]$ be such that the number of solutions to the equation $x + y = z^2$ where $x, y \in S$ is equal to 0. Let P be a big prime, C an integer constant and let*

$$M = C \cdot \prod_{p \ prime, p \leq P, p^{\alpha_p} \leq P^2} p^{\alpha_p}.$$

Then the number of solutions where $x \in S$ and $y \in S + jM$ (for $0 \leq j \leq N^{1-2\epsilon}$, $\epsilon > 0$) is at most $O(\frac{N\sqrt{N}}{\sqrt{P}})$.

Proof :

Given a Set $S \subset [N]$, the number of solutions to the equation $x_1 + x_2 = z^2$ where $x_1, x_2 \in S$ is given by equation (1)

$$\frac{1}{2N}\sum_{t=0}^{2N-1} f_{SQ}(\frac{-t}{2N})f_S(\frac{t}{2N})f_S(\frac{t}{2N})$$

Assuming that there is no solution to this equation then the sum in equation (1) must be 0.

We would like to estimate $|f_{SQ}(\frac{t}{2N})|$, by approximating $\frac{t}{2N}$ by $\frac{a(t)}{b(t)}$. Using Farey fraction with denominator $Q \approx N^{1-\epsilon}$ and letting r to be the error in the approximation we get.

$$r = |\frac{t}{2N} - \frac{a(t)}{b(t)}| \leq |\frac{1}{b(t) \times N^{1-\epsilon}}|$$

$$f_{SQ}(\frac{t}{2N}) = \sum_{x=0}^{\sqrt{2N}} e(x^2\frac{t}{2N}) = \sum_{x=0}^{\sqrt{2N}} e(x^2(\frac{a(t)}{b(t)} + r))$$

Now, replacing x by $j + b(t)k$, where j runs from 0 to $b(t) - 1$ and k runs from 0 to $\frac{\sqrt{2N}}{b(t)}$. We get $x^2 = j^2 + 2kjb(t) + k^2b(t)^2$ and,

$$f_{SQ}(\frac{t}{2N}) = \sum_{k=0}^{\frac{\sqrt{2N}}{b(t)}}\sum_{j=0}^{b(t)-1} e(j^2 + 2kjb(t) + k^2b(t)^2)(\frac{a(t)}{b(t)} + r)$$

$$= \sum_{k=0}^{\frac{\sqrt{2N}}{b(t)}}\sum_{j=0}^{b(t)-1} e(k^2b(t)a(t) + k^2b(t)^2r)e(2kja(t))e(j^2\frac{a(t)}{b(t)})e((j^2 + 2kjb(t))r)$$

$$= \sum_{k=0}^{\frac{\sqrt{2N}}{b(t)}} e(k^2b(t)a(t))e(k^2b(t)^2r)\sum_{j=0}^{b(t)-1} e(2kja(t))e(j^2\frac{a(t)}{b(t)})e((j^2 + 2kjb(t))r)$$

Now using the fact that $e(l) = 1$ when l is integer, we get

$$f_{SQ}(\frac{t}{2N}) = \sum_{k=0}^{\frac{\sqrt{2N}}{b(t)}} e(k^2b(t)^2r)\sum_{j=0}^{b(t)-1} e(j^2\frac{a(t)}{b(t)})e((j^2 + 2kjb(t))r) \tag{2}$$

Assuming that $b(t) \leq \sqrt{2N}$, we get

$$(j^2 + 2kjb(t))r \leq 3N^{(-0.5+\epsilon)} \text{ and } |e(j^2 + 2kjb(t))r)| \leq 1 + 3N^{(-0.5+\epsilon)}.$$

Therefore,

$$\begin{aligned} |\sum_{j=0}^{b(t)-1} e(j^2\frac{a(t)}{b(t)})e((j^2+2kjb(t))r)| &\leq |\sum_{j=0}^{b(t)-1} e(j^2\frac{a(t)}{b(t)})| + \sum_{j=0}^{b(t)-1} 3N^{(-0.5+\epsilon)} \\ &\leq \sqrt{b(t)} + 3b(t)N^{(-0.5+\epsilon)} \end{aligned} \tag{3}$$

In equation (3) we used the fact that the maximum value to $|e(j^2\frac{a(t)}{b(t)})|$ is 1, so we divided the sum into two parts, the first part is the basic part and the second is the error, to find the maximum value of the error we multiplied it by 1 instead of $e(j^2\frac{a(t)}{b(t)})$. Using Gauss sum, we get $|\sum_{j=0}^{b(t)-1} e(j^2\frac{a(t)}{b(t)})| = \sqrt{b(t)}$. Now using equation (2) and equation (3), we get equation (4), which is true for any $b(t) \leq \sqrt{2N}$ (notice that for $b(t) \geq \sqrt{2N}$ we get $|f_{SQ}(\frac{t}{2N})|$ is $o(\sqrt{2N})$ using lemma 2.4 on page $11-12$ of [V-97])

$$\begin{aligned} |f_{SQ}(\frac{t}{2N})| &\leq |\sum_{k=0}^{\frac{\sqrt{2N}}{b(t)}} e(k^2b(t)^2r)(\sqrt{b(t)} + 3b(t)N^{(-0.5+\epsilon)})| \\ &\leq (\sqrt{b(t)} + 3b(t)N^{(-0.5+\epsilon)}) \sum_{k=0}^{\frac{\sqrt{2N}}{b(t)}} 1 \\ &\leq \frac{2\sqrt{2N}}{\sqrt{b(t)}} \end{aligned} \tag{4}$$

We can divide the sum in equation (1) into two parts based on the value of $b(t)$ compared to P,

$$\sum\nolimits^1 = \frac{1}{2N} \sum_{b(t)<P} f_{SQ}(\frac{t}{2N})f_S(\frac{t}{2N})f_S(\frac{t}{2N})$$

$$\sum\nolimits^2 = \frac{1}{2N} \sum_{b(t)\geq P} f_{SQ}(\frac{t}{2N})f_S(\frac{t}{2N})f_S(\frac{t}{2N})$$

We want to upper bound $|\sum^2|$, so using equation (4) and *Parseval* identity , we get

$$\begin{aligned} |\sum\nolimits^2| &= \frac{1}{2N}|\sum_{b(t)\geq P} f_{SQ}(\frac{t}{2N})f_S(\frac{t}{2N})f_S(\frac{t}{2N})| \\ &\leq \frac{\sqrt{2N}}{N\sqrt{P}}|\sum_t |f_S(\frac{t}{2N})|^2| \\ &\leq \frac{\sqrt{2N}}{N\sqrt{P}}N^2 \\ &\leq \frac{N\sqrt{2N}}{\sqrt{P}} \end{aligned}$$

Now since $\sum^1 + \sum^2 = 0$, by the assumption that there is no solution, we conclude that

$$|\sum^1| \le \frac{N\sqrt{2N}}{\sqrt{P}} \tag{5}$$

We want to show the number of solutions to $x_1 + x_2 = z^2$ where $x_1 \in S$, and, $x_2 \in S + jM$, such that $j \le N^{1-2\epsilon}$ is close to the number of solutions to our problem. It is clear, to do that, we need only consider how much $\sum^1$ and $\sum^2$ will change if we replaced equation (1) by equation (6).

$$\frac{1}{2N}\sum_{t=0}^{2N-1} f_{SQ}(\frac{-t}{2N})f_S(\frac{t}{2N})f_{S+jM}(\frac{t}{2N}). \tag{6}$$

From now on we will call $\sum^1$ and $\sum^2$ related to this equation by $\sum^1_j$ and $\sum^2_j$. For $\sum^2_j$, we will use the fact from equation (4) with *Parseval* identity and *Cauchy-Schwartz* inequality to get,

$$\begin{aligned} |\sum^2_j| &\le \frac{\sqrt{2}}{\sqrt{NP}}|\sum_{b(t)\ge P} f_S(\frac{t}{2N})f_{S+jM}(\frac{t}{2N})| \\ &\le \frac{\sqrt{2}}{\sqrt{NP}}\sqrt{\sum |f_S(\frac{t}{2N})|^2}\sqrt{\sum |f_{S+jM}(\frac{t}{2N})|^2} \\ &\le \frac{N\sqrt{2N}}{\sqrt{P}} \end{aligned} \tag{7}$$

For $\sum^1_j$, we will use the fact that $b(t)|M$ if $b(t) < P$ and $jMr < N^{-\epsilon}$ (for the values of M, j and Q chosen), to get,

$$\begin{aligned} f_{S+jM}(\frac{t}{2N}) &= \sum_{x\in S+jM} e(x\frac{t}{2N}) \\ &= \sum_{x\in S+jM} e(\frac{a(t)x}{b(t)} + rx) \\ &= \sum_{y\in S} e(\frac{a(t)(y+jM)}{b(t)} + r(y+jM)) \\ &= \sum_{y\in S} e(\frac{a(t)y}{b(t)} + ry)e(\frac{M}{b(t)}ja(t))e(jMr) \\ &= e(jMr)\sum_{y\in S} e(y\frac{t}{2N}) \\ &= e(jMr)f_S(\frac{t}{2N}) \end{aligned}$$

Therefore,

$$|1 - N^{-\epsilon}||f_S(\frac{t}{2N})| \le |f_{S+jM}(\frac{t}{2N})| \le |1 + N^{-\epsilon}||f_S(\frac{t}{2N})|.$$

Again, using *Parseval* identity we get

$$
\begin{aligned}
|\sum_j^1| &= \frac{1}{2N}|\sum_{b(t)<P} f_{SQ}(\frac{t}{2N}) f_S(\frac{t}{2N}) f_{S+jM}(\frac{t}{2N})| \\
&\le \frac{1}{2N}|\sum_{b(t)<P} f_{SQ}(\frac{t}{2N}) f_S(\frac{t}{2N}) f_S(\frac{t}{2N})| + \frac{1}{2N}|\sum_t f_{SQ}(\frac{t}{2N})|f_S(\frac{t}{2N})||f_S(\frac{t}{2N})|N^{-\epsilon}| \\
&\le |\sum{}^1| + \frac{N^{-\epsilon}}{2N}|\sum_t f_{SQ}(\frac{t}{2N})|f_S(\frac{t}{2N})|^2| \\
&\le |\sum{}^1| + N^{\frac{3}{2}-\epsilon} \\
&\le \frac{N\sqrt{2N}}{\sqrt{P}} \text{ .... using equation (5)} \qquad (8)
\end{aligned}
$$

Using equation (8) and equation (7), we get

$$
\begin{aligned}
|\sum_j^1 + \sum_j^2| &\le |\sum_j^1| + |\sum_j^2| \\
&\le \frac{2N\sqrt{2N}}{\sqrt{P}}
\end{aligned}
$$

□

6 Proof of Main Theorem

Lemmas 2 and 3 imply that

$$(\frac{11}{32} + \delta)^2 \le \alpha_1 \le \alpha_2 \le \ldots \le \alpha_l \le \frac{11}{32} + \delta.$$

It cannot be the case that $\forall i$, $\alpha_{i+1} - \alpha_i \ge \sigma$. Therefore, let's say that $\alpha_{i+1} - \alpha_i < \sigma$ for some $i \ge 1$. By lemma 6, we have a pair of good residue classes mod q_i, denote them by a and b.

Let

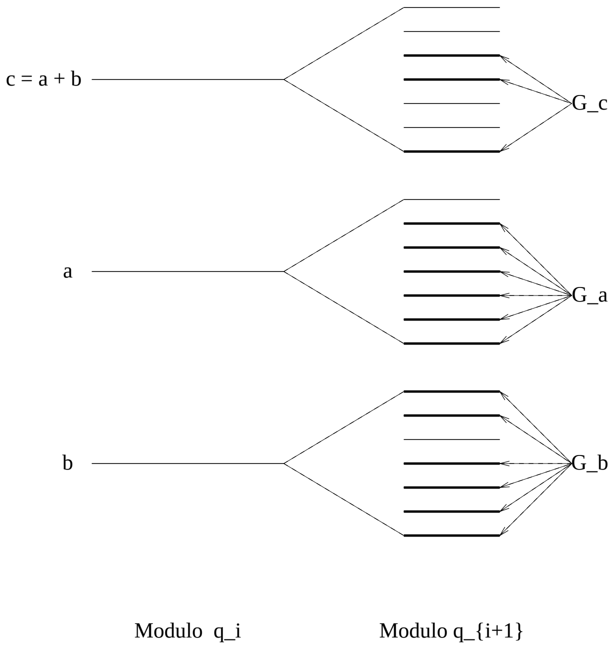
$$G_a = \{w_a | \epsilon_{i+1,a+w_a q_i} \ge \frac{\delta}{4}\}$$

and

$$G_b = \{w_b | \epsilon_{i+1,b+w_b q_i} \ge \frac{\delta}{4}\}$$

Consider residue class $c \equiv a+b$ modulo q_i. It is QR modulo q_i. When class c is subdivided into smaller residue classes modulo q_{i+1}, some of them will be QRs modulo q_{i+1}. Let's denote their index set as

$$G_c = \{w_c | 0 \le w_c < \frac{q_{i+1}}{q_i} \text{ and } c + w_c q_i \text{ is a QR modulo } q_{i+1}\}$$



Let's denote $|G_c| = L$. Note that when we shift through an interval of length $\frac{hq_{i+1}}{q_i}$ in residue class c modulo q_i, we are shifting through an interval of length h through each of the refined residue classes modulo q_{i+1}. Using lemma 7, the number of perfect squares in the interval of length $H = \frac{hq_{i+1}}{q_i}$ is at least $\frac{hq_{i+1}}{4q_i\sqrt{N}}$. Then it is easy to see that the number of perfect squares in the corresponding interval of length h in any of the residue classes in G_c is at least $\frac{hq_{i+1}}{8q_iL\sqrt{N}}$.

Let

$$W_{w_c} = \{(w_a, w_b) | w_a \in G_a \text{ and } w_b = w_c - w_a + \frac{c-a-b}{q_i} \text{ modulo } q_{i+1}\}$$

Note that when we add any residue pair w_a and w_b in W_{w_c}, we get QR w_c modulo q_{i+1}.

From lemma 6 and definition 3, $|G_a| = |W_{w_c}| = |G_b| \geq \frac{3q_{i+1}}{4q_i}$. So, at least 50% elements (w_a, w_b) in W_{w_c} are such that $w_a \in G_a$ and $w_b \in G_b$. That is, if we define

$$G_{w_c} = W_{w_c} \cap (G_a \times G_b)$$

then $|G_{w_c}| \geq \frac{q_{i+1}}{2q_i}$. Choose any pair $(w_a, w_b) \in G_{w_c}$. The number of solutions we get for equation $x + y = z^2$, given that we choose x from residue class $a + w_a q_i$ and y from residue class $b + w_b q_i$ modulo q_{i+1} and do shifting in the interval of length $h = N^{1-2\epsilon}$, $\epsilon > 0$, is at least

$$\frac{\delta N}{4q_{i+1}} \cdot \frac{\delta N}{4q_{i+1}} \cdot \frac{hq_{i+1}}{8q_iL\sqrt{N}}.$$

If we do this for all pairs in G_{w_c} and then repeat it for all different possible $w_c \in G_c$, we find that the number of solutions for $x + y = z^2$ is at least

$$\sum_{w_c \in G_c} \frac{q_{i+1}}{2q_i} \cdot \frac{\delta N}{4q_{i+1}} \cdot \frac{\delta N}{4q_{i+1}} \cdot \frac{hq_{i+1}}{8q_i L\sqrt{N}} = \sum_{w_c \in G_c} \frac{\delta^2 N\sqrt{N}h}{256{q_i}^2 L} = \frac{\delta^2 N\sqrt{N}h}{256{q_i}^2}.$$

This contradicts the upper bound that we get when we use **Shifting Lemma** with $P = p_{i+1}$ and $C = q_i$ (that is $M = q_{i+1}$) since $\frac{2N\sqrt{2N}h}{\sqrt{p_{i+1}}} \ll \frac{\delta^2 N\sqrt{N}h}{256{q_i}^2}$ (owing to lemma 1). $\square$

References

[EG-80] P. Erdős, R. L. Graham, "Old and New Problems and Results in Combinatorial Number Theory", *Monogr. Enseignment Mathe.*, 28, 1980.

[M] J. P. Massias, "Sur les suites dont les sommes des terms 2 á 2 ne sont par des carrés".

[LOS-82] J. P. Lagarias, A. M. Odylzko, J. B. Shearer, "On the Density of Sequences of Integers the Sum of No Two of which Is a Square. I. Arithmetic Progressions", *Journal of Combinatorial Theory. Series A*, 33, pp. 167–185, 1982.

[LOS-83] J. P. Lagarias, A. M. Odylzko, J. B. Shearer, "On the Density of Sequences of Integers the Sum of No Two of which Is a Square. II. General Sequences", *Journal of Combinatorial Theory. Series A*, 34, pp. 123–139, 1983.

[S-78] A. Sárkőzy, "On Difference Sets of Sequences of Integers", I, *Acta Math. Acad. Sci. Hungar.* , 31, pp. 125–149, 1978.

[V-97] R. C. Vaughan, "The Hardy-Littlewood Method", 2nd edition, Cambridge University Press, 1997.