Maximum size of a set of integers with no two adding up to a square

Simao Herdade, Rutgers University
Ayman Khalfalah, Rutgers University
Endre Szemerédi, Rényi Institute

January 5, 2016

Erdős and Sárközy asked in 1981:

What is the maximum cardinality of a subset of the first N integers with the property (NS) that no two elements add up to a perfect square?

Natural candidate

$$S = \{x \in \{1, \ldots, N\},\ x \equiv 1 \pmod 3\}?$$

Massias (1982) found a slightly bigger set with size $11N/32$. Namely,

$$S = \{x \in \{1, \ldots, N\}, x \equiv 1 \pmod 4 \text{ or } x \equiv 14, 26, 30 \pmod{32}\}$$

or

$$S' = \{x \in \{1, \ldots, N\}, x \equiv 1 \pmod 4 \text{ or } x \equiv 10, 14, 30 \pmod{32}\}.$$

In a remarkable work, J. Lagarias, M. Odlyzko and J. Shearer showed that over a finite cyclic group Z_N one cannot find more than $11N/32$ residue classes with the sum of any two a quadratic nonresidue.

In particular they showed that if our set S consists of residue classes for a fixed modulus then $|S| \leq 11N/32$.

In a subsequent paper they proved the first nontrivial upper bound of $0.475N$. ($0.5N$ is trivial)

Using the beautiful result of Lagarias et al., A. Khalfalah, S. Lodha and E. Szemerédi proved that for any $\delta > 0$ and sufficiently large N every subset of $1, \ldots, N$ having at least $(11/32 + \delta)N$ elements contains two elements that add up to a perfect square.

Based on the previous work by Lagarias, Odlyzko, Shearer, Khalfalah and Lodha, Herdade, Khalfalah and Szemerédi proved that

$$|S| \leq 11N/32$$

for large N.
They also proved "stability" results in both of the modular case ($\mathbb{Z}_N$) and for the case of integers. Maybe this is the more important contribution.

"Stability results" briefly

1., The only "large" sets S (that is, $|S| = 11N/32$) are the examples of Massias.

2., If M is one of the examples of Massias and

$$|S \Delta M| \geq \delta N,$$

then

$$|S| \leq \left(\frac{11}{32} - \epsilon(\delta) \right) N$$

is enough.

We could formulate the "stability results" in $\mathbb{Z}_N$, too.

We will sketch only the methods in our presentation.

1., We present the solution to the problem in finite cyclic groups (the results of Lagarias, Odlyzko and Shearer, no originality is claimed).

2., We characterize the subset of a finite cyclic group that achieves maximum size (the two examples of Massias).

3., We explain the argument which reduces the study of a subset of the first N integers to its distribution among residue classes modulo some fixed integer q. An almost tight bound ($\left(\frac{11}{32}+\delta\right) N$) is obtained using the result of Lagarias, Odlyzko and Shearer for cyclic groups (Khalfalah, Lodha, Sz.).

4., We show that the previous argument in 3. can be extended to a set of integers of size at least $\left(\frac{11}{32} - \delta\right) N$ and sufficiently different from the examples of Massias.

5., We prove that any subset of the first N integers of size at least $\frac{11}{32} N$ which is similar to one of the examples of Massias must have two elements that add to a perfect square.

Our main results can be found in 4. and 5.

We introduce a graph Q_N with $V(Q_N) = \mathbb{Z}_N$, and $xy \in \mathbb{Z}_N \times \mathbb{Z}_N$ is an edge if and only if $x + y \equiv t^2 \pmod{N}$ for some $t \in \mathbb{Z}_N$.

A set $S \subset \mathbb{Z}_N$ has property SR if there are no x, y, t such that

$$x + y \equiv t^2 \pmod{N}.$$

Obviously finding the maximum size of subset of $\mathbb{Z}_N$ with property SR is the same as finding the maximum size of independent sets ($\alpha(Q_N)$) in Q_N. Let

$$i(N) = \frac{\alpha(Q_N)}{N},$$

the *independence ratio*. For example, Lagarias et al. proved that $i(N) \leq 11/32$.

We define the product of two graphs G and H to be the graph $G \times H$ with vertex set $V(G) \times V(H)$ and $((x, y)(w, z))$ is an edge of $G \times H$ if and only if $(x, w) \in E(G)$ and $(y, z) \in E(H)$.

Easy lemma: Let

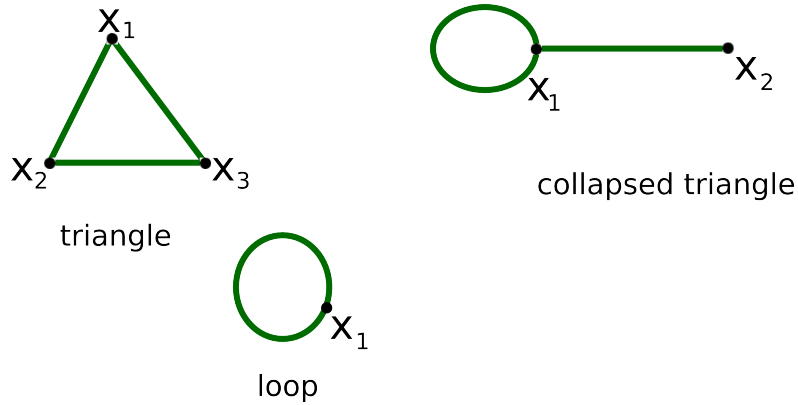
$$N = p_1^{\alpha_1} \cdots p_j^{\alpha_j}$$

then

$$Q_N = Q_{p_1^{\alpha_1}} \times \cdots \times Q_{p_j^{\alpha_j}}.$$

$(z \in V(Q_N),\ z_1, z_2, \ldots, z_j$
$z \equiv z_k \pmod{p_k^{\alpha_k}}$
$i + j \equiv z^2 \pmod{N}$ if and only if $z_k^2 \equiv i + j \pmod{p_k^{\alpha_k}}$
Chinese Remainder Theorem)

$Q_{2^n} \times Q_m$, $i(Q_{2^n} \times Q_m) \leq 11/32$



Difficult theorem of Lagarias, Odlyzko and Shearer: Let N be odd. Then Q_N has a D-uniform covering consisting entirely of triangles, collapsed triangles and loops, for some D depending on N.

Using very elementary facts from linear programming we get $i(Q_N) \leq i(Q_{2^j} \times T)$ where $N = 2^j m$, m is odd, and T is a triangle.

We have to show that

$$i(Q_N \times T) \leq 11/32.$$

Definitions: Let T be a triangle with vertex set $\{a, b, c\}$.
$V(Q_N \times T) = \bigcup_{i=0}^{7} V_i$ where
$V_i = \{(x, t) : \ x \equiv i \ (\mod 8), \ t = a, b, c\}$
I is the independent set of $Q_N \times T$

$$\alpha_i = \frac{|I \cap V_i|}{|V_i|}$$

Our goal is to show that $\sum_{i=0}^{7} \alpha_i \leq 2 + 3/7 \ \ (8 \cdot \frac{11}{32})$.

Observations

(1.)
$\alpha_0 = 0$ OR $\alpha_1 = 0$ OR ($\alpha_0 \leq 1/3$ AND $\alpha_1 \leq 1/3$)
$\alpha_2 = 0$ OR $\alpha_7 = 0$ OR ($\alpha_2 \leq 1/3$ AND $\alpha_7 \leq 1/3$)
$\alpha_3 = 0$ OR $\alpha_6 = 0$ OR ($\alpha_3 \leq 1/3$ AND $\alpha_6 \leq 1/3$)
$\alpha_4 = 0$ OR $\alpha_5 = 0$ OR ($\alpha_4 \leq 1/3$ AND $\alpha_5 \leq 1/3$)

(2.) $\alpha_0 + \alpha_4 \leq 11/16$

(3.)
$\alpha_1 + \alpha_7 \leq 1$
$\alpha_3 + \alpha_5 \leq 1$
$\alpha_1 + \alpha_3 \leq 1$
$\alpha_5 + \alpha_7 \leq 1$

Observations, cont'd

(4.) $2\alpha_6 \leq 1$

(5.) $2\alpha_2 + \alpha_6 \leq 1$

From these observations with some case analysis we can get the required result.

Characterization of extremal sets

$$M_{10} = \{x \in \mathbb{Z}_{2^n} |\ x \equiv 1 \pmod 4 \text{ or } x \equiv 14, 10, 30 \pmod{32}\},$$

$$M_{26} = \{x \in \mathbb{Z}_{2^n} |\ x \equiv 1 \pmod 4 \text{ or } x \equiv 14, 26, 30 \pmod{32}\}.$$

After case analysis $|S| = 2 + 3/7$ occurs only if
$\alpha_0 = 0,\ \alpha_1 = 1,\ \alpha_2 = 1/4,\ \alpha_3 = 0,\ \alpha_4 = 0, \alpha_5 = 1,$
$\alpha_6 = 1/2,\ \alpha_7 = 0.$

Before completing the characterization of the extremal sets we state a proposition:
If $|S| \geq 2 + 3/7 - \delta$ for some $\delta \in [0, 1/16)$ we have
$\alpha_0 = 0,\ \alpha_1 \geq 1 - \delta,\ 1/4 - \delta \leq \alpha_2 \leq 1/4 + \delta,\ \alpha_3 = 0,\ \alpha_4 = 0,$
$\alpha_5 \geq 1 - \delta,\ 1/2 \geq \alpha_6 \geq 1/2 - \delta,\ \alpha_7 \leq 2\delta.$

For the characterization we have to partition $Q_{2^n} \times T$ into 32 classes: $U_0, U_1, \ldots, U_{31}$, where

$$U_i = \{(x,t) |\ x \equiv i(\mod 32),\ t = a, b, c\}.$$

Let

$$\alpha_i^* = \frac{|I \cap U_i|}{|U_i|}.$$

$V_2 = U_2 \cup U_{10} \cup U_{18} \cup U_{26}$
$V_6 = U_6 \cup U_{14} \cup U_{22} \cup U_{30}$

We know that $\alpha_2 = 1/4,\ \alpha_6 = 1/2$. This means

$$\alpha_2^* + \alpha_{10}^* + \alpha_{18}^* + \alpha_{26}^* = 1$$

and

$$\alpha_6^* + \alpha_{14}^* + \alpha_{22}^* + \alpha_{30}^* = 2.$$

Case analysis gives EITHER

$$\alpha^*_{10} = 1,\ \alpha^*_{26} = \alpha^*_2 = \alpha^*_{18} = 0$$

OR

$$\alpha^*_{10} = \alpha^*_2 = \alpha^*_{18} = 0,\ \alpha^*_{26} = 1$$

.

In addition

$$\alpha^*_{30} = \alpha^*_{14} = 1,\ \alpha^*_6 = \alpha^*_{22} = 0.$$

If $|S| \geq 2 + 3/7 - \delta$ for some $\delta \in [0, 1/16)$, then we can derive inequalities for the values of $\alpha^*_j,\ 0 \leq j \leq 31$.

If for some independent set $A \subset V(Q_{2^n} \times T)$ we have

$$|A \Delta M_{10} \times T| \geq \delta 2^3 \cdot 3,$$

$$|A \Delta M_{26} \times T| \geq \delta 2^n \cdot 3$$

then obviously for this A, $\alpha_j^*(A)$ for some j will differ from α_j^* by at least $\delta/16$ so we can conclude that $|A| \geq 11/32 - \delta/100$.

This exactly is a stability theorem for $Q_{2^n} \times T$.

From $Q_{2^n} \times T$ back to $Q_{2^n m}$

Theorem: Let S be a subset of $Q_{2^n} \times Q_m$ with no two elements adding to a perfect square residue and $|S| = 11 \cdot 2^n m/32$. Then either $S = M_{10} \times Q_m$ or $S = M_{26} \times Q_m$.

Proof:

$$A = \{x \in Q_m |\ M_{10} \times \{x\} \subset S\}$$

$$\overline{A} = \{y \in Q_m |\ M_{26} \times \{y\} \subset S\}$$

Observe that no two elements $a \in A,\ b \in \overline{A}$ can be adjacent in Q_m (otherwise $((i, a), (j, b))$ would be an edge in $Q_{2^n} \times Q_m$ between two elements of S for every $i \equiv 10 \pmod{32},\ j \equiv 26 \pmod{32}$. But we will show that it is not possible to partition the vertices of Q_m into two sets with no edges between them.

$a \in A,\ b \in \overline{A},\ t \in \mathbb{Z}_m$, of course there is an edge between A and $\overline{A}$ provided that $\nu > 0$.

Sketch of the proof:
$e(x) = e^{2i\pi x}$
Let $f_A(\alpha) = \sum_{a \in A} e(a\alpha)$, and $f_{\overline{A}}(\alpha) = \sum_{a \in \overline{A}} e(a\alpha)$ and $f_{SQ}(\alpha) = \sum_{z=0}^{m-1} e(z^2\alpha)$. Then

$$\nu = \frac{1}{m} \sum_{t=0}^{m} f_A(t/m) f_{\overline{A}}(t/m) f_{SQ}(-t/m)$$

We divide the above sum into the sum of

$$(1) = \frac{1}{m} f_A(0) f_{\overline{A}}(0) f_{SQ}(0)$$

and

$$(2) = \frac{1}{m} \sum_{t \neq 0} f_A(t/m) f_{\overline{A}}(t/m) f_{SQ}(-t/m).$$

(1) is $|A| \cdot |\overline{A}|$ using Gauss sum, Parseval's identity and some standard computation we can show that $(2) \leq \frac{1}{\sqrt{2}} |A| \cdot |\overline{A}|$.

Assume S is a subset of $\{1, \ldots, N\}$ with size $(11/32 - \delta)N$ such that

$$|S \Delta M_{10}| \geq 10\delta N$$

and

$$|S \Delta M_{26}| \geq 10\delta N.$$

Theorem: There are $x, y \in S$ such that $x + y = z^2$.

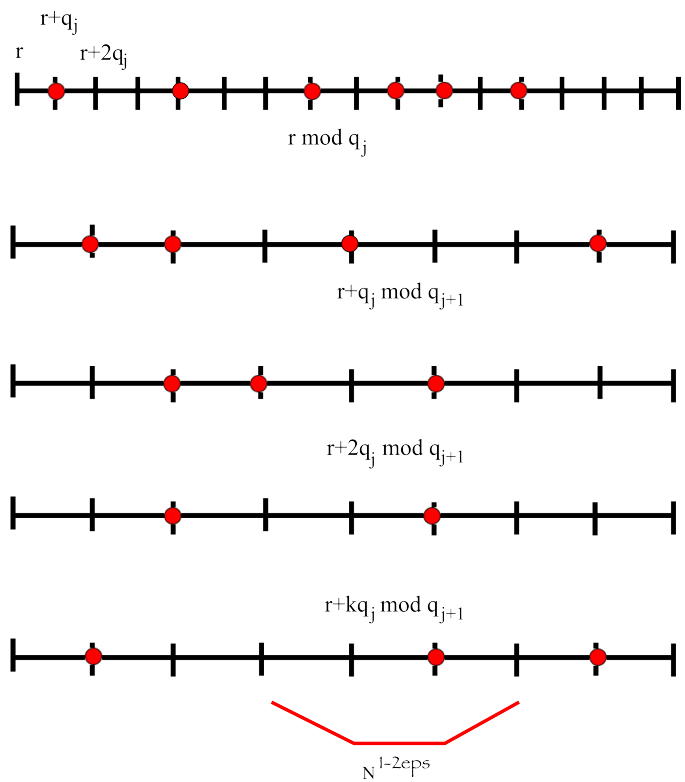
Proof: Let $q_i = m_i!$ and $q_{i+1} = m_{i+1}!$ where $m_{i+1} > 2^{m_i}$ for every i. Easy to see that there is a $j \leq \delta^{-5}$ such that the residue classes of $|S| \mod q_{j+1}$ are well distributed among the residue classes of $|S| \mod q_i$ $(i \leq \delta^{-5})$.

Definition:

$$\varepsilon_{i,j} = \frac{|\{s \in S |\ s \equiv j\ (\mod q_i)\}|}{N/q_i}$$

"Well distributed": $\varepsilon_{i+1,j+kq_i} \approx \varepsilon_{i,j},\ \ 0 \le k \le q_{i+1}/q_i$

Incrementing Method



We can assume that there are at least $(11/32 + \delta/100)q_i$ residue classes such that they contain at least $\delta N/(2q_i)$ elements of S.

Then from the result of Lagarias, Odlyzko and Shearer we can conclude that there are residue classes, say, x, y, z^2 such that $x + y \equiv \ (\mod q_i)$.

If $M = \Pi_{i=1}^{j} q_i$ and $M^* = q_{i+1}$ then by the "shifting" argument we can conclude that the number of solutions of

$$x + y = z^2$$

$$x \in S,\ y \in S + jM,\ 1 \le j \le N^{1-2\varepsilon}$$

is at least $10N\sqrt{N}/\sqrt{P}$, where P is the largest prime divisor of M.

On the other hand

$$\frac{1}{N}\sum_{t=1}^{N} f_S(t/N) f_{S+jM}(t/N) f_{SQ}(-t/N) \approx$$

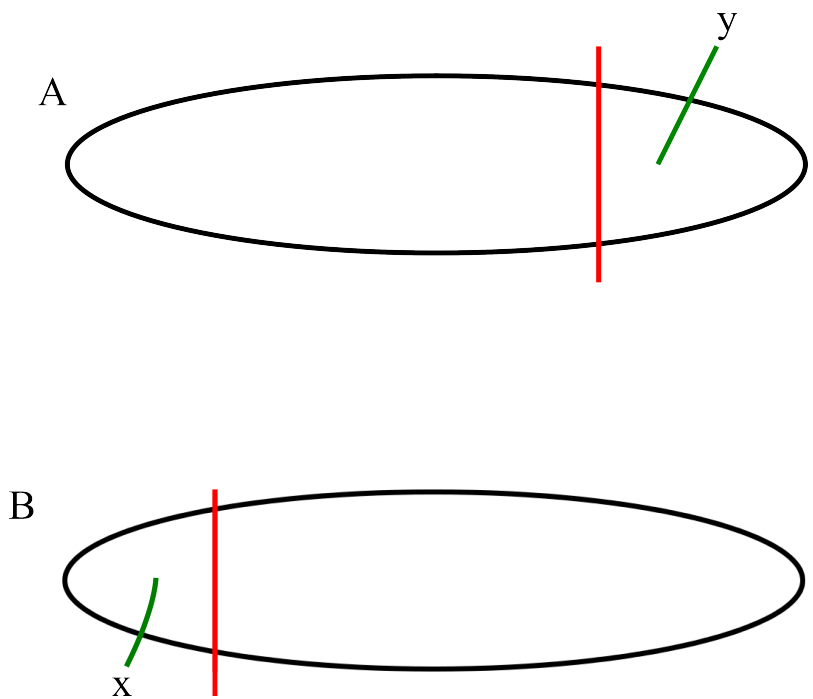
$$\frac{1}{N}\sum_{t=1}^{N} f_S(t/N) f_S(t/N) f_{SQ}(-t/N)$$

(recall, that $f_{SQ}(\alpha) = \sum_{1 \le z^2 \le N} e(z^2\alpha)$).

So from the analytical expression we get that the number of solutions is $\le N^{3/2}/(100\sqrt{P})$, a contradiction.

Let $S \subset \{1, \ldots, N\}$ be a set such that $|S| = 11N/32$ and

$$|S \Delta M_{10}| < \varepsilon N.$$



$x \in S,\ y \notin S,\ x+w^2 = y,\ y \in Y,\ y+w^2 \notin A \backslash Y,\ x+w^2 = y+jM^*$

Assume further that

$a \equiv j \pmod{32},\ j \in M_{10}^*,$

$b \equiv K \pmod{32},\ K \notin M_{10}^*,\ j + K \equiv t^2 \pmod{32}.$

Then the number of solutions for a fixed $y \in A$ is about

$$N^{1-2\varepsilon}/\sqrt{N},$$

the total number of solutions is

$$\frac{|x|N^{1-2\varepsilon}}{\sqrt{N}\sqrt{P}},$$

but it should be

$$\frac{|x| \cdot |y| \cdot N^{1-2\varepsilon}}{\sqrt{N}\sqrt{P}}.$$