Combinatorics and Diophantine equations

Katalin Gyarmati

Eötvös Loránd University
Department of Algebra and Number Theory
Budapest

katalin.gyarmati@gmail.com

We say that a set $\mathcal{B} \subseteq \mathbb{Z}$ forms a multiplicative basis of order h of $\mathcal{S}$ if every element of $\mathcal{S}$ can be written as the product of h members of $\mathcal{B}$.

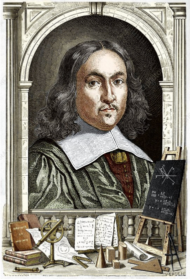
My original plan for this talk was to present non-trivial lower bounds for the size of a multiplicative basis of order 2 in the set $\{f(1), f(2), \ldots, f(n)\}$, where $f(x) \in \mathbb{Z}[x]$ is a polynomial.

However, it turned out that I recently proved some new results in another topic...

Part I

On Diophantine square tuples

Diophantus: the rational numbers $\frac{1}{16}$, $\frac{33}{16}$, $\frac{17}{4}$, and $\frac{105}{16}$ have the following property: the product of any two of them increased by 1 is the square of a rational number.



Fermat: $\{1, 3, 8, 120\}$

Euler:

$\{a, b, a+b+2r, 4r(r+a)(r+b)\}$

where $ab+1=r^2$.

These examples motivate the following definition:

Definition 1

A set $A = \{a_1, a_2, \ldots, a_m\} \subset \mathbb{Z}^+$ is called a Diophantine m-tuple if $a_i a_j + 1$ is a perfect square for all $1 \leq i, j \leq m$.

Dujella: No Diophantine 6-tuple.

He, Togbé, and Ziegler: No Diophantine 5-tuple.

Similar questions

Theorem A [Rivat-Sárközy-Stewart]. *There exists an integer* x_0 *such that if* $x_0 < x \in \mathbb{N}$, $\mathcal{A} \subset \{1, 2, 3, \dots, x\}$ *and for all* $a, a' \in \mathcal{A}$, *the sum* $a + a'$ *is a square, then*

$$|\mathcal{A}| < 37 \log x.$$

Lagrange and Nicolas: found a 6-element set $\mathcal{A}$ satisfying this property:

$$\mathcal{A} = \{ -15863902, 17798783, 21126338, 49064546, 82221218, \\ 447422978\}.$$

Question: How can we estimate the size of sets where the difference between any two elements is always a square?

We can assume the smallest element is 0.

Then all other elements are squares.

Definition 2 (Gy.)

A set $A = \{a_1^2, a_2^2, \ldots, a_m^2\} \subset \mathbb{Z}^+$ is called a Diophantine square m-tuple if $\left|a_i^2 - a_j^2\right|$ is a non-zero square for all $1 \leq i < j \leq m$.

Using computers to consider the interval $[1, 3000^2]$, I found the following triples of this type:

$$\begin{array}{lll}
(153^2, 185^2, 697^2) & (264^2, 520^2, 1105^2) & (264^2, 561^2, 1105^2) \\
(306^2, 370^2, 1394^2) & (448^2, 952^2, 1073^2) & (459^2, 555^2, 2091^2) \\
(495^2, 975^2, 1073^2) & (520^2, 533^2, 925^2) & (528^2, 1040^2, 2210^2) \\
(528^2, 1122^2, 2210^2) & (612^2, 740^2, 2788^2) & (644^2, 725^2, 2165^2) \\
(672^2, 680^2, 697^2) & (756^2, 765^2, 925^2) & (896^2, 1904^2, 2146^2) \\
(952^2, 1073^2, 1105^2) & (975^2, 1073^2, 1105^2) & (990^2, 1950^2, 2146^2) \\
(1040^2, 1066^2, 1850^2) & (1092^2, 2175^2, 2665^2) & (1344^2, 1360^2, 1394^2) \\
(1512^2, 1530^2, 1850^2) & (1540^2, 2431^2, 2665^2) & (1560^2, 1599^2, 2775^2) \\
(1904^2, 2146^2, 2210^2) & (1950^2, 2146^2, 2210^2) & (2016^2, 2040^2, 2091^2) \\
(2040^2, 2067^2, 2165^2) & (2268^2, 2295^2, 2775^2) & (2688^2, 2720^2, 2788^2)
\end{array}$$

Among these 30 triples, there are 14 whose elements are coprimes.

In the following, we give a table with n in the first column, the number of Diophantine square triples in the interval $[1, n^2]$ in the second column, the number of such Diophantine square triples whose elements are coprime in the third column, and the proportion of the number of these coprime Diophantine square triples and $n^{1/2}$ in the fourth column (now for the triples $(a^2, b^2, c^2) \subseteq [1, n^2]$ we suppose $a < b < c \leq n$).

n	# D. s. triples	# coprime D. s. triples	proportion
200000	4626	232	0.5188
400000	9438	334	0.5281
600000	14306	422	0.5448
800000	19170	468	0.5232
1000000	24030	510	0.51

Theorem 1

There are infinitely many Diophantine square triples (a^2, b^2, c^2) *such that* $\gcd(a, b, c) = 1$.

Proof of Theorem 1.

One example: $a_1 = 153, b_1 = 185, c_1 = 697$ (here all integers are odd).

Next, we construct infinitely many Diophantine square triples (a_i, b_i, c_i) by a simple recursion.

Assume that for some $i \in \mathbb{N}$ the Diophantine square triple a_i, b_i, c_i is already given. Then if $a_i + b_i + c_i$ is odd, then let

$$\begin{aligned} a_{i+1} &= \left|a_i^2 + b_i^2 - c_i^2\right| \\ b_{i+1} &= \left|a_i^2 - b_i^2 + c_i^2\right| \\ c_{i+1} &= \left|-a_i^2 + b_i^2 + c_i^2\right|. \end{aligned} \tag{1}$$

Then

$$\begin{aligned} b_{i+1}^2 - a_{i+1}^2 &= (a_i^2 - b_i^2 + c_i^2)^2 - (a_i^2 + b_i^2 - c_i^2)^2 \\ &= 4a_i^2(c_i^2 - b_i^2). \end{aligned}$$

Since (a_i, b_i, c_i) is a Diophantine square triple, $c_i^2 - b_i^2$ is a square, thus $b_{i+1}^2 - a_{i+1}^2$ is a square. Similarly, $c_{i+1}^2 - a_{i+1}^2$, $c_{i+1}^2 - b_{i+1}^2$ are also squares.

It is not very difficult to show that $\gcd(a_{i+1}, b_{i+1}, c_{i+1}) = 1$ and $c_i \to \infty$ as $i \to \infty$.

Thus, $(a_{i+1}, b_{i+1}, c_{i+1})$ is a Diophantine square tuple, whose elements are coprime, and their maximum element tends to infinity as $i \to \infty$. This completes the proof of the theorem.

We remark that a similar recursion can be given if $a_i + b_i + c_i$ is even, namely, let now

$$\begin{aligned} a_{i+1} &= \frac{1}{2}\left|a_i^2 + b_i^2 - c_i^2\right| \\ b_{i+1} &= \frac{1}{2}\left|a_i^2 - b_i^2 + c_i^2\right| \\ c_{i+1} &= \frac{1}{2}\left|-a_i^2 + b_i^2 + c_i^2\right|. \end{aligned} \tag{2}$$

Clearly, $a_i, b_i, c_i \in \mathbb{N}$.

Problem 1

Does exist a parametric system of equations that describes all Diophantine square triples?

Related to the proof of Theorem 1, the following easier question also arises:

Problem 2

Is there a finite set of coprime Diophantine square triples from which all coprime Diophantine square triples can be obtained using only recursions (1) *and* (2)*? Or will this statement only hold if it includes further recursions?*

First, I conjectured that there is no Diophantine square triple (a^2, b^2, c^2) whose elements are pairwise coprime.

This conjecture later turned out to be false when I ran a Python program and found only one such Diophantine square triple in the $[1, 10^{12}]$ interval, namely the triple

$$(40920^2,\ 41449^2,\ 42601^2).$$

Related to this, we ask the following:

Problem 3

Do there exist infinitely many Diophantine square triples whose elements are pairwise coprime?

Another important question is that what upper bound can be given for the size of Diophantine square tuples.

By computer search, I found that there is no Diophantine square quadruple in the interval $[1, 10^{12}]$. Thus, I conjecture the following:

Conjecture 1

There exists a positive integer n such that there is no Diophantine square n-tuple.

Perhaps this conjecure holds in an even sharper version:

Conjecture 2

There is no Diophantine square quadruple.

I have not been able to prove these conjectures.

Thus, instead I will prove a result which can be considered as a partial result in this direction: I will estimate the size of the Diophantine square tuples in terms of the largest element of the set.

Theorem 2

For every $\varepsilon > 0$, there exists an integer $x_0 = x_0(\varepsilon)$ such that if $x_0 < x$, $x \in \mathbb{N}$, $\mathcal{A} \subset \{1, 2, 3, \dots, x\}$, and for all $a, a' \in \mathcal{A},\ a > a'$, the difference $a - a'$ is a square, then

$$|\mathcal{A}| < (1 + \varepsilon) \log x.$$

The proof of this theorem is very similar to the proof of Theorem A.

The main tools are the following:

Lemma 1 (Gallagher's larger sieve)

Suppose that $m, n \in \mathbb{N}$, $\mathcal{A} \subset \{m+1, m+2, \ldots, m+n\}$ *and* $\mathcal{B} \subset \mathbb{N}$ *is a finite set such that its elements are pairwise coprime. For all* $b \in \mathcal{B}$, *denote the number of residue classes* $\mod b$ *that intersect* $\mathcal{A}$ *by* $\nu(b)$. *Then*

$$|\mathcal{A}| \leq \frac{\sum\limits_{b \in \mathcal{B}} \log b - \log n}{\sum\limits_{b \in \mathcal{B}} \frac{\log b}{\nu(b)} - \log n}, \tag{3}$$

provided that the denominator is positive.

Lemma 2 (Hanson-Petridis)

Let p *be a prime. If* $\mathcal{C} \subset \mathbb{Z}_p$ *is a set such that for all* $a, a' \in \mathcal{C}$, $a \neq a'$ *the difference* $a - a'$ *is quadratic residue modulo* p, *then*

$$|\mathcal{C}| \leq \sqrt{p/2} + 1.$$

Probably, Conjecture 1 is a very difficult problem, but perhaps one can make it easier by allowing only certain special subsets.

Proposition 1

There is no Diophantine square triple containing only squares of primes.

The next question is whether we can replace the squares of primes with other sets, for which the answer is less clear. The Fibonacci sequence is a good example of this.

Theorem 3 (Gy.)

There is no Diophantine square triple consisting of squares of Fibonacci numbers.

Related to Theorem 3, we mention that Fujita and Luca proved that there are no Diophantine quadruples of Fibonacci numbers in the sense of the original definition ($aa' + 1$'s are always squares).

Proof of Theorem 1.

The main lemma is the following:

Proof of Theorem 3. The following lemma is true for Fibonacci numbers:

Lemma 3 (Gy)

If $0 < m < n$, $m \equiv n \pmod 2$ *and* $F_n^2 - F_m^2$ *is a square, then*

$$(F_m, F_n) = (F_5, F_7) = (5, 13).$$

The case $\gcd(m, n) = 1$ was proved by Bicknell-Johnson in 1979.

Let us see the proof of the general case:

Proof of Lemma 3. The following identity is due to Ruggles:

Lemma 4

If $0 < m \leq n$ *and* $0 < m \equiv n \pmod 2$, *then*

$$F_n^2 - F_m^2 = F_{n+m} F_{n-m} \tag{4}$$

It is known that

$$\gcd(F_a, F_b) = F_{\gcd(a,b)}.$$

By this, if $d \overset{\text{def}}{=} \gcd(n-m, n+m)$, then

$$\gcd(F_{n+m}, F_{n-m}) = F_d,$$

and by (4), we get

$$\frac{F_n^2 - F_m^2}{F_d^2} = \frac{F_{n+m}}{F_d} \cdot \frac{F_{n-m}}{F_d}.$$

Here $\frac{F_{n+m}}{F_d}$ and $\frac{F_{n-m}}{F_d}$ are coprime, and their product is a square, so both of them are squares.

McDaniel and Ribenboim proved the following:

Lemma 5 (McDaniel, Ribenboim)

Assume u, v *and* y *are positive integers such that* $\dfrac{F_v}{F_u} = y^2$. *Then, either* $u = v$ *or* $(v, u) \in \{(12, 1), (12, 2), (2, 1), (6, 3)\}$.

Now $\dfrac{F_{n+m}}{F_d}$ is a square and $d = \gcd(n + m, n - m) < n + m$, thus

$$(n + m, d) \in \{(12, 1), (12, 2), (2, 1), (6, 3)\}.$$

Then

$$n, m < n + m \leq 12.$$

Using a Python program, it is easy to check that among the first 12 Fibonacci numbers, there are two pairs for which $F_n^2 - F_m^2$ is a square, namely

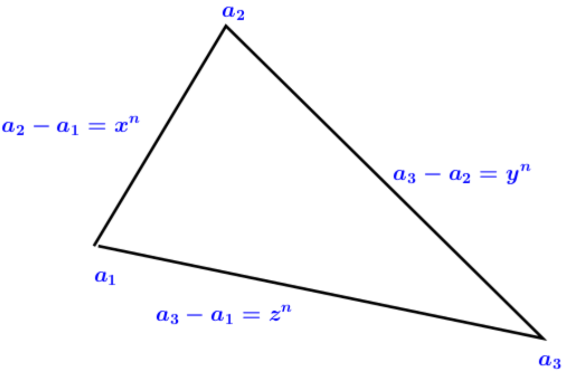
$$(F_2, F_5) = (3, 5) \quad \text{and} \quad (F_5, F_7) = (5, 13).$$

This completes the proof of Lemma 3.

Theorem 4 (Bugeaud, Gy.)

If $n \geq 3$ and $A \subseteq \mathbb{Z}$, then at most $\frac{|\mathcal{A}|^2}{4}$ pairs (a, a') exist such that $a > a'$ and $a - a'$ is an nth power.

$a_1 < a_2 < a_3$



a_2

$a_2 - a_1 = x^n$

$a_3 - a_2 = y^n$

a_1

$a_3 - a_1 = z^n$

a_3

$$z^n = a_3 - a_1 = (a_3 - a_2) + (a_2 - a_1) = y^n + x^n$$

Conjecture 3

There exists a constant $\varepsilon > 0$ such that for all $A \subseteq \mathbb{Z}$, there exist at most $(1-\varepsilon)\frac{|\mathcal{A}|^2}{2}$ pairs (a, a') for which $a - a'$ is a square.

If there is no Diophantine square quadruple, this conjecture is a simple consequence of Turán's theorem.

Using graph theory I proved that if we consider only the squares of Fibonacci numbers, this conjecture is true and can be improved.

Theorem 5 (Gy.)

If $A \subseteq \mathbb{Z}$, then at most $|\mathcal{A}|^{3/2} + |\mathcal{A}|$ pairs (a, a') exist such that $a - a'$ is a square of a (positive) Fibonacci number.

Part II

On multiplicative basis of finite sets

Joint work with Katalin Fried

Introduction

For a set $\mathcal{S} \subseteq \mathbb{Z}$ we denote by $\mathcal{S}(n)$ the cardinality of the set $\mathcal{S} \cap [1, 2, \ldots, n]$.

We say that a set $\mathcal{B} \subseteq \mathbb{Z}$ forms a multiplicative basis of order h of $\mathcal{S}$ if every element of $\mathcal{S}$ can be written as the product of h members of $\mathcal{B}$.

While the study of additive bases is an intensively studied topic in additive number theory, much less attention is devoted to multiplicative bases.

First multiplicative basis of $[n] \overset{\text{def}}{=} [1, 2, \ldots, n]$ were studied. It is easy to see that every multiplicative basis of $[n]$ contains the prime numbers up to n.

On the other hand in 2011 Chan prove that there is a multiplicative basis with less than $\pi(n) + c(h+1)^2 \, \frac{n^{2/(h+1)}}{\log^2 n}$ elements.

This upper bound has been recently sharpened by a factor h by Pach and Sándor. Namely if $G_h(n)$ denotes the size of the smallest multiplicative basis of order h of $[n]$ then

$$\pi(n) + 0.5h\frac{n^{2/(h+1)}}{\log^2 n} \leq G_h(n) \leq \pi(n) + 150.4h\frac{n^{2/(h+1)}}{\log^2 n}.$$

Slightly related problems were studied by Erdős. Next a few definitions follow.

Definition 1

In general for a set $\mathcal{S}$ we denote by $G_h(\mathcal{S})$ the size of the smallest multiplicative basis of order h. A basis $\mathcal{B}$ of order h is a minimal basis of order h of $\mathcal{S}$ if $|\mathcal{B}| = |G_h(\mathcal{S})|$. We call $\mathcal{B}$ a giant basis of order h of $\mathcal{S}$ if $|\mathcal{B}| \geq |\{1\} \cup \mathcal{S}|$.

In this talk we will study multiplicative basis of order 2 of the set $S(f(x), n) \overset{\text{def}}{=} [f(1), f(2), \dots, f(n)]$ where $f(x) \in \mathbb{Z}[x]$ is a polynomial. (A related problem was studied by Hajdu and Sárközy, namely they studied multiplicative decomposability of polynomial sets.)

Clearly, if $f(x)$ is of the form $f(x) = x^r$ then from Chan's result, Pach and Sándor's following result immediately follows

Proposition 1

$$\pi(n) \leq G_h(S(x^r, n)) \leq \pi(n) + 150.4h\frac{n^{2/(h+1)}}{\log^2 n}.$$

So, for these polynomials $f(x) = x^r$ we know the exact order of magnitude of $G_h(S(f(x), n))$.

Now we will study the case of other polynomials. First we study the simplest case $f(x) = x^2 + 1$.

One may conjecture that the set $S(x^2 + 1, n)$ has only giant bases, but it turned out that this is not the case. There exists a basis with slightly less elements than $|\{1\} \cup S(f(x), n)|$.

On the other hand we will prove that every multiplicative basis of $S(x^2 + 1, n)$ has at least as many elements as the number of prime numbers of the form $4k + 1$ between n and $2n$. In other words:

Theorem 1 (Fried, Gy.)

For every $\varepsilon > 0$ there exists a constant $n_0 = n_0(\varepsilon)$ such that for $n > n_0$ we have

$$\left(\frac{1}{2} - \varepsilon\right) \frac{n}{\log n} \leq G_h(S(x^2 + 1, n)) \leq n - n^{1/2} + (1 + \varepsilon) n^{1/4}.$$

There is a huge gap between the lower and upper bound. It is an interesting question which one is closer to the truth.

Problem 1

Does there exist a constant $\varepsilon_1 > 0$ such that

$$\varepsilon_1 n \le G_2(S(x^2+1,n)) \le (1-\varepsilon_1)n$$

is always true?

We will also study the case of general polynomials $f(x)$. In this case we will be able to prove the following:

Theorem 2 (Fried, Gy.)

Let $f(x) \in \mathbb{Z}[x]$ be a polynomial of degree $r \geq 2$ and write $f(x)$ as a product of irreducible polynomials over $\mathbb{Z}[x]$, say

$$f(x) = f_1(x) f_2(x) \cdots f_s(x), \tag{1}$$

where s denotes the number of irreducible factors in (1). Then

$$\frac{n}{(\log n)^{s \log r / \log 2}} \ll G_2(S(f(x), n)).$$

We remark that from Theorem 2 immediately follows the following:

Corollary 1

Let $f(x) \in \mathbb{Z}[x]$ be a polynomial of degree $r \geq 2$. Then

$$\frac{n}{(\log n)^{r \log r / \log 2}} \ll G_2(S(f(x), n)).$$

In case of the polynomial $f(x) = x^2 + 1$, the lower bound in Theorem 2 gives the same result as the one in Theorem 1.

As a general upper bound we are able to give the trivial bound $|\{1\} \cup S(f(x), n)| \leq n + 1$. Related to the upper bound we ask the following questions.

Problem 2

Is there any polynomial $f(x)$ such that for every n the set $S(f(x), n)$ has only giant bases of order 2, in other words do we have for every basis $\mathcal{B}$ of order 2 the following

$$|\mathcal{B}| \geq |\{1\} \cup S(f(x), n)|?$$

Or, is there a general non-trivial upper bound for $G_2(S(f(x), n))$?

Perhaps the lower bound in Theorem 2 can be sharpened. We also ask the following:

Problem 3

Is it possible to give a general better lower bound for $G_2(S(f(x), n))$ *than the bound* $\frac{n}{(\log n)^{s \log r/log 2}}$ *in Theorem 2?*

So far we have been studying multiplicative bases of $S(f(x), n) = \{f(1), f(2), f(3), \ldots, f(n)\}$.

Next we study the multiplicative bases of its subsets, i.e. sets of the form

$$\mathcal{W} \overset{\text{def}}{=} \{f(a_1), f(a_2), f(a_3), \ldots, f(a_k)\}, \qquad (2)$$

where $1 \le a_1 < a_2 < \cdots < a_k \le n$ are integers.

If $\mathcal{B}$ is a multiplicative basis of order 2 of $\mathcal{W}$, then each elements of $\mathcal{W}$ can be written in the form $b_i b_j$ with $b_i, b_j \in \mathcal{B}$, thus

$$|\mathcal{W}| \le |\mathcal{B}|^2 \,,$$

and so

$$|\mathcal{W}|^{1/2} \le |\mathcal{B}| \,. \qquad (3)$$

In case of polynomials $f(x)$ of degree 2, this problem is slightly related to the study of Diophantine tuples.

We will study whether (3) is the best possible general lower bound? Under some not too restrictive conditions on the a_i's in $\mathcal{W}$ we will prove $|\mathcal{W}|^{2/3} \ll |\mathcal{B}|$:

Theorem 3 (Fried, Gy.)

Let $f(x) \in \mathbb{Z}[x]$ be a polynomial of degree $\deg f \geq 2$ and $u, a_1, a_2, \ldots, a_k$ be positive integers such that

$$u \leq a_1 < a_2 < \cdots < a_k < 2u. \tag{4}$$

We define $\mathcal{W}$ by (2). If $\mathcal{B}$ is a multiplicative basis of order 2 of $\mathcal{W}$ then

$$|\mathcal{W}|^{2/3} \ll |\mathcal{B}|. \tag{5}$$

Remark 1

If $f(x)$ is of the form $f(x) = x^r + a_{r-3}x^{r-3} + \cdots + a_{r-4}x^{r-4} + \cdots + a_0$ (so the coefficients of the terms x^{r-1} and x^{r-2} are 0), then Theorem 3 also holds if in place of (4) only $u \leq a_1 < a_2 < \cdots < a_k < u^2$ holds.

Related to Theorem 3 we ask the following

Problem 4

Is it true that the lower bound (5) holds for arbitrary a_i's, i.e. is condition (4) indeed necessary in Theorem 3? In this general case which lower bound can be given for $|\mathcal{B}|$?

Remark 2

Let $\mathcal{B}$ be a multiplicative basis of order 2 of the set $\mathcal{W}$ defined in Theorem 3. Probably, the lower bound (5) in case of certain special polynomials might be sharpened to $|\mathcal{W}|^{3/4} \ll |\mathcal{B}|$. For more details see the end of the proof of Theorem 3.

Finally we will say a few words about sets having only giant bases.

Definition

We call $\mathcal{B}$ a giant basis of order h of $\mathcal{S}$ if $|\mathcal{B}| \geq |\{1\} \cup \mathcal{S}|$.

Clearly the set $I = [a^2, a^2+1, a^2+2, \ldots, a^2+a]$ has only giant bases: Let $\mathcal{B}$ be a multiplicative basis of I of order 2. We split $\mathcal{B}$ into two disjoint subsets, so $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$ where

$$\mathcal{B}_1 \overset{\text{def}}{=} \{b \in \mathcal{B} : \ b \leq a\}$$

$$\mathcal{B}_2 \overset{\text{def}}{=} \{b \in \mathcal{B} : \ b \geq a+1\}.$$

If $b_i b_j \in I$ and $b_i < b_j$, then $b_i \leq a$ and $b_j \geq a+1$. Thus for $b_i b_j \in I$ and $b_i < b_j$, we have $b_i \in \mathcal{B}_1$ and $b_j \in \mathcal{B}_2$.

For each $b \in \mathcal{B}_2$ there exists at most one element i of I for which $b \mid i$ since $|I| = a+1 \leq b$. Thus

$$a+1 = |I| \leq |\mathcal{B}_2| < |\mathcal{B}|,$$

from which the statement follows.

Our final problem is the following:

Problem 5

Let $I = [m+1, m+2, \ldots, m+n]$ and $d \geq 2$ is an integer. For which m and n's does I have only giant bases?

Proof of Theorem 1

First we prove that for $n > n_0(\varepsilon)$ we have

$$\left(\frac{1}{2} - \varepsilon\right) \frac{n}{\log n} \leq G_h(S(x^2+1, n)). \tag{6}$$

Let $\mathcal{B}$ be a multiplicative basis of order h of $S(x^2+1, n)$. Let $\mathcal{P}$ denote the following set

$$\mathcal{P} \overset{\text{def}}{=} \{p : \ p \text{ is a prime of form } 4k+1 \text{ and } n < p < 2n\}. \tag{7}$$

For every prime $p \in \mathcal{P}$ we assign the smallest positive integer $g = g(p)$ with

$$p \mid g(p)^2 + 1.$$

Since for $p \in \mathcal{P}$, p is a prime number of form $4k+1$, the congruence

$$x^2 \equiv -1 \pmod{p}$$

has two different solutions, and one of them is between 1 and $(p-1)/2$, thus

$$1 \le g(p) \le \frac{p-1}{2} < n. \qquad (8)$$

Since $\mathcal{B}$ is a multiplicative basis of $S(x^2+1, n)$ it is also a multiplicative basis of its subsets, namely $\mathcal{B}$ is a multiplicative basis of

$$S_1 \overset{\text{def}}{=} \{g(p)^2 + 1 : \ p \in \mathcal{P}\}$$

since $S_1 \subset S(x^2+1, n)$ by (8).

For every $p \in \mathcal{P}$, S_1 contains a multiple of p since $p \mid g(p)^2 + 1$. Thus $\mathcal{B}$ contains a multiple of p, which we denote by $h(p)$. Thus $h(p) \in \mathcal{B}$ and $p \mid h(p)$.

We will prove that for $p, q \in \mathcal{P}$, $p \neq q$

$$h(p) = h(q)$$

is not possible. Contrary, suppose that $p \neq q$ and $h(p) = h(q)$. Then

$$p \mid h(p), \;\; q \mid h(q).$$

Thus

$$pq \mid h(p) = h(q).$$

Since $p, q \in \mathcal{P}$ we have $n + 1 \leq p, q$ so

$$(n+1)^2 \leq pq \leq h(p) = h(q). \tag{9}$$

But $\mathcal{B}$ is a multiplicative basis of $S(x^2 + 1, n)$ so its elements are less or equal to $n^2 + 1$, thus

$$h(p) = h(q) \leq n^2 + 1,$$

which contradicts (9).

Thus the function $h: \mathcal{P} \to \mathcal{B}$ is injective, so

$$|\mathcal{P}| \leq |\mathcal{B}|,$$

which proves (6).

In order to prove

$$G_h(S(x^2+1,n)) \leq n - n^{1/2} + (1+\varepsilon)n^{1/4}.$$

we will prove a slightly stronger upper bound, namely $G_h(S(x^2+1,n)) \leq n - n^{1/2} + n^{1/4} + 2$.

It is enough to construct a multiplicative basis $\mathcal{B}$ of order h of $S(x^2+1,n)$ with

$$|\mathcal{B}| \leq n - n^{1/2} + n^{1/4} + 2.$$

First observe that

$$\left(a^2+1\right)\left((a+1)^2+1\right) = \left(a^2+a+1\right)^2+1. \tag{10}$$

Let

$$\mathcal{B} \overset{\text{def}}{=} \{x^2+1:\ 0 \le x \le n\} \setminus \{\left(a^2+a+1\right)^2+1:\ n^{1/2}+0.5 \le a^2+a+1 \le n\}$$

In order to prove that $\mathcal{B}$ is a multiplicative basis of order h it is enough to prove that for $1 \le x \le n$ the integer x^2+1 can be written as a product of h elements of $\mathcal{B}$.

If x is not of the form a^2+a+1 where $n^{1/2}+0.5 \le a^2+a+1 \le n$, then it is clear that

$$x^2+1 = b_1 b_2 b_3 \cdots b_h \tag{11}$$

where $b_1 = x^2+1 \in \mathcal{B}$ and $b_2 = b_3 = \cdots = b_h = 1 \in \mathcal{B}$.

If $x = a_1^2 + a_1 + 1$ for some integer a_1 and $n^{1/2} + 0.5 \leq a_1^2 + a_1 + 1 \leq n$, then by (10)

$$x^2 + 1 = \left(a_1^2 + a_1 + 1\right)^2 + 1 = \left(a_1^2 + 1\right)\left((a_1 + 1)^2 + 1\right).$$

Thus

$$x^2 + 1 = b_1 b_2 b_3 \cdots b_h,$$

with $b_1 = a_1^2 + 1,\ b_2 = (a_1 + 1)^2 + 1,\ b_3 = \cdots = b_h = 1$.

Computing the number of elements of $\mathcal{B}$ we get

$$|\mathcal{B}| \leq n - n^{1/2} + n^{1/4} + 2,$$

which was to be proved.

Theorem 2 (Fried, Gy.)

Let $f(x) \in \mathbb{Z}[x]$ be a polynomial of degree $r \geq 2$ and write $f(x)$ as a product of irreducible polynomials over $\mathbb{Z}[x]$, say

$$f(x) = f_1(x) f_2(x) \cdots f_s(x), \tag{12}$$

where s denotes the number of irreducible factors in (12). *Then*

$$\frac{n}{(\log n)^{s \log r / \log 2}} \ll G_2(S(f(x), n)).$$

Proof of Theorem 2

Throughout the proof $c_1, c_2, c_3, \ldots$ will denote constants depending only on the polynomial $f(x)$.

Let $\tau(a)$ denote the number of positive divisors of a positive integer a. It is well-known that

$$\sum_{a=1}^{n} \tau(a) = n \log n + O(n).$$

In 1952 Erdős extended this result to polynomials, namely he proved the following:

Lemma 1 (Erdős)

Let $f(x) \in \mathbb{Z}[x]$ be an irreducible polynomial. There exist positive integers c_1 and c_2 depending on $f(x)$ such that for $n \geq 2$ we have

$$c_1 n \log n < \sum_{a=1}^{n} \tau(f(a)) < c_2 n \log n. \tag{13}$$

Here we mention that Erdős gave an existence proof, and he could not give bounds on the order of magnitude of the constants c_1 and c_2 in Lemma 1.

Lapkova achieved some good bounds in the case of polynomials of degree 2.

In order to prove Theorem 2 we will need only the upper bound in (13).

Let s denote the number of irreducible factors $f_j(x)$ in (1). Using Erdős's lemma we will prove the following:

Lemma 2

There exists a constant c_3 depending only on the polynomial $f(x)$ such that for every integer n large enough we have that the set

$$F(f(x), n) \overset{\text{def}}{=} \{f(a) : n/4 \leq a \leq n \ and \ \tau(f(a)) < c_3(\log n)^s\} \tag{14}$$

has at least $n/(4r)$ different elements.

The proof of the lemma uses Erdős' upper bound (see (13)) and Cauchy Schwarz's inequality.

Next we prove the following:

Lemma 3

Let $\mathcal{B}$ be a multiplicative basis of $F(f(x), n)$ of order 2. Then

$$|\mathcal{B}| \gg \frac{n}{(\log n)^{s \log r / \log 2}}.$$

From Lemma 3 we immediately get Theorem 2. If $\mathcal{B}$ is a multiplicative basis of $S(f(x), n)$ then it is also a multiplicative basis of $F(f(x), n)$ by $F(f(x), n) \subseteq S(f(x), n)$.

Proof of Lemma 3

Define a graph $\mathcal{G}$ by the following: its vertices are the elements of $\mathcal{B}$. Two vertices v_1, v_2 are joined by an edge $\{v_1, v_2\}$ if and only if

$$v_1 v_2 \in F(f(x), n).$$

Then for the number of vertices and edges of $\mathcal{G}$ we have

$$|V(\mathcal{G})| = |\mathcal{B}| \quad \text{and} \quad |E(\mathcal{G})| \geq |F(f(x), n)| > c_6 n. \tag{15}$$

It is easy to prove that there exists a constant c_7 such that if $\{v_1, v_2\}$ is an edge of $\mathcal{G}$, then

$$v_1 > c_7 n \text{ or } v_2 > c_7 n. \tag{16}$$

We split the set of vertices $\mathcal{B}$ into two disjoint sets:

$$\mathcal{B}_1 = \{v \in \mathcal{B} : \ v > c_7 n\}$$
$$\mathcal{B}_2 = \{v \in \mathcal{B} : \ v \leq c_7 n\}$$

By (16) clearly for every edge $e = \{v_1, v_2\}$ of $\mathcal{G}$ we have $v_1 \in \mathcal{B}_1$ or $v_2 \in \mathcal{B}_1$. Thus if we denote by $d(v)$ the degree of a vertex $v \in \mathcal{B}$ in $\mathcal{G}$ then

$$|E(\mathcal{G})| \leq \sum_{v \in \mathcal{B}_1} d(v). \tag{17}$$

In Lemma 4 we give an estimate on the degree of a vertex of $\mathcal{B}_1$:

Lemma 4

For $v \in \mathcal{B}_1$ we have

$$d(v) \ll (\log n)^{s \log r / \log 2}$$

The proof of Lemma 4 is based on standard estimates for the number of solutions of a congruence

$$f(x) \equiv 0 \pmod{m}$$

using the discriminant of $f(x)$ and that for $v \in \mathcal{B}_1$ we have $\tau(v) \ll (\log n)^{s \log r / \log 2}$.

From Lemma 4 we immediately get Lemma 3. From Lemma 4, (15) and (17) follows

$$c_6 n < |E(\mathcal{G})| \leq \sum_{v \in \mathcal{B}_1} d(v) \ll \sum_{v \in \mathcal{B}_1} (\log n)^{s \log r / \log 2}$$
$$\ll |\mathcal{B}_1| (\log n)^{s \log r / \log 2} \ll |\mathcal{B}| (\log n)^{s \log r / \log 2}$$

thus

$$\frac{n}{(\log n)^{s \log r / \log 2}} \ll |\mathcal{B}|$$

which proves Lemma 3. This completes the proof of Theorem 2.

Let us see our last theorem:

Theorem 3 (Fried, Gy.)

Let $f(x) \in \mathbb{Z}[x]$ be a polynomial of degree $\deg f \geq 2$ and $u, a_1, a_2, \dots, a_k$ be positive integers such that

$$u \leq a_1 < a_2 < \cdots < a_k < 2u. \tag{18}$$

We define $\mathcal{W}$ by $\{f(a_1), f(a_2), \dots, f(a_k)\}$. If $\mathcal{B}$ is a multiplicative basis of order 2 of $\mathcal{W}$ then

$$|\mathcal{W}|^{2/3} \ll |\mathcal{B}|\,. \tag{19}$$

Proof of Theorem 3

We define the following graph $\mathcal{G}$. Its vertices are the elements of $\mathcal{B}$, so $V(\mathcal{G}) = \mathcal{B}$. There is an edge between the vertices $b_1 \in \mathcal{B}$ and $b_2 \in \mathcal{B}$ if and only if there exists an $1 \le i \le s$ such that

$$b_1 b_2 = f(a_i).$$

We will denote this edge by $\{b_1, b_2\}$.

Since $\mathcal{B}$ is a multiplicative basis of order 2 of $\mathcal{W}$, for the number of the edges of $\mathcal{G}$ we have

$$|E(\mathcal{G})| \ge |\mathcal{W}|. \tag{20}$$

We will color the edges of $\mathcal{G}$ by different colors where the number of colors depends only on the polynomial $f(x)$.

Based on a technical lemma it is possible to prove that there exist constants c_1 and c_2 such that if we color an edge $\{b_1, b_2\}$ of $\mathcal{G}$ by the first color if $b_1 \leq c_1$ or $b_2 \leq c_1$ and for $i \geq 2$ we color the edge $\{b_1, b_2\}$ of $\mathcal{G}$ by the i-th color if

$$c_2^{i-2} u \leq b_1 b_2 < c_2^{i-1} u, \tag{21}$$

then the graph $\mathcal{G}$ does not contain a cycle of length 4, where the edges of the cycle are colored by the same i-th color for an $i \geq 2$.

By the Kövári-Sós-Turán theorem we have that if a graph $\mathcal{G}$ has n vertices and it does not contain a cycle of length 4, than it has at most

$$1 + n + \left[\frac{1}{2} n^{3/2}\right] \tag{22}$$

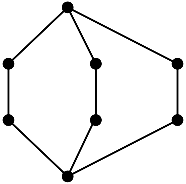
edges.

Since we have at most c_4 different colors we have

$$|\mathcal{W}| \leq |E(\mathcal{G})| \ll |V(\mathcal{G})|^{3/2} = |\mathcal{B}|^{3/2},$$

where the implied constant depend on the polynomial $f(x)$ q.e.d.

Probably, it can be proved that the subgraphs $\mathcal{G}_i$ of $\mathcal{G}$ formed by the edges of $\mathcal{G}$ colored by the i-th color (where $i \geq 2$) do not contain the following graph $\theta_{3,3}$:



From this, using Faudree and Simonovits theorem in extremal graph theory one may obtain the bound

$$|\mathcal{W}| \leq \sum_i E(\mathcal{G}_i) \ll c_1 \, |\mathcal{B}| + \sum_{i\geq 2} |V(\mathcal{G}_i)|^{1+1/3} \ll |\mathcal{B}|^{4/3} ,$$

from which

$$|\mathcal{B}| \gg |\mathcal{W}|^{3/4} \tag{23}$$

follows.

Here, we remark that the proof that these subgraphs of $\mathcal{G}$ do not contain $\theta_{3,3}$ can be rather lengthly and complicated, and the desired lower bound (23) is just slightly stronger than the one in Theorem 3 and it is also far from the truth. Thus we did not work out the details of the proof here.

Thank you for your attention!