Riley Becker and M. Ram Murty

*Diophantine $m$-tuples with the property $D(n)$*

Accepted manuscript

This is a preliminary PDF of the author-produced manuscript that has been peer-reviewed and accepted for publication. It has not been copyedited, proofread, or finalized by Glasnik Production staff.
DIOPHANTINE \( m \)-TUPLES WITH THE PROPERTY \( D(n) \)

RILEY BECKER AND M. RAM MURTY
Queen’s University, Canada

Abstract. Let \( n \) be a non-zero integer. A set of \( m \) positive integers \( \{a_1, a_2, \cdots, a_m\} \) such that \( a_ia_j + n \) is a perfect square for all \( 1 \leq i < j \leq m \) is called a Diophantine \( m \)-tuple with the property \( D(n) \). In a series of papers, Dujella studied the quantity \( M_n = \sup\{|S| : S \text{ has the property } D(n)\} \) and showed for \( |n| \geq 400 \) that \( M_n \leq 15.476 \log |n| \) and if \( |n| > 10^{100} \), then \( M_n < 9.078 \log |n| \). We refine his argument to show that \( C_n \leq 2 \log |n| + O\left(\frac{\log |n|}{\log \log |n|}\right) \), where the implied constant is effectively computable and \( C_n = \sup\{|S \cap [1, n^2]| : S \text{ has the property } D(n)\} \). Together with earlier work of Dujella, this implies \( M_n \leq 2.6071 \log |n| + O\left(\frac{\log |n|}{(\log \log |n|)^2}\right) \), where the implied constant is effectively computable.

1. Introduction

Let \( n \) be an non-zero integer. A set of \( m \) positive integers \( \{a_1, a_2, \cdots, a_m\} \) such that \( a_ia_j + n \) is a perfect square for all \( 1 \leq i < j \leq m \) is called a Diophantine \( m \)-tuple with the property \( D(n) \). Diophantus found that the quadruple \( \{1, 33, 68, 105\} \) has property \( D(256) \). Fermat found that \( \{1, 3, 8, 120\} \) has property \( D(1) \) [6]. Baker and Davenport [1] showed that Fermat’s quadruple cannot be extended to a quintuple with property \( D(1) \), and recently Dujella and others showed there is no sextuple with property \( D(1) \) and only finitely many, effectively computable quintuples with property \( D(1) \) using sophisticated methods including Baker’s theory of linear forms in logarithms [7, 8]. Finally, in a recent

2000 Mathematics Subject Classification. 11D25, 11N36.

Key words and phrases. Diophantine \( m \)-tuples, Gallagher’s sieve, Vinogradov’s inequality.
paper, He, Togbé and Ziegler [16] have shown that there are no Diophantine quintuples. Thus, the study of Diophantine \( m \)-tuples has ancient roots, and over the centuries the methods used to study them have ranged from the very elementary to the profoundly deep.

For example, for any natural number \( n \), we cannot have an infinite set \( \{a_1, a_2, \ldots \} \) with property \( D(n) \) because, by a famous theorem of Siegel [20], there are only finitely many integral points on the elliptic curve

\[
y^2 = (a_1 x + n)(a_2 x + n)(a_3 x + n).
\]

However, known bounds for the number of integral solutions to (1.1) depend on \( n, a_1, a_2, \) and \( a_3 \).

On the other hand, if we consider the hyperelliptic curve

\[
y^2 = (a_1 x + n)(a_2 x + n)(a_3 x + n)(a_4 x + n)(a_5 x + n),
\]

which has genus 2, a celebrated conjecture of Caporaso, Harris, and Mazur [3] predicts that (1.2) has a bounded number of integral points independent of \( n \) and \( a_1, \ldots, a_5 \). Thus, the quantity

\[
M_n = \sup\{|S| : S \text{ has the property } D(n)\}
\]

is conjectured to be bounded by an absolute constant independent of \( n \). These observations were made by Dujella [9].

Some progress was made recently towards this boundedness conjecture. In [11], Dujella and Luca show that \( M_n \) is bounded in terms of the number of prime factors of \( n \) for squarefree values of \( n \). In particular, \( M_p \) is absolutely bounded for primes \( p \). They also show that for almost all \( n \) (in the sense of natural density), one has the estimate \( M_n < \log \log |n| \).

There are other results related to this problem relevant to our discussion. For example, there is the paper by Evertse and Silverman that gives a bound for the number of integral solutions in terms of \( n, a_1, a_2, a_3 \) [13]. Another paper of Silverman [20] connects ranks of elliptic curves to the number of integral points. In particular, if \( M_n \) is unbounded, this would suggest that there are infinitely many elliptic curves of arbitrarily large rank, resolving a celebrated open question. This remark is not a rigorous statement but is inspired by the general feeling that “an elliptic curve with many integral points must have large rank” as expressed in [20].

By very elementary congruence consideration, one can deduce that \( M_n \leq 3 \) if \( n \equiv 2 \) (mod 4). To see this, note that

\[
a_i a_j + n \equiv 0 \text{ or } 1 \pmod{4}
\]

since any square is congruent to either 0 or 1 (mod 4). If \( n \equiv 2 \pmod{4} \), we must have

\[
a_i a_j \equiv 2 \text{ or } 3 \pmod{4},
\]
implying that at most one $a_i$ can be even. Then by a simple application of
the pigeonhole principle, $M_n \leq 3$. This observation was made independently
by Brown [2], Gupta and Singh [15] and Mohanty and Ramasamy [17].

Dujella [7, 9, 10] has written several papers dealing with estimates for $M_n$
as a function of $n$. He proved that $M_n \leq 15.476 \log |n|$ if $|n| > 400$ and that if
$n > 10^{100}$, then $M_n < 9.078 \log |n|$. Dujella uses a fundamental inequality due
to Vinogradov (see page 193 of [21]) and it seems that a sharper inequality
was obtained in the Russian edition (see pages 82 and 150 of [22]). If the
weaker inequality is used, then 15.476 should be replaced by 29.310 and 9.078
by 20.927. Since this subtle discrepancy may not be known to those who
have the Dover English translation of [22], we will elaborate and explain the
refinement. At the same time, we offer another approach to the estimate
of Vinogradov which is applicable not only to quadratic characters, but also
other characters as well. This will be discussed in the next section.

Dujella [8] decomposes his analysis into three components. He defines

$$
A_n = \sup \{ |S \cap [|n|^3, \infty)| : S \text{ has } D(n) \}
$$

$$
B_n = \sup \{ |S \cap (|n|^2, |n|^3)| : S \text{ has } D(n) \}
$$

$$
C_n = \sup \{ |S \cap [1, n^2]| : S \text{ has } D(n) \}
$$

He proves that $A_n \leq 21$ and $B_n < 0.65 \log |n| + 2.24$
for all $n$. He then improves these bounds in [10] to

$$
B_n < 0.6071 \log |n| + 2.152
$$

and

$$
C_n < 11.006 \log |n|.
$$

for $|n| > 400$. If $|n| > 10^{100}$, he shows that $C_n < 8.37 \log |n|$ and the final
result is derived by combining all of these estimates. Since the size of $A_n$
is globally bounded and $B_n$ is effectively bounded by $0.65 \log |n| + 2.24$, we will
define focus on $C_n$.

In this paper, our goal is to improve the estimate for $C_n$:

**Theorem 1.1.**

$$
C_n \leq 2 \log |n| + O \left( \frac{\log |n|}{(\log \log |n|)^2} \right).
$$

2. Preliminaries

In [10], the estimate for $C_n$ was derived using Gallagher’s larger sieve
[14] together with an estimate of Vinogradov [21]. As noted earlier Dujella
[8] states Vinogradov’s theorem from the Russian edition, omitting the 2 in Vinogradov’s original statement that

\[ \left| \sum_{a \in A} \sum_{b \in B} \left( \frac{ab + n}{p} \right) \right| \leq \sqrt{2p|A||B|}, \]

which appears in [21] and instead writing

(2.3) \[ \left| \sum_{a \in A} \sum_{b \in B} \left( \frac{ab + n}{p} \right) \right| \leq \sqrt{p|A||B|}. \]

In a later section, we will show that (2.3) holds, but it does not follow from Vinogradov’s original method [21] but rather from a later method explained in [22]. We will use the method of Gauss sums to derive (2.3) and it will be evident that the result holds in a more general context.

An important role is played by Gallagher’s sieve estimate:

**Proposition 2.1.** Let \( S \) be a set of integers contained in an interval of length \( N \). Let \( P \) be a finite set of primes. If for each prime \( p \in P \), we define \( u(p) = |S \pmod{p}| \), then

\[ |S| \leq \frac{\sum_{p \in P} \log p - \log N}{\sum_{p \in P} \frac{\log p}{\log(p)} - \log N}, \]

provided the denominator is positive.

**Proof.** This is Theorem 1 in [14] (see also Theorem 2.2.1 in [4]). \( \square \)

We have written this following the formulation in [4] for ease of application and clarity.

**Proposition 2.2.** Let \( \theta(x) = \sum_{p \leq x} \log p \), where the sum is over primes. Then

\[ |\theta(x) - x| < 3.965 \frac{x}{\log^2 x} \]

for \( x \geq 2 \).

**Proof.** This is Theorem 4 of Dusart [12]. \( \square \)

**Proposition 2.3.**

\[ \sum_{p \leq Q} \frac{\log p}{\sqrt{p} + 4} = 2\sqrt{Q} + O \left( \frac{\sqrt{Q}}{\log Q} \right). \]

The constant implied in the \( O \)-estimate is effectively computable.
Proof. By partial summation, we have that
\[ \sum_{p \leq x} f(p) \log p = \theta(x) f(x) - \int_2^x \theta(t) f'(t) dt \]
for any function \( f(t) \) which is differentiable on \([1, x]\). Applying this with \( f(t) = 1/(\sqrt{t} + 4) \) and using Proposition 2.2 gives the result. Indeed, we have that the sum equals
\[ \frac{Q}{\sqrt{Q} + 4} + O \left( \frac{\sqrt{Q}}{\log^2 Q} \right) + \int_2^Q \frac{dt}{2(\sqrt{t} + 4)^2 \sqrt{t}} + O \left( \int_2^Q \frac{dt}{\sqrt{t} \log^2 t} \right). \]
The first term is clearly
\[ \sqrt{Q} + O(1). \]
The third term is evidently
\[ \sqrt{Q} + O(\log Q). \]
Finally, the last term is estimated as follows. We break the integral into two parts:
\[ \int_2^{\sqrt{Q}} + \int_{\sqrt{Q}}^Q. \]
Then, the first of these integrals is \( O(Q^{1/4}) \). In the second integral, we see that as \( 1/\log^2 t \) is a decreasing function of \( t \), we obtain a final estimate of
\[ O \left( \frac{\sqrt{Q}}{\log^2 Q} \right). \]
Using Proposition 2.2, one can derive an explicit constant implied by our \( O \)-estimate. Of course, if we invoke the prime number theorem with the best known error term, the error in our assertion can be substantially improved. 

As Dujella pointed out to us, the version of Vinogradov’s theorem he used came from a later edition [22]. The improved version in the Russian edition is quite elegant and short, so we include it here for the benefit of those who do not have access to the Russian edition or do not know Russian. By Cauchy-Schwarz,
\[ \left| \sum_{a \in A} \sum_{b \in B} \left( \frac{ab + n}{p} \right) \right|^2 \leq |A| \sum_{a=0}^{p-1} \sum_{b \in B} \left( \frac{ab + n}{p} \right)^2. \]
Expanding the square on the right hand side, we get
\[ \left| \sum_{a \in A} \sum_{b \in B} \left( \frac{ab + n}{p} \right) \right|^2 \leq |A| \sum_{b,c \in B} S_{b,c}, \]
where
\[
S_{b,c} = \sum_{a=0}^{p-1} \left( \frac{ab + n}{p} \right) \left( \frac{ac + n}{p} \right).
\]
A straightforward calculation now shows that
\[
S_{b,c} = p \text{ if } b = c = 0 \text{ and zero if } bc = 0, b \neq c.
\]
Furthermore, if \( b = c \neq 0 \), then \( S_{b,c} = p - 1 \) and is zero otherwise. Injecting this into our sum gives
\[
\left| \sum_{a \in A} \sum_{b \in B} \left( \frac{ab + n}{p} \right) \right|^2 \leq |A| ((p - 1)|B| + \delta_B(0)),
\]
where
\[
\delta_B(0) = \begin{cases} 
1, & \text{if } 0 \in B \\
0, & \text{if } 0 \notin B.
\end{cases}
\]
Of course, this implies the bound \( \sqrt{p|A||B|} \) and is even better if \( 0 \notin B \). A modification of this method to deal with character sums instead of quadratic sums does not lead to similar estimates (as can be verified by the reader). Indeed, the difficulty arises in the explicit computation of \( S_{b,c} \) above. In the case of a non-quadratic character, one encounters Jacobi sums of the form
\[
\sum_{a=0}^{p-1} \chi(a)\chi(a + t).
\]
Thus, we offer below another derivation based on Gauss sums that does generalize to arbitrary characters (see Proposition 2.7).

**Proposition 2.4.** Let \( \chi \) be a Dirichlet character (i.e. a homomorphism from \((\mathbb{Z}/p\mathbb{Z})^\times\) to \(\mathbb{C}^\times\)). Let
\[
\tau = \sum_{c=1}^{p} \chi(c) e^{2\pi ic/p}.
\]
If \( \chi \) is non-trivial, then \(|\tau| = \sqrt{p}\). Moreover,
\[
\overline{\chi}(n) = \frac{1}{\tau} \sum_{c=1}^{p-1} \chi(c) e^{2\pi i nc/p}.
\]
**Proof.** This is Theorem 5.3.3 in [18].

With this proposition we can improve upon Vinogradov’s original theorem.

**Proposition 2.5.** Let \( p \) be an odd prime and \( n \) be an integer with \( \gcd(n, p) = 1 \). If \( A \subseteq \{1, \ldots, p - 1\} \) and \( B \subseteq \{0, 1, \ldots, p - 1\} \), then
\[
\sum_{a \in A} \sum_{b \in B} \left( \frac{ab + n}{p} \right) \leq \sqrt{p|A||B|}.
\]
Proof. Since it will be easier to deal with the sum if the \( a \) and \( b \) are separated, we write

\[
\sum_{a \in A} \sum_{b \in B} \left( \frac{ab + n}{p} \right) = \sum_{a \in A} \sum_{b \in B} \left( \frac{a}{p} \right) \left( \frac{b + na^{-1}}{p} \right).
\]

Replacing \( \left( \frac{b + na^{-1}}{p} \right) \) with its Gauss sum, we have

\[
\sum_{a \in A} \sum_{b \in B} \left( \frac{ab + n}{p} \right) = \sum_{a \in A} \sum_{b \in B} \left( \frac{a}{p} \right) \sum_{c \in \mathbb{Z}} \left( \frac{c}{p} \right) e^{2\pi i (b + na^{-1})c/p}.
\]

Applying the Cauchy-Schwarz inequality to the right hand side, we deduce

\[
\left| \sum_{a \in A} \sum_{b \in B} \left( \frac{ab + n}{p} \right) \right| \leq \frac{1}{\sqrt{p}} \sum_{c=0}^{p-1} \left( \left| \sum_{a \in A} \left( \frac{a}{p} \right) e^{2\pi i na^{-1}c/p} \right|^2 \right)^{1/4} \left( \left| \sum_{b \in B} e^{2\pi i bc/p} \right|^2 \right)^{1/4}
\]

and the right hand side is

\[
= \frac{1}{\sqrt{p}} \left( \sum_{a,a' \in A} \left( \frac{aa'}{p} \right) \sum_{c=0}^{p-1} e^{2\pi i (a^{-1} - a'^{-1})c/p} \right)^{1/4} \left( \sum_{b,b' \in B} \sum_{c=0}^{p-1} e^{2\pi i (b - b')c/p} \right)^{1/4}
\]

\[
= \frac{1}{\sqrt{p}} (|A|)^{1/2} (|B|)^{1/2}
\]

\[
= \sqrt{p} |A||B|.
\]

Now we correct this argument to include the case when \( 0 \in A \).

Proposition 2.6. Let \( p \) be an odd prime and \( n \) be an integer with \( \gcd(n, p) = 1 \). If \( A, B \subseteq \{0, \ldots, p-1\} \), then

\[
\sum_{a \in A} \sum_{b \in B} \left( \frac{ab + n}{p} \right) \leq \sqrt{p}|A||B| + |B|.
\]

Proof. Using the result of the previous proposition, we have

\[
\sum_{a \in A} \sum_{b \in B} \left( \frac{ab + n}{p} \right) = \sum_{a \in A \setminus \{0\}} \sum_{b \in B} \left( \frac{ab + n}{p} \right) + \delta_A(0) |B| \left( \frac{n}{p} \right)
\]

\[
\leq \sqrt{p}|A||B| + |B|,
\]
where
\[ \delta_A(a) = \begin{cases} 1, & \text{if } a \in A \\ 0, & \text{if } a \not\in A \end{cases} \]

As hinted earlier, Proposition 2.5 is true in a wider context and we record it here:

**Proposition 2.7.** Let \( p \) be an odd prime and \( n \) be an integer with \( \gcd(n, p) = 1 \). If \( A \subseteq \{1, \ldots, p-1\} \) and \( B \subseteq \{0, 1, \ldots, p-1\} \), then
\[
\sum_{a \in A} \sum_{b \in B} \chi(ab + n) \leq \sqrt{p}|A||B|,
\]
for any non-trivial character \( \chi \pmod{p} \).

### 3. Proof of the Main Theorem

Let \( S = \{a_1, \ldots, a_m\} \), with all \( a_i \leq n^2 = N \). As noted earlier, this estimate for \( C_n \) combined with the estimates for \( A_n \) and \( B_n \) of [10] gives the final estimate. This is consonant with Remark 4.1 of [10].

In Dujella [10], it is shown for each prime \( p \leq Q \) with \( (p, n) = 1 \) that
\[
|S_p| \leq \min\{\sqrt{p} + 3, p\},
\]
where \( S_p = S \pmod{p} \). This is done using the Legendre symbol and Vinogradov’s theorem as stated in equation (2.3) and showing that
\[
|S_p|(|S_p| - 3) \leq |S_p|\sqrt{p}.
\]

Using the version of the Vinogradov’s theorem given in Proposition 2.6 gives instead that \( |S_p|(|S_p| - 3) \leq |S_p|\sqrt{p} + 1 \), or that
\[
|S_p| \leq \sqrt{p} + 4.
\]
Thus, we use \( u(p) = \min\{\sqrt{p} + 4, p\} \) in Gallagher’s sieve, noting that for \( p \geq 7 \), \( |S_p| \leq \sqrt{p} + 4 \). In the denominator, we have from Proposition 2.3
\[
\sum_{p \leq Q} \frac{\log p}{\sqrt{p} + 4} = 2\sqrt{Q} + O\left(\frac{\sqrt{Q}}{\log^2 Q}\right) - \sum_{p \mid n} \frac{\log p}{\sqrt{p} + 4}
\]
and the final term is bounded as follows. Let \( p_i \) denote the \( i \)-th prime. Then, letting \( \omega(n) \) denote the number of distinct prime divisors of \( n \), we have again using Proposition 2.3, and noting that \( \frac{\log x}{\sqrt{x} + 4} \) is eventually a decreasing function,
\[
\sum_{p \mid n} \frac{\log p}{\sqrt{p} + 4} \leq \sum_{i \leq \omega(n)} \frac{\log p_i}{\sqrt{p_i} + 4} + O(1) = 2p^{1/2} + O\left(\frac{\sqrt{\omega(n)}}{\log^2 \omega(n)}\right) \ll (\log |n|)^{1/2},
\]
the final estimate coming from the bound that \( p_i \ll i \log i \) and Ramanujan’s bound

\[
\omega(n) \ll \frac{\log n}{\log \log n}.
\]

Therefore

\[
\sum_{p \leq Q} \frac{\log p}{\sqrt{p} + 4} \geq 2\sqrt{Q} + O((\log |n|)^{1/2}) + O\left(\frac{\sqrt{Q}}{\log^2 Q}\right).
\]

We insert this into Gallagher’s sieve to get an estimate for \(|S|\) of

\[
\theta(Q) - \log N \leq 2\sqrt{Q} + O((\log |n|)^{1/2}) + O\left(\frac{\sqrt{Q}}{\log^2 Q}\right) - \log N.
\]

We choose \( Q = \frac{(1+\delta)^2}{4} (\log N)^2 \) so that the denominator is

\[
\delta \log N + O\left(\frac{\log N}{(\log \log N)^2}\right) + O((\log |n|)^{1/2}) = 2\delta \log n + O((\log |n|)^{1/2}),
\]

since \( N = n^2 \). Now

\[
\theta(Q) = Q + O\left(\frac{Q}{\log^2 Q}\right)
\]

so that the set \( S \) has size at most

\[
2 \log |n| + O\left(\frac{\log |n|}{(\log \log |n|)^2}\right),
\]

where we have chosen \( \delta = 1 \) to minimize the coefficient of \( \log |n| \). This completes the proof.

4. Concluding remarks

There are several ways in which one can attempt to improve our estimate in the main theorem. One way is to improve the estimate of Proposition 2.5. If for example, one could show that \( |S_p| = o(\sqrt{p}) \), then our proof would lead to a final estimate of \( o(\log |n|) \), as is readily verified by the reader.

On the other hand, the sum in Proposition 2.5 is trivially less than or equal to \( |A||B| \). If one could improve this to \( \delta |A||B| \), for any \( \delta < 1 \), then one can deduce boundedness of \( M_n \) for all values of \( n \). Indeed, following Dujella [10],

\[
|S_p|^2 - 3|S_p| \leq \sum_{a \in S_p} \sum_{b \in S_p} \left( \frac{ab + n}{p} \right) < \delta |S_p|^2,
\]

implies \( |S_p| \leq 3/(1-\delta) \) so that \( S_p \) is bounded. Injecting this into our argument above leads to the boundedness of \( C_n \) and \( B_n \) and consequently \( M_n \). The point is that to include \( B_n \) in the analysis, we need to take \( N = n^3 \) and \( Q \) is chosen to be a sufficiently large multiple of \( \log N \).
It is possible to determine the implied constant in the $O$-estimate of our main theorem using Proposition 2.2 and partial summation. To keep our derivation reasonably elegant, we have not done so here. Of course, using the prime number theorem with error term, one could replace this error by

$$O((\log |n|)e^{-c\sqrt{\log \log |n|}})$$

for some $c > 0$. Though there are sharp effective error terms, computing precise numerical constants is quite cumbersome. For the ambitious reader, we suggest reasonably painless path to derive explicit constants.

We need a lower bound for

$$\sum_{7 \leq p \leq Q, (p,n) = 1} \frac{\log p}{\sqrt{p} + 4}.$$ 

This is easily seen to be

$$\geq \sum_{7 \leq p \leq Q, (p,n) = 1} \frac{\log p}{\sqrt{p}} - 4 \sum_{7 \leq p \leq Q} \frac{\log p}{p}.$$ 

These sums are more amenable for the insertion of Dusart’s explicit estimates via partial summation. The only snag is that we also need to estimate the sum

$$\sum_{p|n} \frac{\log p}{\sqrt{p}}$$

which of course can be dealt with the more explicit estimates for the $i$-th prime and $\omega(n)$ in the form

$$p_i < i \log i + i \log \log i$$

for $i \geq 6$. Also, one can apply an estimate due to Robin [19]: for $n \geq 3$,

$$\omega(n) \leq \frac{\log n}{\log \log n} + 1.45743 \frac{\log n}{(\log \log n)^2}.$$ 

We leave the details to the assiduous reader.

Acknowledgements.

The first author is supported by an NSERC Undergraduate Student Research Award. The second author’s research is partially supported by an NSERC Discovery grant. We thank Andrej Dujella for his comments on an earlier version of our paper as well as bringing [22] to our attention. We also thank Ahmet Güloğlu and the referees for their helpful and meticulous corrections and suggestions.
DIOPHANTINE $m$-TUPLES WITH THE PROPERTY $D(n)$

References


Riley Becker
Department of Mathematics, Queen’s University, Kingston, Ontario, K7L 3N6, Canada.
E-mail: rileydebecker@gmail.com
M. Ram Murty
Department of Mathematics, Queen’s University, Kingston, Ontario, K7L 3N6, Canada.
E-mail: murty@queensu.ca