Sums and products with smooth numbers

William D. Banks, David J. Covert *

Department of Mathematics, University of Missouri, Columbia, MO 65211, USA

A R T I C L E   I N F O

Article history:
Received 22 September 2010
Accepted 2 November 2010
Available online 12 February 2011
Communicated by Robert C. Vaughan

MSC:
11B75
11N25

Keywords:
Sumset
Productset
Smooth numbers
Combinatorics
Number theory

A B S T R A C T

We estimate the sizes of the sumset \(A + A\) and the productset \(A \cdot A\) in the special case that \(A = S(x, y)\), the set of positive integers \(n \leq x\) free of prime factors exceeding \(y\).

1. Background

For any nonempty subset \(A\) of a ring, the sumset and productset of \(A\) are defined as

\[ A + A = \{ a + a' : a, a' \in A \} \quad \text{and} \quad A \cdot A = \{ a \cdot a' : a, a' \in A \}, \]

respectively. A famous problem of Erdős and Szemerédi [6] asks one to show that the sumset and productset of a finite set of integers cannot both be small.

Conjecture (Erdős–Szemerédi). For any fixed \(\delta > 0\) the lower bound

\[ \max \left\{ |A + A|, |A \cdot A| \right\} \gg |A|^{2-\delta} \]

holds for all finite sets \(A \subset \mathbb{Z}\).

* Corresponding author.
E-mail addresses: BanksWD@missouri.edu (W.D. Banks), CovertDj@missouri.edu (D.J. Covert).

0022-314X/$ – see front matter Published by Elsevier Inc.
doi:10.1016/j.jnt.2010.11.001
Erdős and Szemerédi [6] took the first step towards this conjecture by showing that for some \( \epsilon > 0 \), one has a lower bound of the form
\[
\max\{ |A + A|, |A \cdot A| \} \geq c(\epsilon)|A|^{1+\epsilon}
\]  
for all finite sets \( A \subset \mathbb{Z} \). Nathanson [10] gave the first explicit bound by showing that one can take \( \epsilon = \frac{1}{31} \) and \( c(\epsilon) = 0.00028 \ldots \) in this inequality, and later, Ford [8] showed that \( \epsilon = \frac{1}{15} \) is acceptable. Establishing an important connection between the sum–product problem and geometric incidence theory, Elekes [3] showed that one can take \( \epsilon = \frac{1}{4} \) via a clever application of the Szemerédi–Trotter incidence theorem (which counts incidences between points and lines in the plane); moreover, his argument readily extends to finite sets of real numbers. Further improvements, including the best known bound to date, have been given by Solymosi [12,13]; he has shown that (1) holds with any \( \epsilon < \frac{1}{3} \) for all finite sets \( A \subset \mathbb{R} \).

Although the Erdős–Szemerédi conjecture remains open, it is known that the productset must be large whenever the sumset is sufficiently small. In fact, Nathanson and Tenenbaum [11] have shown that
\[
|A \cdot A| \geq c|A|^2 \log |A|
\]
if \( |A + A| \leq 3|A| - 4 \).

The aforementioned best known bound to date, given by Solymosi [13], follows from his more general inequality
\[
|A + A|^2|A \cdot A| \geq \frac{|A|^4}{4|A|}\log |A|.
\]  
Note that (3) provides a quantitative generalization of the Nathanson–Tenenbaum result (2) (see also the results in [3,4,12]); it implies that \( |A \cdot A| \geq |A|^{2-\delta} \) whenever \( |A + A| < |A|^{1+\epsilon} \), where \( \delta \to 0 \) as \( \epsilon \to 0 \).

In the opposite direction, Chang [2] has shown that the sumset must be large whenever the productset is sufficiently small. More precisely, she has shown that
\[
|A + A| > 36^{-\alpha} |A|^2 \quad \text{if} \quad |A \cdot A| < \alpha|A| \quad \text{for some constant} \ \alpha.
\]  

A great deal of attention has also been given to the sum–product problem in other rings, including (but not limited to) finite fields, polynomial rings, and matrix rings. For a thorough account of the subject, we refer the reader to [14] and the references contained therein.

2. Statement of results

Let \( \Omega \) be any infinite collection of finite sets within a given ring. We shall say that \( \Omega \) has the Erdős–Szemerédi property if
\[
\max\{ |A + A|, |A \cdot A| \} = |A|^{2+o(1)} \quad \text{as} \quad |A| \to \infty \quad \text{with} \quad A \in \Omega.
\]

Then, the Erdős–Szemerédi conjecture is the assertion that the collection consisting of all finite sets of integers has the Erdős–Szemerédi property.

In this paper, we study the Erdős–Szemerédi property with collections of sets of smooth numbers, i.e., sets of the form
\[
S(x, y) = \{ n \leq x: P^+(n) \leq y \} \quad (x \geq y \geq 2),
\]
where $P^+(n)$ denotes the largest prime factor of an integer $n \geq 2$, and $P^+(1) = 1$. These sets are well known in analytic number theory; for a background on integers free of large prime factors, we refer the reader to [15, Chapter III.5] (see also the survey [9]).

**Theorem 1.** There is an absolute constant $c > 0$ for which the collection
\[
\Omega = \{S(x, y) : 2 \leq y \leq c \log x\}
\]
has the Erdős–Szemerédi property.

**Remarks.** In Theorem 4 we show that for values of $y$ of size $o(\log x)$, the productset of $A = S(x, y)$ has size $|A|^{1+o(1)}$; thus, only the sumset is large in this region. Using only Theorem 4 and Chang’s result (4), one can show that the smaller collection
\[
\Omega = \{S(x, y) : 2 \leq y \leq C(\log \log \log x)(\log \log \log \log x)\}
\]
has the Erdős–Szemerédi property for any constant $C < 1/\log 2$.

**Theorem 2.** Let $f$ be an arbitrary real-valued function such that $f(x) \to \infty$ as $x \to \infty$. Then, the collection
\[
\Omega = \{S(x, y) : f(x) \log x \leq y \leq x\}
\]
has the Erdős–Szemerédi property.

**Remark.** For slightly larger values of $y$ exceeding $(\log x) f(x)$ we show that the sumset of $A = S(x, y)$ has size $|A|^{1+o(1)}$ (see Theorem 5), and hence only the productset is large in this region.

Since each set $S(x, y)$ is multiplicatively defined, it is quite difficult to estimate the size of the sumset $S(x, y) + S(x, y)$ for values of $y$ close to $\log x$. It is reasonable to expect that for every fixed $\kappa > 0$ one has
\[
|S(x, y) + S(x, y)| = |S(x, y)|^{2+o(1)} \quad (x \to \infty, \ y = \kappa \log x).
\]

In view of (12), the Erdős–Szemerédi conjecture implies that this is true. A partial result in this direction is provided by (13). We also expect that for any fixed $A > 1$ one has
\[
|S(x, y) + S(x, y)| = |S(x, y)|^{\beta_A + o(1)} \quad (x \to \infty, \ y = (\log x)^A)
\]
for some constant $\beta_A$ in the open interval $(1, 2)$. For $A > 2$, a partial result in this direction is provided by Theorem 8.

**3. Preliminaries**

As before, we write
\[
S(x, y) = \{n \leq x : P^+(n) \leq y\} \quad (x \geq y \geq 2),
\]
and we now set
\[
\Psi(x, y) = |S(x, y)| \quad (x \geq y \geq 2).
\]
We also put
\[ G(t) = \log(1 + t) + t \log(1 + t^{-1}) \quad (t > 0). \]

From this definition we immediately derive the crude estimates
\[ G(t) = \log t \left\{ 1 + O \left( \frac{1}{\log t} \right) \right\} \quad (t \geq 2) \quad (5) \]
and
\[ G(t) = t \log t^{-1} \left\{ 1 + O \left( \frac{1}{\log t} \right) \right\} \quad (0 < t \leq 1/2). \quad (6) \]

The following result is due to de Bruijn [1].

**Lemma 1.** Uniformly for \( x \geq y \geq 2 \) we have
\[ \log \Psi(x, y) = \frac{\log x}{\log y} G \left( \frac{y}{\log x} \right) \left\{ 1 + O \left( \frac{1}{\log \log 2 x} \right) \right\}. \]

For smaller values of \( y \), we need the following result of Ennola [5].

**Lemma 2.** Uniformly for \( 2 \leq y \leq \sqrt{\log x \log \log x} \) we have
\[ \Psi(x, y) = \frac{1}{\pi(y)!} \prod_{p \leq y} \log x \log p \left\{ 1 + O \left( \frac{y^2}{\log x \log y} \right) \right\}, \]
where \( \pi(y) = |\{p \leq y\}| \).

For any finite set of primes \( S \), let \( O^*_S \) denote the group of \( S \)-units in \( \mathbb{Q}^* \); that is,
\[ O^*_S = \left\{ a/b \in \mathbb{Q}^* : p \mid ab \Rightarrow p \in S \right\}. \]

The next statement is a special case of a more general result of Evertse on solutions to \( S \)-unit equations (see [7, Theorem 3]).

**Lemma 3.** Given \( a_1 \cdots a_n \in \mathbb{Q}^* \) and a finite set of primes \( S \) of cardinality \( |S| = s \), the \( S \)-unit equation
\[ a_1 u_1 + \cdots + a_n u_n = 1 \quad (u_1, \ldots, u_n \in O^*_S) \]
has at most \( (2^{35} n^2)^{n^2} \) solutions \((u_1, \ldots, u_n)\) with \( \sum_{j \in \mathcal{J}} a_j u_j \neq 0 \) for every nonempty subset \( \mathcal{J} \subseteq \{1, \ldots, n\} \).

To get a better handle on productsets of smooth numbers, we shall apply the following technical lemma.

**Lemma 4.** We have
\[ \Psi \left( \frac{x^2}{y}, y \right) \leq |S(x, y) \cdot S(x, y)| \leq \Psi \left( \frac{x^2}{y} \right) \quad (x \geq y \geq 2). \]
Proof. It is easy to see that \( S(x, y) \cdot S(x, y) \subseteq S(x^2, y) \), which yields the second inequality. For the first inequality, it suffices to show that \( S(x^2/y, y) \) is contained in the productset \( S(x, y) \cdot S(x, y) \). To this end, let \( n \in S(x^2/y, y) \), and let \( d \) be the largest divisor of \( n \) that does not exceed \( x \). Note that \( \max\{P^+(d), P^+(n/d)\} \leq y \). There are three possibilities for the number \( d \):

(i) \( d > x/y \);
(ii) \( d = n \leq x/y \);
(iii) \( d \leq x/y \) and \( d < n \).

In case (i) we have \( n/d \leq x \), hence we can write \( n = d \cdot (n/d) \) where \( d \) and \( n/d \) both lie in \( S(x, y) \); this shows that \( n \in S(x, y) \cdot S(x, y) \) as required. In case (ii) the number \( n \) lies in the set \( S(x/y, y) \), which is a subset of \( S(x, y) \cdot S(x, y) \). To finish the proof, we need only show that the case (iii) is not possible. Indeed, suppose \( d \leq x/y \) and \( d < n \), and let \( p \) be any prime factor of \( n/d \); then \( p \leq P^+(n/d) \leq y \), \( dp \mid n \), and \( dp \leq x \), which contradicts the maximal property of \( d \). \( \square \)

4. Small values of \( y \)

Theorem 3. There is an absolute constant \( c > 0 \) such that the estimate

\[
|S(x, y) + S(x, y)| \sim \frac{1}{2} \Psi(x, y)^2 \quad (x \to \infty)
\]

holds uniformly for \( 2 \leq y \leq c \log x \).

Proof. We have

\[
\Psi(x, y)^2 = |S(x, y)|^2 = \sum_{n \in S(x, y) + S(x, y)} \sum_{m_1, m_2 \in S(x, y)} 1.
\]

Using the Cauchy inequality it follows that

\[
\Psi(x, y)^4 \leq |S(x, y) + S(x, y)| \cdot |T|,
\]

where \( T \) is the set of quadruples \((m_1, m_2, m_3, m_4)\) with entries in \( S(x, y) \) such that \( m_1 + m_2 = m_3 + m_4 \). It is easy to see that there are precisely \( 2\Psi(x, y)^2 - \Psi(x, y) \) quadruples in \( T \) for which \( m_1 = m_2 \) or \( m_1 = m_4 \). Let \( T^* \) be the set of quadruples in \( T \) with \( m_1 \neq m_2 \) and \( m_1 \neq m_4 \) (thus, \( m_2 \neq m_3 \) and \( m_2 \neq m_4 \) as well). If we put \( a_1 = a_2 = 1 \) and \( a_3 = -1 \), the equation \( m_1 + m_2 = m_3 + m_4 \) becomes

\[
a_1 u_1 + a_2 u_2 + a_3 u_3 = 1, \quad \text{(7)}
\]

where

\[
u_1 = \frac{m_1}{m_4}, \quad u_2 = \frac{m_2}{m_4} \quad \text{and} \quad u_3 = \frac{m_3}{m_4}. \quad \text{(8)}
\]

Let \( S \) be the set of primes \( p \leq y \), and let \( O^*_S \) be the group of \( S \)-units in \( \mathbb{Q}^* \). According to Lemma 3, there are at most \((2^{359})^{27\pi(y)} \) solutions to the \( S \)-unit equation (7) with \( u_j \in O^*_S \), \( j = 1, 2, 3 \), and \( \sum_{j \in T} a_j u_j \neq 0 \) for each nonempty subset \( T \subseteq \{1, 2, 3\} \). On the other hand, for every fixed solution \((u_1, u_2, u_3)\) to (7) there are at most \( \Psi(x, y) \) quadruples \((m_1, m_2, m_3, m_4)\) in \( T^* \) for which (8) holds (since each choice of \( m_4 \in S(x, y) \) determines \( m_1, m_2, m_3 \) uniquely). Putting everything together, it follows that the bound

\[
\Psi(x, y)^4 \leq |S(x, y) + S(x, y)| \cdot \left(2\Psi(x, y)^2 - \Psi(x, y) + \exp(c_1 y/\log y)\Psi(x, y)\right)
\]
holds with some absolute constant $c_1 > 0$. Taking into account the trivial upper bound

$$|S(x, y) + S(x, y)| \leq \frac{1}{2}(\Psi(x, y)^2 + \Psi(x, y)),$$

it suffices to show that there is an absolute constant $c > 0$ such that for all sufficiently large $x$, we have

$$\exp(c_1 y/\log y) \leq \Psi(x, y)^{1/2} \quad (2 < y < c \log x). \quad (9)$$

For every sufficiently large integer $N$, Lemma 1 implies that:

$$\log \Psi(x, y) \geq \frac{1}{2} \log x G\left(\frac{y}{\log x}\right) \quad (x > N)$$

if $x$ is sufficiently large. Let $N \geq 2$ be fixed with this property. For every sufficiently small constant $c > 0$ we also have by (6):

$$G(t) \geq \frac{1}{2} t \log t^{-1} \quad (0 < t < c).$$

Let $0 < c \leq e^{-8c_1}$ be fixed with this property. Combining the two bounds, we see that

$$\log \Psi(x, y) \geq \frac{\log(1/c)}{4} \frac{y}{\log y} \geq 2c_1 \frac{y}{\log y} \quad (N < y < c \log x)$$

if $x$ is large enough; this implies (9) in the range $N < y < c \log x$. For the smaller values of $y$ in the range $2 < y \leq N$, we simply observe that $\exp(c_1 y/\log y) = O(1)$, whereas

$$\Psi(x, y) \geq \Psi(x, 2) = 1 + \left\lfloor \frac{\log x}{\log 2} \right\rfloor \to \infty \quad \text{as} \quad x \to \infty.$$

Hence, (9) also holds for these values of $y$ if $x$ is sufficiently large. This completes the proof. \[\square\]

**Theorem 4.** Suppose that $y \geq 2$ and $y = o(\log x)$. Then

$$|S(x, y) \cdot S(x, y)| = \Psi(x, y)^{1+o(1)}.$$

**Proof.** By Lemma 4 we have

$$\Psi(x, y) \leq \Psi\left(\frac{x^2}{y}, y\right) \leq |S(x, y) \cdot S(x, y)| \leq \Psi(x^2, y),$$

hence it suffices to show that $\Psi(x^2, y) = \Psi(x, y)^{1+o(1)}$ as $x \to \infty$.

First, suppose that $2 < y \leq \sqrt{\log x}$. By Lemma 2 we have

$$\Psi(x, y) \sim \frac{1}{\pi(y)!} \prod_{p \leq y} \frac{\log x}{\log p} \quad (x \to \infty)$$

and

$$\Psi(x^2, y) \sim \frac{1}{\pi(y)!} \prod_{p \leq y} \frac{\log x^2}{\log p} \sim 2^{\pi(y)}\Psi(x, y) \quad (x \to \infty).$$
Since the inequality $\pi(y)! \leq y^{\pi(y)}$ implies

$$\Psi(x, y) \geq (1 + o(1)) \left( \frac{\log x}{y \log y} \right)^{\pi(y)} \geq (1 + o(1)) \left( \frac{2\sqrt{\log x}}{\log \log x} \right)^{\pi(y)},$$

it follows that $2^{\pi(y)} = \Psi(x, y)^{o(1)}$; thus, $\Psi(x^2, y) = \Psi(x, y)^{1+o(1)}$ as required.

Next, suppose that $y > \sqrt{\log x}$ and $y = o(\log x)$ as $x \to \infty$. Using Lemma 1 together with (6) we see that the estimate

$$\log \Psi(z, y) = \frac{y}{\log y} \left\{ \log \left( \frac{\log z}{y} \log \left( \frac{y}{\log x} \right) \right) \{1 + o(1)\} \right\},$$

holds uniformly for all $z$ in the range $x \leq z \leq x^2$. Applying this estimate with $z = x$ and with $z = x^2$, we derive that $\Psi(x^2, y) = \Psi(x, y)^{1+o(1)}$ in this case as well. □

5. Large values of $y$

For values of $y$ exceeding any fixed power of $\log x$, we have:

**Theorem 5.** Suppose that $(\log y) / \log \log x \to \infty$. Then,

$$\left| S(x, y) + S(x, y) \right| = \Psi(x, y)^{1+o(1)} \ (x \to \infty).$$

**Proof.** Using Lemma 1 and (5) we see that

$$\log \Psi(x, y) \sim \frac{\log x}{\log y} G \left( \frac{y}{\log x} \right) \sim \frac{\log x}{\log y} (\log y - \log \log x) \sim \log x \ (x \to \infty),$$

since $(\log \log x) / \log y \to 0$; that is,

$$\Psi(x, y) = x^{1+o(1)} \ (x \to \infty).$$

Using the trivial bounds

$$\Psi(x, y) \leq \left| S(x, y) + S(x, y) \right| \leq 2x$$

together with the previous estimate, we obtain the desired result. □

**Theorem 6.** Let $y / \log x \to \infty$. Then,

$$\left| S(x, y) \cdot S(x, y) \right| = \Psi(x, y)^{2+o(1)} \ (x \to \infty). \ (10)$$

**Proof.** In the case that $(\log y) / \log \log x \to \infty$, we can apply Theorem 5 together with (3) to obtain (10) immediately. Thus, we can assume that $\log y \asymp \log \log x$. Since $y / \log x \to \infty$, we derive from Lemma 1 and (5) the estimate

$$\log \Psi(x, y) = \frac{\log x}{\log y} \log \left( \frac{y}{\log x} \right) \{1 + o(1)\}, \ (11)$$

W.D. Banks, D.J. Covert / Journal of Number Theory 131 (2011) 985–993 991
whereas both \( \log \Psi\left( \frac{x^2}{y}, y \right) \) and \( \log \Psi\left( x^2, y \right) \) are of the size

\[
\frac{\log x}{\log y} \log \left( \frac{y}{\log x} \right) \left( 2 + o(1) \right).
\]

Therefore,

\[
\Psi\left( \frac{x^2}{y}, y \right) = \Psi\left( x, y \right)^{2 + o(1)} \quad \text{and} \quad \Psi\left( x^2, y \right) = \Psi\left( x, y \right)^{2 + o(1)},
\]

and the estimate (10) follows from Lemma 4. \( \square \)

6. Intermediate values of \( y \)

**Theorem 7.** Suppose that \( y = \kappa \log x \), where \( \kappa > 0 \) is fixed. Then,

\[
\left| S(x, y) \cdot S(x, y) \right| = \Psi(x, y)^{\alpha_k + o(1)} \quad (12)
\]

and

\[
\left| S(x, y) + S(x, y) \right| \geq \Psi(x, y)^{(4 - \alpha_k)/2 + o(1)}, \quad (13)
\]

where

\[
\alpha_k = \frac{2 \log(1 + \kappa/2) + \kappa \log(1 + 2/\kappa)}{\log(1 + \kappa) + \kappa \log(1 + 1/\kappa)}.
\]

**Remark.** For every positive real number \( \kappa \) we have \( 1 < \alpha_k < 2 \). Also, \( \alpha_k \to 1 \) as \( \kappa \to 0^+ \) and \( \alpha_k \to 2 \) as \( \kappa \to \infty \).

**Proof.** First note that (13) follows from combining (12) and (3). It remains to prove (12). By Lemma 1 we have

\[
\log \Psi(x, y) = (G(\kappa) + o(1)) \frac{\log x}{\log \log x} \quad (x \to \infty)
\]

and

\[
\log \Psi(x^2, y) = (2G(\kappa/2) + o(1)) \frac{\log x}{\log \log x} \quad (x \to \infty),
\]

where the functions implied by \( o(1) \) depend only on \( \kappa \). Since \( G \) is continuous it is also easy to see that

\[
\log \Psi\left( \frac{x^2}{y}, y \right) = (2G(\kappa/2) + o(1)) \frac{\log x}{\log \log x} \quad (x \to \infty).
\]

Using Lemma 4, the above estimates, and the fact that \( \alpha_k = 2G(\kappa/2)/G(\kappa) \), the result follows. \( \square \)

**Theorem 8.** Suppose that \( y \asymp (\log x)^A \), where \( A > 2 \) is fixed. Then,

\[
\left| S(x, y) + S(x, y) \right| \leq \Psi(x, y)^{\frac{A}{A-1} + o(1)} \quad (x \to \infty).
\]
Proof. If \( y \approx (\log x)^A \) for some \( A > 1 \), then the estimate \( \Psi(x, y) = x^{A-1+o(1)} \) follows immediately from (11). Taking into account the trivial bound \( |S(x, y) + S(x, y)| \leq 2x \), we obtain the stated result (which is nontrivial in the range \( A > 2 \)). \( \square \)

Acknowledgments

The authors would like to thank Derrick Hart, Alex Iosevich, and Igor Shparlinski for helpful conversations.

References