EXTREMAL PROPERTIES OF PRODUCT SETS

KEVIN FORD

To Sergei Vladimirovich Konyagin on the occasion of his 60th birthday

ABSTRACT. We find the nearly optimal size of a set \( A \subset [N] := \{1, \ldots, N\} \) so that the product set \( AA \) satisfies either (i) \( |AA| \sim |A|^2/2 \) or (ii) \( |AA| \sim |[N]|^2 \). This settles problems recently posed in a paper of Cilleruelo, Ramana and Ramaré.

1. INTRODUCTION

For \( A, B \subset \N \) let \( AB \) denote the product set \( \{ab : a \in A, b \in B\} \). In the special case \( [N] = \{1, 2, 3, \ldots, N\} \), denote by \( M_N = |[N][N]| \) the number of distinct products in an \( N \) by \( N \) multiplication table. In a recent paper [CRR17] of Cilleruelo, Ramana and Ramaré (see also Problems 15,16 in [CRS18]), the following problems were posed:

1. [CRR17, Problem 1.2]. If \( A \subset [N] \) and \( |AA| \sim |A|^2/2 \), is \( |A| = o(N/\log^{1/2} N) \)?
2. [CRR17, Problem 1.4]. If \( A \subset [N] \) and \( |AA| \sim M_N \), is \( |A| \sim N \)?

In this note, we answer both questions in the negative. Our results are based on a careful analysis of the structure of \([N][N]\) developed in [For08a] and [For08b]. Let

\[
\theta = \frac{1}{2} - \frac{\log\log 4}{\log 4} = 1 - \frac{1 + \log\log 4}{\log 4} = 0.04303566 \ldots
\]

From [For08a], we have

\[
M_N \approx \frac{N^2}{(\log N)^{3/2}}.
\]

In light of the elementary inequalities \( |AA| \leq \min(|A|^2, M_N) \), it follows that if \( |AA| \sim \frac{1}{2}|A|^2 \), then \( |A| \) cannot be of order larger than \( M_N^{1/2} \), and if \( |AA| \sim M_N \), then \( |A| \) cannot have order of growth smaller than \( M_N^{1/2} \). As we shall see, \( M_N^{1/2} \) turns out to be close the threshold value of \( |A| \) for each of these properties to hold.

Theorem 1. Let \( D > 7/2 \). For each \( N \geq 10 \) there is a set \( A \subset [N] \) of size

\[
|A| \geq \frac{N}{(\log N)^\theta (\log\log N)^D},
\]

for which \( |AA| \sim |A|^2/2 \) as \( N \to \infty \).

Consequently, the largest size \( T_N(\varepsilon) \) of a set \( A \) with \( |AA| \geq (1 - \varepsilon)|A|^2/2 \) satisfies

\[
\frac{N}{(\log N)^\theta (\log\log N)^{7/2 + o(1)}} \ll T_N(\varepsilon) \ll \frac{N}{(\log N)^\theta (\log\log N)^{3/4}}.
\]

\(^1\)

Research of the author supported in part by individual NSF grant DMS-1501982. Some of this work was carried out at MSRI, Berkeley during the Spring semester of 2017, partially supported by NSF grant DMS-1440140.
Theorem 2. For each $N \geq 10$ there is a set $A \subset [N]$ of size
\[
|A| \leq \frac{N}{(\log N)^p} \exp \left\{ (2/3) \sqrt{\log \log N \log \log \log N} \right\},
\]
for which $|AA| \sim M_N$ as $N \to \infty$.

The construction of extremal sets satisfying the required properties in either Theorem 1 or 2 requires an analysis of the structure of integers in the “multiplication table” $[N][N]$, as worked out in [For08a]. From this work, we know that most elements of $[N][N]$ have $\log \log t + O(1)$ prime factors, and moreover, these prime factors are not “compressed at the bottom”, meaning that for most $n \in [N][N]$ we have
\[
\#\{p|n : p \leq t\} \leq \frac{\log \log t}{\log 2} + O(1) \quad (3 \leq t \leq N).
\]
Here the terms $O(1)$ should be interpreted as being bounded by a sufficiently large constant $C = C(\epsilon)$, where $\epsilon$ is the relative density of exceptional elements of $[N][N]$. This suggests that candidate extremal sets $A$ should consist of integers with about half as many prime factors; that is, $\omega(n) \approx \log \log 4$.

In a sequel paper, we will refine the estimates in Theorems 1 and 2. In particular, we will show that the threshold size of $A$ for the property $|AA| \sim |A|^2/2$ is genuinely smaller than the threshold size of $|A|$ for the property $|AA| \sim M_N$. More precisely, we will show that if $|A| \leq \frac{N}{(\log N)^p} \exp\{O(\sqrt{\log \log N})\}$, then $|AA| \not\sim M_N$. The proof requires a much more intricate analysis of the arguments in the papers [For08a] and [For08b].

Acknowledgements. The author is grateful to Sergei Konyagin for bringing the paper [CRR17] to his attention, and for helpful conversations.

2. Preliminaries

Here $\omega(n)$ is the number of distinct prime factors of $n$. $\omega(n,t)$ is the number of prime factors $p|n$ with $p \leq t$. $\Omega(n)$ is the number of prime power divisors of $n$, $\Omega(n,t)$ is the number of prime powers $p^a|n$ with $p \leq t$. We analyze the distribution of these functions using a simple, but powerful technique known as the parametric method (or the “tilting method” in probability theory).

For brevity, we use the notation $\log_k x$ for the $k$-th iterate of the logarithm of $x$.

Lemma 2.1. Let $f$ be a real valued multiplicative function such that $0 \leq f(p^a) \leq 1.9^a$ for all primes $p$ and positive integers $a$. Then, for all $x > 1$ we have
\[
\sum_{n \leq x} f(n) \ll \frac{x}{\log x} \exp \left( \sum_{p \leq x} \frac{f(p)}{p} \right).
\]

Proof. This is a corollary of a more general theorem of Halberstam and Richert; see Theorem 01 of [HT88] and the following remarks.

In the special case $f(n) = \lambda^{\Omega(n)}$, where $0 < \lambda \leq 1.9$, we get by Mertens’ estimate the uniform bound
\[
\sum_{n \leq x} \lambda^{\Omega(n,t)} \ll x(\log t)^{\lambda^{-1}}.
\]
This is useful for bounding the tails of the distribution of $\Omega(n, t)$.

3. **Proof of Theorem 1**

Define

$$k = \left\lfloor \log_2 N \right\rfloor \log 4$$

and let

$$B = \left\{ \frac{N}{2} < m \leq N : m \text{ squarefree, } \omega(m) = k, \omega(m, t) \leq \frac{\log t}{\log 4} + 2 \ (3 \leq t \leq N) \right\}.$$ 

Our proof of Theorem 1 has three parts:

(i) establish a lower bound on the size of $B$, showing that the upper bound on $\omega(n, t)$ affects the size of $B$ only mildly;

(ii) give an upper bound on the multiplicative energy $E(B)$, which shows that there are few nontrivial solutions of $b_1b_2 = b_3b_4$; consequently, the product set $BB$ is large; and

(iii) select a thin random subset $A$ of $B$ that has the desired properties, an idea borrowed from Proposition 3.2 of [CRR17].

**Lemma 3.1.** We have

$$|B| \gg \frac{N}{(\log N)^\theta (\log_2 N)^{3/2}}.$$ 

**Lemma 3.2.** Let $E(B) = |\{(b_1, b_2, b_3, b_4) \in B^4 : b_1b_2 = b_3b_4\}|$ be the multiplicative energy of $B$. Then

$$E(B) \ll |B|^2(\log_2 N)^4.$$ 

**Lemma 3.3.** Given $B \subset [N]$ with $E(B) \leq |B|^2 f(N)$ and $f(N) \leq |B|^{1/2}$, let $A$ be a subset of $B$ where the elements of $A$ are chosen at random, each element $b \in B$ chosen with probability $\rho$ satisfying $\rho^2 = o(1/f(N))$ and $\rho|B|^2 \gg |N|^{1.1}$ as $N \to \infty$. Then with probability $\to 1$ as $N \to \infty$, we have $|A| \sim \rho|B|$ and $|AA| \sim \frac{1}{2}|A|^2$.

Assuming these three lemmas, it is easy to prove Theorem 1. We apply Lemma 3.3 with $f(N) = C(\log_2 N)^4$, invoking the energy estimate from Lemma 3.2 and the size bound from Lemma 3.1. For any function $g(N) \to \infty$ as $N \to \infty$, we take

$$\rho = \frac{1}{(\log_2 N)^2 g(N)}$$

and deduce that there is a set $A \subset [N]$ of size

$$|A| \sim \rho|B| \gg \frac{N}{(\log N)^\theta (\log_2 N)^{7/2} g(N)},$$

such that $|AA| \sim \frac{1}{2}|A|^2$.

Now we prove the three lemmas.
Hence, by the proof of the aforementioned theorem, we obtain
\[ \log_2 p_j(m) \geq (j - 2) \log 4 \quad (1 \leq j \leq \omega(m)). \]
Indeed, the assertion is trivial if \( t < p_1(m) \) since in this case \( \omega(m, t) = 0 \). If \( p_1(m) \leq t \leq N \), set \( j = \max\{i : t \geq p_i(m)\} \). Then
\[ \omega(m, t) = j \leq \frac{\log_2 p_j(m)}{\log 4} + 2 \leq \frac{\log_2 t}{\log 4} + 2. \]
Thus,
\[ |B| \geq |\{N/2 < m \leq N : \omega(m) = k, m \text{ squarefree}, \log_2 p_j(m) \geq j \log 4 - 2 \log 4 (1 \leq j \leq \omega(m))\}|. \]
This is closely related to the quantity
\[ N_k(x; \alpha, \beta) = |\{m \leq x : \omega(m) = k, \log_2 p_j(m) \geq \alpha j - \beta(1 \leq j \leq k)\}|, \]
as defined in [For07]. In fact, the lower bound in [For07, Theorem 1] for \( N_k(x; \alpha, \beta) \) is proved under the additional conditions that \( m \) is squarefree and lies in a dyadic range ([For07, §4]), although this is not stated explicitly. Thus, the proof of [For07, Theorem 1] applies to lower-bounding \( |B| \). In the notation of [For07], we have
\[ k = \left\lfloor \frac{\log_2 N}{\log 4} \right\rfloor, \quad A = \frac{1}{\log 4}, \quad \alpha = \log 4, \quad \beta = 2 \log 4, \quad u = 2, \quad v = \frac{\log_2 N}{\log 4}, \quad w = \frac{\log_2 N}{\log 4} - k + 3. \]
Taking \( \varepsilon = 0.1 \), one easily verifies the required conditions for [For07, Theorem 1]:
\[ \alpha - \beta \leq A, \quad w \geq 1 + \varepsilon, \quad e^{\alpha(w-1)} - e^{\alpha(w-2)} \geq 1 + \varepsilon. \]
Hence, by the proof of the aforementioned theorem, we obtain
\[ |B| \gg \frac{N(\log_2 N)^{k-2}}{(\log N)(k-1)!}, \]
from which the conclusion follows by Stirling’s formula. \( \square \)

**Proof of Lemma 3.2** Set
\[ \beta_{13} = \gcd(b_1, b_3), \quad \beta_{14} = \gcd(b_1, b_4), \quad \beta_{23} = \gcd(b_2, b_3), \quad \beta_{24} = \gcd(b_2, b_4), \]
so that
\[ b_1 = \beta_{13} \beta_{14}, \quad b_2 = \beta_{23} \beta_{24}, \quad b_3 = \beta_{13} \beta_{23}, \quad b_4 = \beta_{14} \beta_{24}. \]
Since \( 1/2 \leq b_1/b_4 \leq 2 \), it follows that \( 1/2 \leq \beta_{13}/\beta_{24} \leq 2 \) and likewise that \( 1/2 \leq \beta_{14}/\beta_{23} \leq 2 \). By reordering variables, we may assume without loss of generality that \( \min(\beta_{13}, \beta_{24}) \gg N^{1/2} \). For some parameter \( T \), which is a power of 2 and satisfies \( T = O(N^{1/2}) \), we have
\[ T \leq \beta_{14} < 2T. \]
This implies that \( T/2 \leq \beta_{23} \leq 4T \) and \( N/8T \leq \beta_{13}, \beta_{24} \leq 2N/T \). We also note that
\[ \omega(b_j, 4T) = \Omega(b_j, 4T) \leq z_T (1 \leq j \leq 4), \quad z_T = \frac{\log_2 (4T)}{\log 4} + 2. \]
Let \( \lambda_1, \lambda_2 \in (0, 1) \) be two parameters to be chosen later. Let \( U_T \) be the number of solutions of
\[ b_1 b_2 = b_3 b_4 \quad (b_j \in B) \]
also satisfying (3.1). Using (3.2), we see that

\[ U_T \leq \sum_{\beta_{14}, \beta_{23} \leq 4T} \sum_{\beta_{24}, \beta_{13} \leq 2N/T} \prod_{j=1}^{2} \lambda_1^{\Omega(\beta_{j4}, 4T)} + z_T \prod_{j=3}^{4} \lambda_1^{\Omega(\beta_{j3}, 4T)} - z_T \lambda_2^{\Omega(\beta_{j1}, \beta_{j2})} - k \]

\[ = \lambda_1^{-4zT} \lambda_2^{-4k} \sum_{\beta_{14}, \beta_{23} \leq 4T} \sum_{\beta_{24}, \beta_{13} \leq 2N/T} \lambda_1^{2\Omega(\beta_{14}, \beta_{23}) + 2\Omega(\beta_{13}, \beta_{24})} \lambda_2^{2\Omega(\beta_{13}, \beta_{24}, \beta_{23})} \]

\[ = \lambda_1^{-4zT} \lambda_2^{-4k} \left( \sum_{\beta \leq 4T} (\lambda_1^2 \lambda_2^2)^{\Omega(\beta)} \right)^2 \left( \sum_{\beta \leq 2N/T} \lambda_1^{2\Omega(\beta, 4T)} \lambda_2^{2\Omega(\beta)} \right)^2 . \]

An application of Lemma 2.1 yields

\[ U_T \ll \lambda_1^{-4zT} \lambda_2^{-4k} \left( T(\log T) \lambda_1^2 \lambda_2^2 - 1 \right)^2 \left( \frac{N}{T} (\log N) \lambda_1^2 - 1 (\log T) \lambda_1^2 \lambda_2^2 - \lambda_2^2 \right)^2 \]

\[ = \lambda_1^{-4zT} \lambda_2^{-4k} N^2 (\log N)^{2\lambda_2^2 - 2} (\log T)^{4\lambda_1^2 \lambda_2^2 - 2} \lambda_2^2 - 2\lambda_2^2 . \]

We optimize by taking \( \lambda_1^2 = \frac{1}{2} \) and \( \lambda_2^2 = \frac{1}{\log 4} \), so that

\[ U_T \ll \frac{N^2}{(\log N)^{2\theta}(\log T)} . \]

Summing over \( T = 2^r \ll N^{1/2} \) yields

\[ E(B) \ll \frac{N^2 \log N}{(\log N)^{2\theta}} \ll |B|^2 (\log_2 N)^4 , \]

using Lemma 3.1.

\[ \square \]

Proof of Proposition 3.3 This is similar to the proof of Proposition 3.2 of [CRK17]. First, if elements of \( A \) are chosen from \( B \) with probability \( \rho \), then by easy first and second moment calculations,

\[ E|A| = \rho |B| , \quad E(|A| - \rho |B|)^2 = O(\rho |B|) , \]

where \( E \) denotes expectation. By Chebyshev’s inequality, \( |A| \sim \rho |B| \) with probability tending to 1 as \( N \to \infty \). By the proof of Proposition 3.2 of [CRK17], we also have

\[ E|AA| = \sum_x \left( 1 - (1 - \rho^2)^{\tau_B(x)/2} \right) + O(\rho N) , \]

where

\[ \tau_B(x) = | \{ x = b_1 b_2 : b_1, b_2 \in B \} | . \]

Now \( (1 - z)^k = 1 - k z + O((kz)^2) \) uniformly for \( 0 \leq z \leq 1 \) and \( k \geq 1 \), and so

\[ E|AA| = (\rho^2/2) \sum_x \tau_B(x) + O(\rho^4 \sum_x \tau_B^2(x)) + O(\rho N) \]

\[ = (\rho^2/2) |B|^2 + O(\rho^4 E(B) + \rho N) \]

\[ = \left( \frac{1}{2} + o(1) \right) (\rho |B|)^2 . \]
Since $|A| \sim \rho |B|$ with probability tending to 1 as $N \to \infty$, and also $|AA| \leq \frac{1}{2} |A|(|A| + 1)$ for all $|A|$, we conclude that $|AA| \sim \frac{1}{2} |A|^2$ with probability tending to 1 as $N \to \infty$. \hfill \Box

4. PROOF OF THEOREM 2

Again let

$$k = \left\lfloor \log_2 N \over \log 4 \right\rfloor.$$

Define

$$A = \{ m \leq N : \Omega(m) \leq k + r \}, \quad r = 2 \sqrt{\log_2 N \log_3 N}.$$

By (2.1), we have the size bound

$$|A| \leq \sum_{m \leq N} \left( \frac{1}{\log 4} \right)^{\Omega(m)-(k+r)} \leq \frac{N(\log 4)^r}{(\log N)^\theta} \leq \frac{N}{(\log N)^\theta} \exp\{2/3 \sqrt{\log_2 N \log_3 N}\}$$

using (1.1). Next, we show that $|AA| \sim M_N$. Let $B = [N] \setminus A$. It suffices to show that

$$|B[N]| \leq |AB| + |BB| = o(M_N).$$

Let $c = ab$, where $a \leq N$ and $b \in B$, and consider two cases: (i) $\Omega(c) > 2k + h$, where $h = \lfloor 5 \log_3 N \rfloor$, and (ii) $\Omega(c) < 2k + h$. We then have $|B[N]| \leq D_1 + D_2$, where $D_1$ is the number of integers $c \leq N^2$ with $\Omega(c) > 2k + h$, and $D_2$ is the number of pairs $(a, b) \in [N]^2$ with $\Omega(ab) \leq 2k + h$ and $\Omega(b) \geq k + r$. We will show that each of these is small, essentially by exploiting the imbalance in prime factors of $a$ and $b$ implied in the conditions on $D_2$. By (2.1) and (1.1),

$$D_1 \ll \sum_{c \leq N^2} \left( \frac{1}{\log 2} \right)^{\Omega(c)-(2k+h)} \leq \frac{N^2}{(\log N)^{2\theta}(1/\log 2)^h} = o(M_N),$$

in light of estimate (1.2). Next, choose parameters $0 < \lambda_2 < 1 < \lambda_1 < 1.9$. Then

$$D_2 \ll \sum_{a, b \leq N} \lambda_2^{\Omega(ab)-(2k+h)} \lambda_1^{\Omega(b)-(k+r)} \leq \lambda_2^{-(2k+h)} \lambda_1^{-(k+r)} N^2 (\log N)^{\lambda_2 + \lambda_1 - 2},$$

invoking (2.1) again. A near-optimal choice for the parameters is

$$\lambda_2 = 1 - \frac{x}{\log 4}, \quad \lambda_1 = 1 + \frac{x}{1 - x}, \quad x = r \frac{\log 4}{\log_2 N}.$$

A little algebra reveals that the previous upper bound on $D_2$ is bounded by

$$\ll N^2 (\log N)^{-2\theta - \frac{2}{\log 4} \log \log (1 + x) \log (1 + x) + (1 - x) \log (1 - x)} - \frac{h}{\log_2 N} \log \frac{1}{|x|}. $$

By Taylor’s expansion,

$$(1 + x) \log(1 + x) + (1 - x) \log(1 - x) \geq x^2 \quad (|x| < 1)$$

and therefore the exponent of $\log N$ is at most

$$-2\theta - \frac{x^2}{\log 4} + \frac{h \log \log (4 + o(1))}{\log_2 N} \leq -2\theta - 4 \log 4 \log_3 N \log_2 N + 1.7 \frac{\log_3 N}{\log_2 N} \leq -2\theta - 3.8 \frac{\log_3 N}{\log_2 N}.$$

We get that $D_2 \ll N^2 (\log N)^{-2\theta} (\log_2 N)^{-3.8} = o(M_N)$ and Theorem 2 follows.
REFERENCES

[CRR17] J. Cilleruelo, D. S. Ramana, and O. Ramaré. Quotient and product sets of thin subsets of the positive integers. *Tr. Mat. Inst. Steklova*, 296(Analiticheskaya i Kombinatornaya Teoriya Chisel):58–71, 2017.

[CRS18] Pablo Candela, Juanjo Rué, and Oriol Serra. Memorial to Javier Cilleruelo: A problem list. *INTEGERS: The electronic journal of combinatorial number theory*, 18(A28):1–9, 2018.

[For07] Kevin Ford. Generalized Smirnov statistics and the distribution of prime factors. *Funct. Approx. Comment. Math.*, 37(part 1):119–129, 2007.

[For08a] K. Ford. The distribution of integers with a divisor in a given interval. *Ann. of Math.* (2), 168(2):367–433, 2008.

[For08b] K. Ford. Integers with a divisor in $(y, 2y]$. In *Anatomy of integers*, volume 46 of *CRM Proc. Lecture Notes*, pages 65–80. Amer. Math. Soc., Providence, RI, 2008.

[HT88] Richard R. Hall and Gérald Tenenbaum. *Divisors*, volume 90 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1988.

DEPARTMENT OF MATHEMATICS, 1409 WEST GREEN STREET, UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN, URBANA, IL 61801, USA

E-mail address: ford@math.uiuc.edu