COLLOQUIA MATHEMATICA SOCIETATIS JÁNOS BOLYAI
51. NUMBER THEORY, BUDAPEST (HUNGARY), 1987

ON PRODUCTS OF SEQUENCES OF INTEGERS

C. POMERANCE[1] and A. SÁRKÖZY[2]

1. INTRODUCTION

In this paper our goal is to show that if A and B are "dense" sets of integers then there are "many" distinct products of the form ab where $a \in A$, $b \in B$. Furthermore, we will show that this fact can be applied to study certain multiplicative "hybrid problems", i.e., multiplicative problems involving both general sets and special sets.

In 1960, Paul Erdős [1] showed the following, surprising result: the number of distinct integers of

[1] Research supported in part by an NSF grant.
[2] Research partially supported by Hungarian National Foundation for Scientific Research, Grant No. 1811.

the form ab where a, b are natural numbers not exceeding x is

$$(1.1) \quad x^2(\log x)^{-\alpha+o(1)} ,$$

where $\alpha = 1 - \log(e \log 2)/\log 2 = 0.0860\ldots$. Thus we have the seemingly paradoxical result that only $o(x^2)$ integers may be found in the "multiplication table" of the integers up to x.

This paradox may be explained via the function $\nu(n)$, the number of distinct prime factors of n. It has been known since Hardy and Ramanujan [3] that the normal order of $\nu(n)$ is loglog n. Thus a "normal product" of integers $a,b \le x$ would have about 2 loglog x prime factors, which is quite *abnormal* for integers below x^2. In fact, the bulk of the products ab making up the count (1.1) come from factors a, b with less than the normal number of prime factors.

Thus it should be expected that if a certain thin subset is deleted from the set of integers up to x, then in fact there should be considerably fewer products ab than the count in (1.1). Indeed this is true and easy to see using the results in the aforementioned paper of Hardy and Ramanujan (see (2.14) below). By taking

$$A = B = \{n \le x \ : \ \nu(n) \ge \log\log x - (\log\log x)^{2/3}\},$$

then $|A| = |B| = (1 + o(1))x$, while

$$(1.2) \quad |AB| \le x^2(\log x)^{1-2\log 2 + o(1)}$$

where $|\ |$ denotes the cardinality of the enclosed set and AB denotes the set of products ab where $a \in A$, $b \in B$.

It may be asked if the drop in the exponent from $-\alpha$ to $1 - 2\log 2$ between (1.1) and (1.2) can be induced to drop still further by choosing the sets A, B a bit thinner. The principal result in this paper is that the expression on the right of (1.2) essentially gives the correct lower bound for $|AB|$ so long as A, B are "dense". We shall prove the following result.

THEOREM 1. If $\varepsilon > 0$, $\delta > 0$ are arbitrary, then there exists some $x_0 = x_0(\varepsilon,\delta)$ such that if $x > x_0$, $A, B \subset \{1,2,\ldots,[x]\}$ and

$$(1.3) \quad |A| > \varepsilon x, \ |B| > \varepsilon x,$$

then

$$(1.4) \quad |AB| > x^2(\log x)^{1-2\log 2 - \delta}$$

Section 2 below will be devoted to proving Theorem 1. It should be noted that the method of proof could also be used to treat the case when the ε in the theorem is not fixed, but allowed to tend to 0 "slowly". In section 3 it will be shown that the numerousness of the product set AB given by (1.4) implies there are products that are "close" to some number of a special sequence. Specifically, the special sequences considered are the primes and the integers free of large prime factors.

Fianally we remark that Theorem 1 implies we have equality in (1.2)

2. PROOF OF THEOREM 1

We begin by showing that we may replace A by a "dense" subset A_0 such that the elements of A_0 do not have too many prime factors, all have the same largest square factor and are all in an interval of the form $[u,2u]$ (and similarly for the set B).

Let $t(n)$ denote the largest integer such that $t(n)^2 | n$. If t is a natural number, then the number of $n \le x$ with $t(n) > t$ is at most

$$\sum_{i=t+1}^{\infty} \left[\frac{x}{i^2}\right] < x \sum_{i=t+1}^{\infty} \frac{1}{i^2} < x \int_t^{\infty} \frac{1}{s^2}\, ds = \frac{x}{t} .$$

Thus if A' is the set of $a \in A$ with $t(a) \le 1 + 2/\varepsilon$, then

$$|A'| > \frac{\varepsilon}{2} x .$$

Hence, if $A(t)$ is the set of $a \in A$ with $t(a) = t$, then there is some

$$t_0 \le 1 + 2/\varepsilon$$

with

$$|A(t_0)| > \frac{\varepsilon}{2(1+2/\varepsilon)} x > \frac{\varepsilon^2}{6} x .$$

Let R be an integer so that

$$2^{-R} \le \frac{\varepsilon^2}{12} < 2^{-R+1} ,$$

i.e., $R = [\log_2(12/\varepsilon^2)]$. Then

$$\frac{\varepsilon^2}{6} x < |A(t_0)| \le 2^{-R}x + \sum_{r=1}^{R} |A(t_0) \cap [2^{-r}x,\ 2^{-r+1}x]|$$

$$\le \frac{\varepsilon^2}{12} x + R \cdot \max_{1\le r\le R} |A(t_0) \cap [2^{-r}x,\ 2^{-r+1}x]|$$

implies there is some u of the form $2^{-r}x$ with

$$|A(t_0) \cap [u,2u]| > \frac{\varepsilon^2}{12R} x .$$

Let A_0 be the set of $a \in A(t_0) \cap [u,2u]$ with

(2.1) $\quad \nu(a) \le [\log\log x + (\log\log x)^{2/3}] := T$.

Since by [3] the number of $a \le x$ for which (2.1) fails is $o(x)$, we have, for large x,

(2.2) $\quad |A_0| > \frac{\varepsilon^2}{13R} x$.

To summarize, A_0 is a subset of A for which (2.2) holds and every member a of A_0 satisfies $t(a) = t_0$, $u \le a \le 2u$ and (2.1). Similarly there are numbers t_1, v such that if B_0 is the set of $b \in B$ for which $t(b) = t_1$, $v \le b \le 2v$ and $\nu(b)$ satisfies (2.1), then $|B_0| > (\varepsilon^2/(13R))x$.

Let D denote the set of integers $d = a/t_0^2$ where $a = A_0$ and similarly let E denote the set of integers $e = b/t_1^2$ where $b \in B_0$. Then every member of A_0B_0 is of the form $det_0^2t_1^2$ where $d \in D$, $e \in E$, so that

(2.3) $\quad |AB| \ge |A_0 B_0| = |DE|$.

Thus is suffices to estimate $|DE|$.

Note that

(2.4) $\quad |D|, \ |E| > \frac{\varepsilon^2}{13R} x$,

every member of D, E is squarefree with at most T distinct prime factors and

(2.5) $\quad D \subset \left[\frac{u}{t_0^2}, \frac{2u}{t_0^2}\right], \ E \subset \left[\frac{v}{t_1^2}, \frac{2v}{t_1^2}\right]$

where $2u, 2v \le x$.

Let us denote the number of solutions of

$$de = n, \ d \in D, \ e \in E$$

by $f(n)$. By the Cauchy-Schwarz inequality we have

(2.6) $\quad |DE| = \sum_{f(n)>0} 1 \ge (\sum_n f(n))^2 / \sum_n f(n))^2$.

In view of (2.4) we have

(2.7) $\quad \sum_n f(n) = |D \times E| = |D||E| > \frac{\varepsilon^4}{169R^2} x^2$,

so that it suffices to give an upper bound for the mean square

$$M_2 := \sum_n (f(n))^2 = \sum_n \Big(\sum_{\substack{de=n \\ d\in D, e\in E}} 1 \Big)^2 = \sum_{\substack{d_1e_1=d_2e_2 \\ d_1,d_2\in D \\ e_1,e_2\in E}} 1 .$$

In other words, M_2 denotes the number of solutions of

$$d_1e_1 = d_2e_2,\ d_1,\ d_2 \in D \ \text{ and } \ e_1,\ e_2 \in E$$

or, in equivalent form,

$$\frac{d_1}{d_2} = \frac{e_2}{e_1}\ ,\ d_1,\ d_2 \in D \ \text{ and } \ e_1,\ e_2 \in E \ .$$

Let us write the rational number in this equation in reduced form:

$$(2.8) \qquad \frac{d_1}{d_2} = \frac{e_2}{e_1} = \frac{p}{q}\ ,\ \text{where } (p,q) = 1\ .$$

It follows from (2.5) that $1/2 \le p/q \le 2$ and thus there is a positive integer $k \le [1 + \log x]$ with

$$(2.9) \qquad e^{k-2} < p,q \le e^k\ .$$

By (2.8) there exist positive integers r, s such that

$$(2.10) \qquad d_1 = rp,\ d_2 = rq,\ e_2 = sp,\ e_1 = sq\ .$$

By (2.9) we have

$$(2.11) \qquad r = \frac{d_1}{p} \le \frac{x}{e^{k-2}}\ ,\ s = \frac{e_2}{p} \le \frac{x}{e^{k-2}}\ .$$

By (2.10) and the fact that $d_1,\ d_2 \in D,\ e_1,\ e_2 \in E$ we have

$$\nu(d_1) = \nu(r) + \nu(p) \le T,\ \nu(d_2) = \nu(r) + \nu(q) \le T,$$

$$\nu(e_2) = \nu(s) + \nu(p) \le T,\ \nu(e_1) = \nu(s) + \nu(q) \le T$$

and hence

$$(2.12) \qquad \max(\nu(r),\ \nu(s)) + \max(\nu(p),\ \nu(q)) \le T\ .$$

From (2.9), (2.11) and (2.12) there exist integers $1 \le k \le [1 + \log x]$ and $0 \le \ell \le T$ with

$$p,q \in \{n \le e^k:\ \nu(n) \le \ell\}\ ,$$

$$r,s \in \{n \le x/e^{k-2}:\ \nu(n) \le T - \ell\}\ .$$

Hence the number M_2 of quadruplets $p,\ q,\ r,\ s$ is not greater than

$$M_2 \le \sum_{k=1}^{[1+\log x]} \sum_{\ell=0}^{T} |\{n \le e^k : \nu(n) \le \ell\}|^2\, |\{n \le e^{k-2} : \nu(n) \le T - \ell\}|^2$$

$$(2.13) \qquad \le T^3 \sum_{k=1}^{[1+\log x]} \max_{i+j\le T} \left(\pi_i(e^k)\ \pi_j\left(\frac{x}{e^{k-2}}\right)\right)^2\ ,$$

where by $\pi_t(y)$ we mean the number of $n \le y$ with $\nu(n) = t$.

From [3] we have absolute positive constants c_1, c_2 such that for all natural numbers t and every $y \geq 3$

$$(2.14) \qquad \pi_t(y) \leq c \frac{y}{\log y} \frac{(\log\log y + c_2)^{i-1}}{(i-1)!} .$$

Thus for $1 \leq k \leq [1 + \log x]$ and $i + j \leq T$ we have

$$\pi_i(e^k)\ \pi_j\left(\frac{x}{e^{k-2}}\right) \ll \frac{e^k}{k} \cdot \frac{(\log k + c_2)^{i-1}}{(i-1)!} \cdot$$

$$\cdot \frac{x\, e^{-k+2}}{\log(xe^{-k+2})} \cdot \frac{(\log\log(xe^{-k+2}) + c_2)^{j-1}}{(j-1)!} \ll$$

$$(2.15) \qquad \ll \frac{x}{k(\log x - k)} \cdot \frac{1}{(i+j-2)!} \binom{i+j-2}{i-1} (\log k + c_2)^{i-1}$$

$$(\log\log(x^{e^{-k+2}}) + c_2)^{j-1} \leq \frac{x}{k(\log x - k)} \cdot$$

$$\cdot \frac{1}{(i+j-2)!} (\log k + \log\log(xe^{-k+2}) + 2x_2)^{i+j-2} .$$

Since $\log k + \log\log(xe^{-k+2}) > \log\log x$, we have from (2.15) that

$$\pi_i(e^k)\ \pi_j\left(\frac{x}{e^{k-2}}\right) \ll \frac{x}{k(\log x - k)}$$

$$(2.16) \qquad \left(\frac{T}{\log\log x}\right)^{T-(i+j-2)} \frac{(\log k + \log(\log x - k) + c_3)^T}{T!}$$

$$\ll L(x) \frac{x}{k(\log x - k)} \frac{(\log k + \log(\log x - k))^T}{T!}$$

where $L(x) := \exp\{(\log\log x)^{2/3} + (\log\log x)^{1/3}\} = (\log x)^{o(1)}$.

Since the terms $k = \ell$ and $k = \log x - \ell$ play a symmetric role in (2.16) we obtain from (2.13) that

$$M_2 \ll L(x)^3 \sum_{k=1}^{[\frac{1}{2}\log x + 1]} \left(\frac{x}{k(\log x - k)} \frac{e^T(\log k + \log(\log x - k))^T}{T^T}\right)^2$$

$$\ll \frac{x^2 L(x)^3}{(\log x)^2} \sum_{k=1}^{[\frac{1}{2}\log x + 1]} \frac{e^{2T}}{k^2} \left(\frac{\log k + T}{T}\right)^{2T}$$

$$\leq x^2 L(x)^5 \sum_{1 \leq \ell \leq \log[\frac{1}{2}\log x + 1]} \sum_{k=[e^{\ell-1}]}^{[e^\ell]} \frac{1}{k^2} \left(1 + \frac{\log k}{T}\right)^{2T}$$

$$(2.17)\quad \ll x^2 L(x)^5 \sum_{1\le \ell \le \log[1/2 \log x + 1]} \frac{1}{e^{\ell}} \left(1 + \frac{\ell}{T}\right)^{2T} .$$

Writing $F(u) = e^{-u}(1 + u/T)^{2T}$ it is easy to see that the function $F(u)$ is increasing on $0 \le u \le T$. Thus from (2.17) we obtain

$$(2.18)\quad \begin{aligned} M_2 &\ll x^2 L(x)^5 (\log\log x) \frac{2^{2T}}{e^T} \ll x^2 L(x)^6 \left(\frac{4}{e}\right)^{\log\log x} \\ &= x^2 (\log x)^{2\log 2 - 1} L(x)^6 . \end{aligned}$$

Finally, putting (2.18) into (2.6) and using (2.3) and (2.7) we have

$$(2.19)\quad |AB| \gg \frac{\varepsilon^4}{169R^2 L(x)^6} x^2(\log x)^{1-2 \log 2}$$

where the implied constant is absolute. Sinde $L(x) = (\log x)^{o(1)}$ and $R \ll \log(1/\varepsilon)$, we have our theorem.

REMARK. By changing the definition of T in (2.1) to $[(1 + \eta(x))\log\log x]$ where $\eta(x) \to 0^+$ sufficiently slowly, the above proof would give a stronger result where ε is allowed to be any function of the form $1/(\log x)^{o(1)}$. In

fact if $\delta > 0$ is fixed in the theorem, we may choose ε as small as $1/(\log x)^{c_4\delta}$ where $c_4 > 0$ is some absolute constant.

3. APPLICATIONS

In a recent paper, Iwaniec and Sárközy [6] proved that if A, B are "dense" sets of integers, then there is a product $ab \in AB$ which is "near" a square. Specifically, they showed that if $A, B \subset \{1,2,\dots,[x]\}$, $|A|, |B| > \varepsilon x$, then for $x \ge x_1(\varepsilon)$ there is a solution to the inequality

$$|ab - n^2| < (x \log x)^{1/2}, \quad a \in A,\ b \in B,\ n \in \mathbb{Z} .$$

A result such as this does not immediately follow from Theorem 1 above since the number of integers $m \le x^2$ such that

$$|m - n^2| < (x \log x)^{1/2}$$

for some integer n is $\ll x^{3/2}(\log x)^{1/2}$.

However, there are special sets other than the squares for which an easy application of Theorem 1 gives a result that appears to be worth stating.

THEOREM 2. Let $\varepsilon > 0$, $\delta > 0$ be arbitrary. Then there exists a constant $x_0 = x_0(\varepsilon,\delta)$ such that if $x \geq x_0$, $A, B \subset \{1,2,\dots,[x]\}$ and $|A|, |B| > \varepsilon x$, then there exist $a \in A$, $b \in B$ and a prime p such that

$$|ab - p| < x^{1/5 + \delta} . \tag{3.1}$$

Assuming the Riemann hypothesis, the right side of (3.1) may be replaced with $(\log x)^{1+2 \log = + \delta}$

THEOREM 3. With the same hypotheses as Theorem 2 there is an $x_0 = x_0(\varepsilon,\delta)$ such that if $x \geq x_0$ then for any $y \geq \exp\{(\log x)^{5/6 + \delta}\}$ there is a solution to the inequality

$$|ab - n| \leq y \cdot \exp\{(2 \log x)^{1/6}\}, \quad a \in A, \ b \in B$$

where no prime factor of n exceeds y. Moreover if $\eta > 0$ is arbitrary, then there is an $x_1 = x_1(\varepsilon,\delta,\eta)$ such that if $x \geq x_1$, there is a solution to the inequality

$$|ab - n| \leq x^{\delta}, \quad a \in A, \ b \in B$$

where no prime factor of n exceeds x^{η}.

(Note that the full strength of Theorem 1 is needed only in the proof of the second half of Theorem 2, while

the first half of Theorem 2 and Theorem 3 follow from any estimate of the type $|AB| > x^2(\log x)^{-c}$, and, in fact, this can be proved much more easily with, say, $c = 2 \log 2 + \delta$.)

PROOF OF THEOREM 2. Let p_n denote the n-th prime. A theorem of Harman [4] is that

$$\sum_{\substack{p_n \leq x \\ p_{n+1} - p_n \geq x^{\frac{1}{10}+\delta}}} (p_{n+1} - p_n) \ll \frac{x}{\log x} \tag{3.2}$$

for any fixed $\delta > 0$. Replacing δ by $\frac{1}{2}\delta$ and x by x^2 in this result, Theorem 1 implies there is some ab with $a \in A$, $b \in B$ that is not in any interval $[p_n, p_{n+1}]$ with $p_n \leq x^2$ and $p_{n+1} - p_n \geq x^{\frac{1}{5}+\delta}$. Then the closest prime p to ab satisfies (3.1).

Now assume the Riemann hypothesis holds. From Selberg [7] we have

$$\sum_{\substack{p_n \leq x \\ p_{n+1} - p_n \geq H}} (p_{n+1} - p_n) \ll \frac{x}{H} (\log x)^2$$

uniformly for all $H \geq 1$. Thus our result follows by replacing x with x^2 and choosing $H = (\log x)^{1+2 \log 2 + \delta}$.

REMARK. In [5], Harman announced a result analogous to (3.2) with 1/10 replaced by 1/12 and with $x/\log x$ replaced by $o(x)$. If this $o(x)$ can be improved to $x(\log x)^{1-2\log 2-\delta}$, then the 1/5 in (3.1) can be replaced with 1/6.

PROOF OF THEOREM 3. The first result follows immediately from Theorem 1 above and Theorem 6 in Friedlander and Lagarias [2] while the second result follows from Theorem 1 above and Theorem 5 in [2].

REFERENCES

[1] P. ERDŐS, An asymptotic inequality in the theory of numbers, *Vestnik Leningrad Univ.* 15 (1960), 41-49 (in Russian).

[2] J.B. FRIEDLANDER and J.C. LAGARIAS, On the distribution in short intervals of integers having no large prime factor, *J. Number Theory* 25 (1987), 249-273.

[3] G.H. HARDY and S. RAMANUJAN, The normal number of prime factors of a number n, *Quarterly J. Math.* 48 (1920), 76-92.

[4] G. HARMAN, Primes in short intervals, *Math. Z.* 180 (1982), 335-348.

[5] G. HARMAN, On the distribution of αp mod one, *J. London Math. Soc.* (2) 27 (1983), 9-18.

[6] H. IWANIEC and A. SÁRKÖZY, On a multipliative hybrid problem, *J. Number Theory* 26 (1987), 89-95.

[7] A. SELBERG, On the normal density of primes in short intervals and the difference between consecutive primes, *Arch. Math. Naturvid.* 47 (1943), 87-105.

A. SÁRKÖZY
MTA MKI
Budapest
Reáltanoda u. 13-15.
H-1053

C. POMERANCE
Department Mathematics
University of Georgia
Athens Georgia
U S A